DATA CENTER HANDBOOK

Hwaiyu Geng

WILEY

DATA CENTER HANDBOOK

DATA CENTER HANDBOOK

HWAIYU GENG, P.E.
Amica Association
Palo Alto, CA, USA

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Data center handbook / edited by Hwaiyu Geng.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-43663-9 (cloth)
1. Electronic data processing departments–Design and construction–Handbooks, manuals, etc.
2. Electronic data processing departments–Security measures–Handbooks, manuals, etc. I. Geng, Hwaiyu.
TH4311.D368 2015
004.068′4–dc23

2014013900

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To "Our Mothers Who Cradle the World," and To "Our Earth Who Gives Us Life."

BRIEF CONTENTS

CONTENTS

CONTRIBUTORS

Guy AlLee, Intel Corporation, Hillsboro, OR, USA

Hanwook Bae, Samsung SDS, Seoul, South Korea

Ken Baudry, P.E., K.J Baudry, Inc., Atlanta, GA, USA

Sofia Bergqvist, IBM Corporation, Stockholm, Sweden

David Bonneville, P.E., S.E., Degenkolb Engineers, San Francisco, CA, USA

Lynn Brown, P.E., LEED AP, QCxP, Encotech Engineering, Austin, TX, USA

Omar Cherkaoui, Ph.D., University of Quebec à Montréal (UQÀM), Montreal, Quebec, Canada

Henry Coles, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Chris Crosby, Compass Datacenters, Dallas, TX, USA

Chris Curtis, Compass Datacenters, Dallas, TX, USA

Nicolas H. Des Champs, Ph.D., Munters Corporation, Buena Vista, VA, USA

Jay Dietrich, Distinguished Engineer, IBM Corporation, Essex Junction, VT, USA

Sean S. Donohue, P.E., Hughes Associates, Inc., Colorado Springs, CO, USA

Keith Dunnavant, P.E., Munters Corporation, Buena Vista, VA, USA

Chris Gabriel, Logicalis Group, London, UK

Venkata Vish Ganti, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Hwaiyu Geng, P.E., Amica Association, Palo Alto, CA, USA

Girish Ghatikar, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Steve Greenberg, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Taewon Han, Ph.D., Rutgers, The State University of New Jersey, New Brunswick, NJ, USA

Sarah Hanna, MSEE, Facebook, Inc., Menlo Park, CA, USA

Mark Harris, Nlyte Software, San Mateo, CA, USA

Ching-I Hsu, Ph.D., Raritan, Inc., Somerset, NJ, USA

Jingyi Hu, China Electronics Standardization Institute, Beijing, China

Phil Hughes, Clustered Systems Company, Inc., Santa Clara, CA, USA

Phil Isaak, P.E., P.Eng., DCDC, RCDD Isaak Technology, Inc., Minneapolis, MN, USA

Alexander Jew, J&M Consultants, Inc., San Francisco, CA, USA

Masatoshi Kajimoto, ISACA, Tokyo, Japan

William Kao, Ph.D., University of California Santa Cruz, Silicon Valley Extension, Santa Clara, CA, USA

Mingoo Kim, Ph.D., Samsung SDS, Seoul, South Korea

William J. Kosik, P.E., CEM, LEED AP, BEMP, Hewlett-Packard Company, Chicago, IL, USA

Minseok Kwon, Samsung SDS, Seoul, South Korea

Ian Levine, Hewlett-Packard Company, Albany, NY, USA

Haibo Li, China Electronics Standardization Institute, Beijing, China

Zhe Liu, China Electronics Standardization Institute, Beijing, China

Chris Loeffler, Eaton, Raleigh, NC, USA

Malik Megdiche, Ph.D., Schneider Electric, Grenoble, France

Ramesh Menon, IBM Corporation, Gaithersburg, MD, USA

Dave Moody, Schneider Electric ITB, O'Fallon, MO, USA

Veerendra Mulay, Ph.D., Facebook, Menlo Park, CA, USA

Liam Newcombe, Romonet, London, UK

Jay S. Park P.E., Facebook, Inc., Menlo Park, CA, USA

Robert Peckelnicky, P.E., S.E., Degenkolb Engineers, San Francisco, CA, USA

Mary Ann Piette, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Joe Prisco, IBM Corporation, Rochester, MN, VT, USA

Peter Sacco, PTS Data Center Solution, Inc., Oakland, NJ, USA

Mark Seymour, Future Facilities Limited, London, UK

Matt Slaby, Hewlett-Packard Company, Houston, TX, USA

Hongru Song, China Electronics Standardization Institute, Beijing, China

Vali Sorell, P.E., Syska Hennessy Group, Charlotte, NC, USA

Ed Spears, Eaton, Raleigh, NC, USA

Wade Vinson, Hewlett-Packard Company, Houston, TX, USA

John Weale, P.E., LEED AP, The Integral Group, Oakland, CA, USA

Yutao Yang, China Electronics Standardization Institute, Beijing, China

PREFACE

Designing and operating a sustainable data center (DC) requires technical knowledge and skills from strategic planning, complex technologies, available best practices, optimum operating efficiency, disaster recovery, and more.

Engineers and managers all face challenges operating across functionalities, for example, facilities, IT, engineering, and business departments. For a mission-critical, sustainable DC project, we must consider the following:

- What are the goals?
- What are the givens?
- What are the constraints?
- What are the unknowns?
- Which are the feasible solutions?
- How is the solution validated?
- How does one apply technical and business knowledge to develop an optimum solution plan that considers emerging technologies, availability, scalability, sustainability, agility, resilience, best practices, and rapid time to value?

The list can go on and on. Our challenges may be as follows:

- To prepare a strategic location plan
- To design and build a mission critical DC with energy efficient infrastructure
- To apply best practices thus consuming less energy
- To apply IT technologies such as cloud and virtualization and
- To manage DC operations thus reducing costs and carbon footprint

A good understanding of DC components, IT technologies, and DC operations will enable one to plan, design, and implement mission-critical DC projects successfully.

The goal of this handbook is to provide DC practitioners with essential knowledge needed to implement DC design and construction, apply IT technologies, and continually improve DC operations. This handbook embraces both conventional and emerging technologies, as well as best practices that are being used in the DC industry. By applying the information contained in the handbook, we can accelerate the pace of innovations to reduce energy consumption and carbon emissions and to "Save Our Earth Who Gives Us Life."

The handbook covers the following topics:

- DC strategic planning
- Hosting, colocation, site selection, and economic justifications
- Plan, design, and implement a mission critical facility
- IT technologies including virtualization, cloud, SDN, and SDDC
- DC rack layout and MEP design
- Proven and emerging energy efficiency technologies
- DC project management and commissioning
- DC operations
- Disaster recovery and business continuity

Each chapter includes essential principles, design and operations considerations, best practices, future trends, and further readings. The principles cover fundamentals of a technology and its applications. Design and operational considerations include system design, operations, safety, security, environment issues, maintenance, economy, and best practices. There are useful tips for planning, implementing, and controlling operational processes. The future trends and further reading sections provide visionary views and lists of relevant books, technical papers, and websites for additional reading.

This *Data Center Handbook* is specifically designed to provide technical knowledge for those who are responsible for the design, construction, and operation of DCs. It is also useful for DC decision makers who are responsible for strategic decisions regarding capacity planning and technology investments. The following professionals and managers will find this handbook to be a useful and enlightening resource:

- C-level Executives (Chief Information Officer, Chief Technology Officer, Chief Operating Officer, Chief Financial Officer)
- Data Center Managers and Directors
- Data Center Project Managers
- Data Center Consultants
- Information Technology and Infrastructure Managers
- Network Operations Center and Security Operations Center Managers
- Network, Cabling, and Communication Engineers
- Server, Storage, and Application Managers
- IT Project Managers
- IT Consultants
- Architects and MEP Consultants
- Facilities Managers and Engineers
- Real Estate Portfolio Managers
- Finance Managers

This *Data Center Handbook* is prepared by more than 50 world-class professionals from eight countries around the world. It covers the breadth and depth of DC planning, designing, construction, and operating enterprise, government, telecommunication, or R&D Data Centers. This *Data Center Handbook* is sure to be the most comprehensive single-source guide ever published in its field.

Hwaiyu Geng, CMfgE, P.E.

ACKNOWLEDGMENTS

The *Data Center Handbook* is a collective representation of an international community with scientists and professionals from eight countries around the world. Fifty-one authors, from data center industry, R&D, and academia, plus fifteen members at Technical Advisory Board have contributed to this book. Many suggestions and advice were received while I prepared and organized the book.

I gratefully acknowledge the contributors who dedicated their time in spite of their busy schedule and personal lives to share their wisdom and valuable experience.

I would also like to thank the members at Technical Advisory Board for their constructive recommendations on the structure of this handbook and thorough peer review of book chapters.

My thanks also go to Brett Kurzman, Alex Castro, Katrina Maceda at Wiley and F. Pascal Raj at SPi Global whose can do spirit and teamwork were instrumental in producing this book.

Thanks and appreciation must go to the following individuals for their advice, support, and contributions:

Sam Gelpi	Hewlett-Packard Company
Dongmei Huang, Ph.D.	Rainspur Technology, China
Madhu Iyengar, Ph.D.	Facebook Inc.
Johnathan Jew	J&M Consultants
Jonathan Koomey, Ph.D.	Stanford University
Tomoo Misaki	Nomura Research Institute, Ltd., Japan
Veerendra Mulay, Ph.D.	Facebook Inc.
Jay Park, P.E.	Facebook Inc.
Roger Schmidt, Ph.D.	IBM Corporation
Hajime Takagi	GIT Associates, Ltd., Japan
William Tschudi, P.E.	Lawrence Berkeley National Laboratory
Kari Capone	John Wiley & Sons, Inc.

This book benefited from the following organizations and institutes:

- 7 × 24 Exchange International
- American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE)
- Building Industry Consulting Service International (BICSI)
- Datacenter Dynamics
- European Commission Code of Conduct
- The Green Grid
- Japan Data Center Council
- Open Compute Project
- Silicon Valley Leadership Group
- Telecommunications Industry Association (TIA)
- Uptime Institute/451 Research
- U.S. Department of Commerce, National Institute of Standards and Technology
- U.S. Department of Energy, Lawrence Berkeley National Laboratory
- U.S. Department of Energy, Oak Ridge National Laboratory
- U.S. Department of Energy, Office of Energy Efficiency & Renewable Energy
- U.S. Department of Homeland Security, Federal Emergency Management Administration
- U.S. Environmental Protection Agency, ENERGY STAR Program
- U.S. Green Building Council, Leadership in Energy & Environmental Design

My special thanks to my wife, Limei, my daughters, Amy and Julie, and grandchildren for their understanding, support, and encouragement when I was preparing this book.

TECHNICAL ADVISORY BOARD

David Bonneville, S.E., Degenkolb Engineers San Francisco, California

John Calhoon, Microsoft Corporation Redmond, Washington

Yihlin Chan, Ph.D., OSHA (Retiree) Salt Lake City, Utah

Sam Gelpi, Hewlett-Packard Company Palo Alto, California

Hwaiyu Geng, P.E., Amica Association Palo Alto, California

Magnus Herlin, Ph.D., ANCIS Incorporated San Francisco, California

Madhu Iyengar, Ph.D., Facebook Inc. Menlo Park, California

Jonathan Jew, J&M Consultants San Francisco, California

Jacques Kimman, Ph.D., Zuyd University Heerlen, Netherlands

Jonathan Koomey, Ph.D., Stanford University Stanford, California

Veerendra Mulay, Ph.D., Facebook Inc. Menlo Park, California

Dean Nelson, eBay Inc. San Jose, California

Jay Park, P.E., Facebook Inc. Menlo Park, California

Roger Schmidt, Ph.D., IBM Corporation Poughkeepsie, New York

Jinghua Zhong, China Electronics Engineering Design Institute Beijing, China

CHAPTER ORGANIZATION

This book is designed to cover following five major parts:

- Part 1: Data Center Overview and Strategic Planning
- Part 2: Data Center Design and Construction
- Part 3: Data Center Technology
- Part 4: Data Center Operations and Management
- Part 5: Disaster Recovery and Business Continuity

This organization allows readers to have an overview of data centers including strategic planning, design and construction; the available technologies and best practices; how to efficiently and effectively manage a data center and close out with disaster recovery and business continuity. Within 5 parts, there are 36 chapters.

PART 1: DATA CENTER OVERVIEW AND STRATEGIC PLANNING

Chapter 1—Data Centers—Strategic Planning, Design, Construction, and Operations: This chapter provides high-level discussion of some key elements in planning and designing data centers. It covers the definition of data centers; vision; principles in preparing a roadmap and strategic planning; global location planning; sustainable design relating to reliability, computational fluid dynamics, DCIM, and PUE; best practices; proven and emerging technologies; and operations management. It concludes with disaster recovery and business continuity. All of these subjects are described in more detail within the handbook.

Chapter 2—Energy and Sustainability in Data Centers: This chapter gives an overview of best practices in designing and operating data centers that would reduce energy consumption and achieve sustainability.

Chapter 3—Hosting or Colocation Data Centers: This chapter describes the definition of hosting, colocation, and data center. It explores 'build vs. buy" with financial considerations. It also describes the elements to consider in evaluating and selecting hosting or colocation providers.

Chapter 4—Modular Data Centers: Design, Deployment, and other Considerations: An anatomy of modular data center using ISO container standards is presented. The benefits and applications using MDC as well as site preparation, installation, and commissioning are introduced.

Chapter 5—Data Center Site Search and Selection: This chapter gives you a roadmap for site search and selection, process, and team members, and critical elements that lead to a successful site selection are described.

Chapter 6—Data Center Financial Analysis, ROI, and TCO: This chapter starts with fundaments of financial analysis (NPV, IRR), return on investment, and total cost of ownership. Case studies are used to illustrate NPV, breakeven, and sensitivity analysis in selecting different energy savings retrofits. It also includes an analysis of "Choosing to build, reinvest, least, or rent" of data centers, colocation, and cloud.

Chapter 7—Overview Data Centers in China: Overview of policies, laws, regulations, and GB (standards) of China's data centers is presented. Development status, distribution and energy efficiency of data centers, and cloud are discussed.

Chapter 8—Overview of Data Centers in Korea: Overview of policies, laws, regulations, codes and standards, and market of Korea's data centers is presented. Design and construction practices of Korea's data centers are discussed.

PART 2: DATA CENTER DESIGN AND CONSTRUCTION

Chapter 9—Architecture Design: Data Center Rack Floor Plan and Facility Layout Design: An overview of server rack, cabinet, network, and large frame platform is introduced. Computer room design with coordination of HVAC system, power distribution, fire detection and protection system, lighting, raised floor vs. overhead system, and aisle containment is discussed. Modular design, CFD modeling, and space planning are also addressed.

Chapter 10—Mechanical Design in Data Centers: Design criteria including reliability, security, safety, efficiency, and flexibility are introduced. Design process with roles and responsibilities from predesign, schematics design, design development, construction documents, and construction administration are well explained. Considerations in selecting key mechanical equipment and best practices on energy efficiency practices are also discussed.

Chapter 11—Electrical Design in Data Centers: Electrical design requirements, uptime, redundancy, and availability are discussed.

Chapter 12—Fire Protection and Life Safety in Data Centers: Fundamentals of fire protection, codes and stands, local authorities, and life safety are introduced. Passive fire protection, early detection, and alarm and signaling systems are discussed. Hot and cold aisle ventilations are reviewed.

Chapter 13—Structural Design in Data Centers: Natural Disaster Resilience: Strengthening building structural and nonstructural components are introduced. Building design using code based vs. performance based is discussed. New design considerations and mitigation strategies relating to natural disasters are proposed. This chapter concludes with comprehensive resiliency strategies with pre- and postdisaster planning.

Chapter 14—Data Center Telecommunication Cabling: Telecommunication cabling organizations and standards are introduced. The spaces, cabling topology, cable type, cabinet/rack placement, pathways, and energy efficiency are discussed. It concludes with discussion on patch panel, cable management, and reliability tiers.

Chapter 15—Dependability Engineering for Data Center Infrastructures: This chapter starts with definition of system dependability analysis. System dependability indexes including reliability, availability, and maintainability are introduced. Equipment dependability data including MTTF, MTBF, and failure rate are also introduced. System dependability, redundancy modeling, and system dysfunctional analysis are discussed.

Chapter 16—Particulate and Gaseous contamination in Data Centers: IT equipment failure rates between using outside air vs. recirculated air are discussed. ISO standards addressing particulate cleanliness, ANSI standards evaluating gaseous contamination, and ASHRAE TC9.9 Committee on particulate and gaseous contaminations are addressed.

Chapter 17—Computational Fluid Dynamics Applications in Data Centers: Fundamentals and theory of CFD are introduced. Applying CFD in data centers including design, troubleshooting, upgrade, and operations management are discussed. Modeling data centers that include CRAC/CRAH, cooling infrastructure, control system, time-dependent simulation, and failure scenarios are performed. This chapter concludes with benefits of CFD and future virtual facility.

Chapter 18—Environment Control of Data Centers: Thermal management of data centers including structural parameters, placement of CRAC units, cooling system design and control, and data center design are discussed. Energy management of data centers including airside or waterside economizer, CRAH, liquid cooling, and dynamic cooling are discussed.

Chapter 19—Data Center Project Management and Commissioning: This chapter describes project management that involves planning, scheduling, safety and security, tracking deliverables, test and commissioning, and training and operations. Commissioning tasks starting from design stage all the way through test and commissioning to final occupancy phases are discussed. This chapter details how to select a commissioning team, what equipment and systems to be tested and commissioned, and roles and responsibilities of commissioning team at different stage of project life cycle.

PART 3: DATA CENTER TECHNOLOGY

Chapter 20—Virtualization, Cloud, SDN, and SDDC: Fundamentals of virtualization, cloud, SDN, and SDDC are described. What benefits and challenges of those technologies to data center practitioners are described.

Chapter 21—Green Microprocessor and Server Design: This chapter concerns itself with microprocessor and server design on how to judge and select them as the best fit to sustainable data centers. This chapter starts with guiding principles to aid your server selection process. It follows in detail by the prime criteria for the microprocessor and server system, as well as, considerations with respect to storage, software, and racks.

Chapter 22—Energy Efficiency Requirements in Information Technology Equipment Design: This chapter addresses energy efficiency of servers, storage system, and uninterruptible power supply (UPS) being used in data centers. Each device is being examined at component level and in operating condition as regards how to improve energy efficiency with useful benchmark.

Chapter 23—Raised Floors versus Overhead Cooling in Data Centers: This chapter discusses benefits and challenges between raised floors cooling vs. overhead cooling in the areas of air delivery methodology, air flow dynamics, and underfloor air distribution.

Chapter 24—Hot Aisle versus Cold Aisle Containment: This chapter covers design basics of models for airflow architecture using internal and external cooling units. Fundamentals of hot/cold aisle containments and airflow management systems are presented. The effects of increased return air temperatures at cooling units from HAC are discussed. Concerns with passive ducted return air systems are discussed. HAC and CAC impacts on cooling fan power and redundancy with examples are provided. Consideration is given to peripheral equipment and economizer operations.

Chapter 25—Free Cooling Technologies in Data Centers: This chapter describes how to use ambient outside air to cool a data center. What is economizer thermodynamic process with dry-bulb and wet-bulb temperatures has been discussed. Air to air heat exchanger vs. an integer air to air and cooling tower is reviewed. Comparative energy savings and reduced mechanical refrigeration are discussed.

Chapter 26—Rack-Level Cooling and Cold Plate Cooling: Fundamentals and principles of rack level cooling are introduced. Energy consumption for conventional room cooling vs. rack level cool is discussed. Advantages and disadvantages of rack level cooling including enclosed, in flow, rear door, and cold plate cooling are discussed.

Chapter 27—Uninterruptible Power Supply System: UPSs are an important part of the electrical infrastructure where high levels of power quality and reliability are required. In this chapter, we will discuss the basics of UPS designs, typical applications where UPS are used, considerations for energy efficiency UPS selection, and other components and options for purchasing and deploying a UPS system.

Chapter 28—Using Direct Current Networks in Data Centers: This chapter addresses why AC power, not DC power, is being used. Why DC power should be used in data centers and trending in using DC power.

Chapter 29—Rack PDU for Green Data Centers: An overview of PDU fundamentals and principles are introduced. PDUs for data collection that includes power energy, temperature, humidity, and air flow are discussed. Considerations in selecting smart PDUs are addressed.

Chapter 30—Renewable and Clean Energy for Data Centers: This chapter discusses what is renewable energy, the differences between renewable and alternative energy, and how they are being used in data centers.

Chapter 31—Smart Grid-Responsive Data Centers: This chapter examines data center characteristics, loads, control systems, and technologies ability to integrate with the modern electric grid (Smart Grid). The chapter also provides information on the Smart Grid architecture, its systems, and communication interfaces across different domains. Specific emphasis is to understand data center hardware and software technologies, sensing, and advanced control methods, and how they could be made responsive to identify demand response (DR) and automated DR (auto-DR) opportunities and challenges for Smart Grid participation.

PART 4: DATA CENTER OPERATIONS AND MANAGEMENT

Chapter 32—Data Center Benchmark Metrics: This chapter provides information on PUE, xUE, RCI, and RTI. This chapter also describes benchmark metrics being developed or used by SPEC, the Green 500, and EU Code of Conduct.

Chapter 33—Data Center Infrastructure Management: This chapter covers what DCiM is, where it stands in hype cycle, why it is important to deploy DCiM in data centers, what are modules of a DCiM, what are future trends, and how to select and implement a DCiM system successful.

Chapter 34—Computerized Maintenance Management System for Data Centers: This chapter covers the basics of CMMS, why it is important to deploy CMMS, what CMMS modules included, maintenance service process, management and reporting, and how to select, implement, and operate a CMMS in a data center successfully.

PART 5: DISASTER RECOVERY AND BUSINESS CONTINUITY

Chapter 35—Data Center Disaster Recovery and High Availability: This chapter aims to give a sense of the key design elements, planning and process approaches to maintain the required level of service and business continuity from the data centre and the enterprise architectures residing within disaster recovery and high availability.

Chapter 36—Lessons Learned from Natural Disasters and Preparedness of Data Centers: This chapter covers lessons learned from two major natural disasters that will broaden data center stakeholders toward natural disaster awareness, prevention, and preparedness. Detailed lessons learned from the events are organized in the following categories: Business Continuity/Disaster Recovery Planning, Communications, Emergency Power, Logistics, Preventive Maintenance, Human Resources, and Information Technology. They can be easily reviewed and applied to enhance your BC/DR planning.

PART I

DATA CENTER OVERVIEW AND STRATEGIC PLANNING

1

DATA CENTERS—STRATEGIC PLANNING, DESIGN, CONSTRUCTION, AND OPERATIONS

Hwaiyu Geng
Amica Association, Palo Alto, CA, USA

1.1 INTRODUCTION

In a typical data center, electrical energy is used to operate Information and Communication Technology (ICT) equipment and its supporting facilities. About 45% of electrical energy is consumed by ICT equipment, which includes servers, storages, and networks. The other 55% of electrical energy is consumed by facilities, which include power distribution system, uninterruptible power supplies, chillers, computer room air conditioners, lights, and so on. Improving power consumption by ICT equipment and facilities is imperative for efficient use of energy. Many studies have proven increasing greenhouse gases due to human activities resulting in global warming.

1.1.1 Data Centers and Global Warming

A study by the journal *Science* estimates that, from 1992 to 2012, the melting ice from Greenland and Antarctica has raised the global sea level by 11.1 mm (0.43 in.). Rising sea levels have gained more attention from the flooding caused by the superstorm *Sandy* in 2012 that struck the heavily populated U.S. East Coast.

A report titled *Climate Change 2013: The Physical Science Basis* [1], prepared by the Intergovernmental Panel on Climate Change (IPCC), set up by the World Meteorological Organization and the UN's Environment Program, states as follows: "Warming of the climate system is unequivocal. Since the 1950s, many of the observed changes are unprecedented over decades to millennia. The atmosphere and ocean have warmed, the amounts of snow and ice have diminished, sea level has risen, and the concentrations of greenhouse gases have increased". "The rate of sea level rise since the mid-nineteenth century has been larger than the mean rate during the previous two millennia (*high confidence*). Over the period 1901–2010, global mean sea level rose by 0.19 [0.17–0.21] m."

The World Bank issued a report in November 2012, titled *Turn Down the Heat: Why a 4°C Warmer World Must be Avoided* [2]. The report describes what the world would be like if it warmed by 4°C (7.2°F). "The 4°C world scenarios are devastating: the inundation of coastal cities; increasing risks for food production potentially leading to higher malnutrition rates; many dry regions becoming dryer, wet regions wetter; unprecedented heat waves in many regions, especially in the tropics; substantially exacerbated water scarcity in many region, increase frequency of high-intensity tropical cyclones; and irreversible loss of biodiversity, including coral reef system."

"The science is unequivocal that humans are the cause of global warming, and major changes are already being observed: global mean warming is 0.8°C above pre-industrial levels; oceans have warmed by 0.09°C since the 1950s and are acidifying; sea levels rose by about 20 cm since pre-industrial times and are now rising at 3.2 cm per decade; an exceptional number of extreme heat waves occurred in the last decade; major food crop growing areas are increasingly affected by drought."

Human beings generate all kinds of heat from cooking food, manufacturing goods, building houses, passenger and freight transport, and ICT activities. ICT continues as a pervasive force in the global economy, which includes Internet surfing, computing, online purchase, online banking, mobile phone, social networking, medical services, and exascale

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

machine (supercomputer). They all require energy in data centers and give out heat as a result. One watt input to process data results in 1 W of heat output. As a result, all data centers take energy and give out heat. We can't stop giving out heat, but we can reduce heat output by efficiently managing energy input.

1.1.2 Data Center Definition

The term "data center" means differently to different people. Some of the names used include data center, data hall, data farm, data warehouse, computer room, server room, R&D software lab, high-performance lab, hosting facility, colocation, and so on. The U.S. Environment Protection Agency defines a data center as:

- "Primarily electronic equipment used for data processing (servers), data storage (storage equipment), and communications (network equipment). Collectively, this equipment processes, stores, and transmits digital information."
- "Specialized power conversion and backup equipment to maintain reliable, high-quality power, as well as environmental control equipment to maintain the proper temperature and humidity for the ICT equipment."

Data centers are involved in every aspect of life running Amazon, AT&T, CIA, Citibank, Disneyworld, eBay, FAA, Facebook, FEMA, FBI, Harvard University, IBM, Mayo Clinic, NASA, NASDAQ, State Farm, U.S. Government, Twitter, Walmart, Yahoo, Zillow, etc. This A–Z list reflects the "basic needs" of food, clothing, shelter, transportation, health care, and social activities that cover the relationships among individuals within a society.

A data center could consume electrical power from 1 to over 500 MW. Regardless of size and purpose (Table 1.1), all data centers serve one purpose, and that is to process information. In this handbook, we use "data center" that refers to all names stated earlier.

1.1.3 Energy Consumption Trends

"Electricity used in global data centers during 2010 likely accounted for between 1.1 and 1.5% of total electricity use, respectively. For the U.S., that number was between 1.7 and 2.2%" [3].

IDC IVIEW, sponsored by EMC Corporation, stated [4] as follows: "Over the next decade, the number of servers (virtual and physical) worldwide will grow by a factor of 10, the amount of information managed by enterprise data centers will grow by a factor of 50, and the number of files the data center will have to deal with will grow by a factor of 75, at least."

Gartner estimated [5], "In 2011, it is believed that 1.8 Zettabytes of data was created and replicated. By 2015, that number is expected to increase to 7.9 Zettabytes. That is equivalent to the content of 18 million Libraries of Congress. The majority of data generation originates in North America and Europe. As other global regions come online more fully, data generation is expected to increase exponentially."

Evidently, as a result of increasing activities such as big data analytics, online services, mobile broadband, social activities, commercial business, manufacturing business, health care, education, medicine, science, and engineering, energy demand will continue to increase.

1.1.4 Using Electricity Efficiently

A data center houses ICT equipment and facilities that are used to cool ICT equipment. While air cooling is still the most economical way to cool servers in racks, water cooling is the most efficient way to remove heat generated by processors.

Based on "Power Usage Effectiveness, March 2012"[2] prepared by LBNL, 33.4% of total energy is used in power and cooling a data center and 66.6% by IT load (Fig. 1.1). For a typical server, 30% of power is consumed by a processor and 70% by peripheral equipment that includes power supply, memory, fans, drive, and so on. A server's utilization efficiency is estimated to be at a disappointing 20% [6].

TABLE 1.1 Data center type, server volume, and typical size

Facility types	Volume servers	Estimated servers per facility	Typical size in sq. ft. (m^2)	Estimated number of facilities (in the United States)	2006 electric use (billion kWh)
Server closets	1,798,000	1–2	<200 (19)	900,000–1,500,000	3.5
Server rooms	2,120,000	3–36	<500 (46)	50,000–100,000	4.3
Localized data center	1,820,000	36–300	<1000 (93)	10,000–13,000	4.2
Midtier data center	1,643,000	300–800	<5000 (465)	2,000–4,000	3.7
Enterprise-class data center	3,215,000	800–2000+	5000+ (465+)	1,000–2,500	8.8

Sources: EPA, 2007; CHP in Data Centers, ICF International, Oak Ridge National Laboratory, 2009.

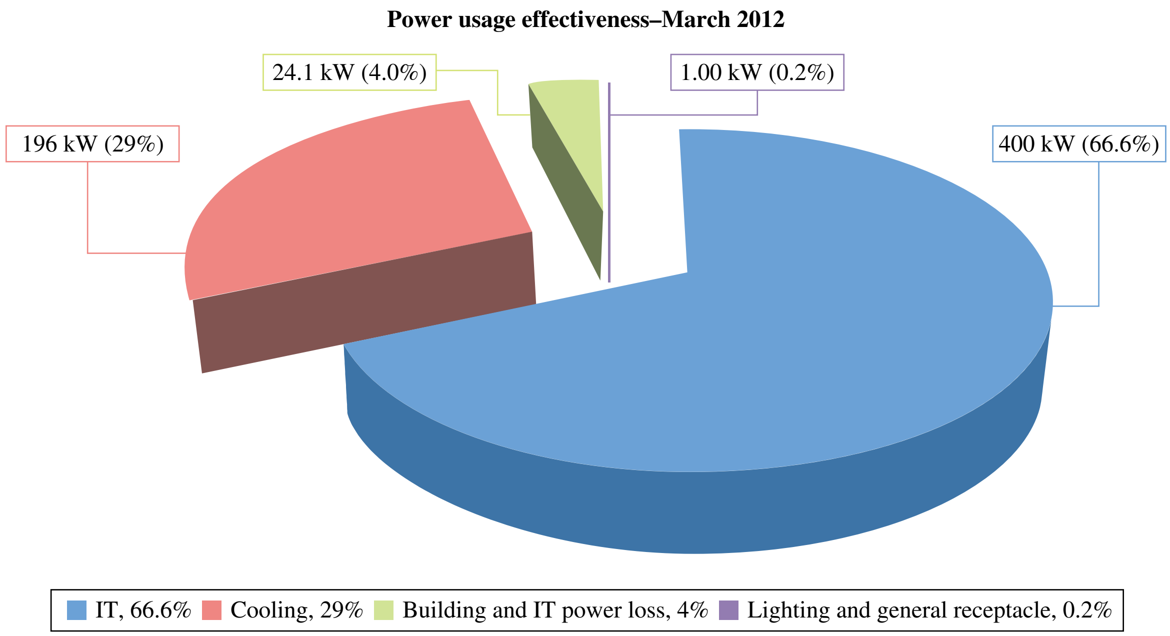


FIGURE 1.1 The DOE national average PUE for data centers is 1.75. 50B-1275 data center has evolved from an average PUE of 1.65 (calculated in 2009) to today's 1.47. Getting there, staying there, and further improving the PUE is an ongoing effort (Source: Nina Lucido, Data Center Utilization Report, March 2012, LBNL, U.S. Department of Energy. https://commons.lbl.gov/display/itdivision/2012/04).

Opportunities of saving energy at the server level include the use of ENERGY STAR-rated equipment, water cooling server, solid-state drive, and variable-speed fan in servers. Virtualization could be applied to improve the server's utilization efficiency.

1.1.5 Virtualization, Cloud, Software-Defined Data Centers

As illustrated in Figure 1.2, "Virtualization is a method of running multiple independent virtual operating systems on a single physical computer. It is a way of allowing the same amount of processing to occur on fewer servers by increasing server utilization. Instead of operating many servers at low CPU utilization, virtualization combines the processing power onto fewer servers that operate at higher utilization [7]."

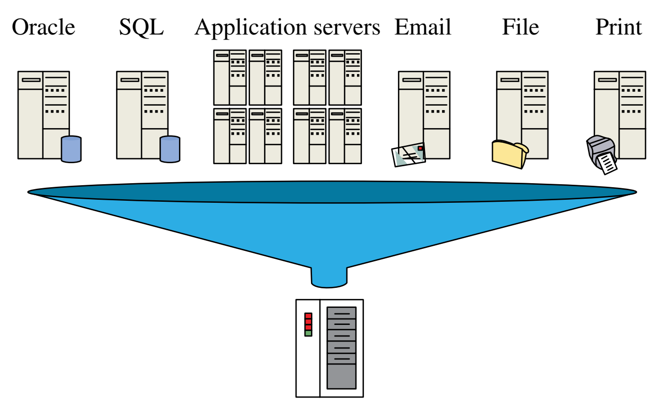


FIGURE 1.2 Virtualization (Source: https://www.energystar.gov/index.cfm?c=power_mgt.datacenter_efficiency_virtualization).

Cloud computing is an evolving model [8]. It is characterized as easy access, on demand, rapidly adaptable, flexible, cost-effective, and self-service to share pool of computing resources that include servers, storage, networks, applications, and services. Cloud capacity could be rapidly provisioned, controlled, and measured.

Cloud computing provides various service models including Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS). HP's "Everything as a Service" provides service model as follows: "Through the cloud, everything will be delivered as a service, from computing power to business processes to personal interactions." Cloud computing is being deployed in public, private, community, or hybrid cloud models. It benefits data center managers by offering resource pooling and optimizing resource uses with lower costs. IDC estimate that by 2015, 20% of the information will be "touched" by cloud computing.

The Software-Defined Data Center (SDDC), pioneered by VMware, is an architectural approach that has all ICT infrastructure (server, storage, networking, and security) virtualized through hardware-independent management system. SDDC can be a building block to Cloud, or Cloud can be an extension of an SDDC [9]. Virtual machines can be deployed in minutes with little human involvement. Provisioning applications can be operational in minutes that shorten time to value. SDDC maximizes the utilization of physical infrastructure [10]. As a result, SDDC reduces capital spending, advances asset utilization, improves operational efficiency, and enhances ICT productivity. SDDC is likely to drive down data center hardware costs.

1.2 DATA CENTER VISION AND ROADMAP

Table 1.2 provides a framework of vision, possible potential technology solutions, and key benefits. This table consolidates the ideas and solutions from 60 experts who attended the Vision and Roadmap Workshop on Routing Telecom and Data Centers Toward Efficient Energy Use. The table could be tailored to individual needs by enhancing with emerging technologies such as SDDC, fuel cell technology, etc.

1.2.1 Strategic Planning and Roadmap

Strategic planning for a holistic data center could encompass a global location plan, site selection, design, construction, and operations that support ICT and emerging technology. There is no one "correct way" to prepare a strategic plan. Depending on data center acquisition strategy (i.e., host, colocation, expand, lease, buy, or build) of a new data center, the level of deployments could vary from minor modifications of a server room to a complete build out of a green field project.

Professor Michael E. Porter's "How Competitive Forces Shape Strategy" [12] described the famous "Five Forces" that lead to a state of competition in an industry. They are threat of new entrants, bargaining power of customers, threat of substitute products or services, bargaining power

TABLE 1.2 ICT vision and roadmap summary [11]

Equipment and software	Power supply chain	Cooling
	Visions	
ICT hardware and software will increase the computing power of a watt by at least an order of magnitude, meeting future demand without increasing energy consumption or total cost of ownership and substantially decreasing the environmental footprint of ICT facilities	Reduce power losses in date centers and telecommunications central offices by 50% from service entrance to end use—while maintaining or improving reliability the total cost of ownership	Reduce cooling energy as a percentage of ICT power to a global average of ≤20% for retrofit and <5% for new construction. Cooling systems will be adaptable, scalable, and able to maximize utilization and longevity of all assets over their lifetime—while maintaining system resiliency and lowering total cost of ownership
	Potential technology solutions	
• Advanced power management in ICT hardware • Dynamic network power management • New data storage technologies • Free cooling and equipment standards • Hardened ICT equipment • Novel computing architectures *Game-Changing Technologies* • Nanoelectronic circuitry • All-optical networks • Superconducting components	• Eliminate voltage conversion steps • High-efficiency power system components • Efficiency-optimized control systems • Transition to DC operation • On-site DC generation and microgrid	• Advanced air cooling • Liquid cooling of hardware • Advanced cooling of individual hardware components • Efficiency-optimized control systems
	Key benefits	
• Efficiency gains in ICT equipment as software drive savings in all areas of ICT facilities by reducing loads for the power supply chain and cooling systems • Hardening equipment to perform reliably in extreme environments may obviate or greatly reduce ICT cooling	• Improved efficiency will reduce power system losses and associated cooling loads • Most strategies to reduce power losses focus on reducing the number of voltage steps, which likely will reduce the number and cost of power system components • Green energy can avoid carbon output	• New approaches for cooling can lower energy costs and facilitate greater ICT hardware densities

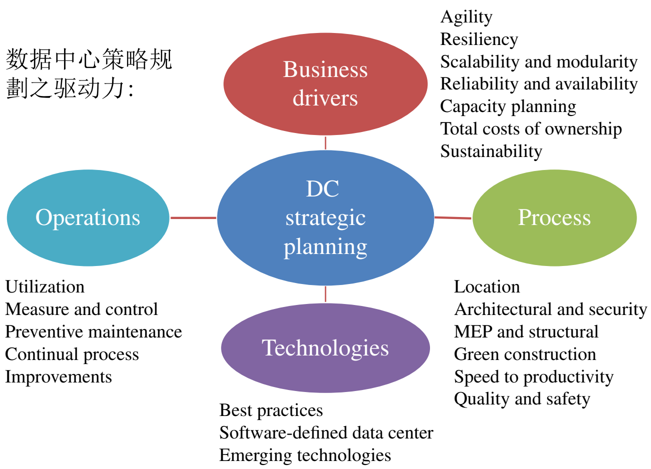


FIGURE 1.3 Data center strategic planning forces (Courtesy of Amica Association).

of suppliers, and the industry jockeying for position among current competitors. Chinese strategist Sun Tzu, in *The Art of War*, stated five factors: the Moral Law, Heaven, Earth, the Commander, and Methods and Discipline. Key ingredients in both strategic planning reflect the following [13]:

- What are the goals
- What are the knowns and unknowns
- What are the constraints
- What are the feasible solutions
- How the solutions are validated
- How to find an optimum solution

In preparing a strategic plan for a data center, Figure 1.3 [14] shows four forces: business driver, process, technologies, and operations. "Known" business drivers and philosophies of a data center solution include the following:

- Agility: Ability to move quickly.
- Resiliency: Ability to recover quickly from an equipment failure or natural disaster.
- Modularity and Scalability: "Step and repeat" for fast and easy scaling of infrastructures.
- Reliability and Availability: Reliability is the ability of equipment to perform a given function. Availability is the ability of an item to be in a state to perform a required function.
- Sustainability: Apply best practices in green design, construction, and operations of data centers to reduce environmental impacts.
- Total cost of ownership: Total life cycle costs of CapEx (e.g., land, building, green design, and construction) and OpEx (e.g., energy costs) in a data center.

Additional "knowns" to each force could be added to suit the needs of individual data center project. It is clear that "knowns" Business Drivers are complicated and sometimes conflicting. For example, increasing resiliency, or flexibility, of a data center will inevitably increase the costs of design and construction as well as continuous operating costs. Another example is that the demand for sustainability will increase the Total Cost of Ownership. "He can't eat his cake and have it too," so it is essential to prioritize business drivers early on in the strategic planning process.

A strategic plan should also consider emerging technologies such as using direct current power, fuel cell as energy source, or impacts from SDDC.

1.2.2 Capacity Planning

Gartner's study indicated that data center facilities rarely meet the operational and capacity requirements of their initial design [15]. It is imperative to focus on capacity planning and resource utilization. Microsoft's top 10 business practices estimated [16] that if a 12 MW data center uses only 50% of power capacity, then every year approximately US$4–8 million in unused capital is stranded in UPS, generators, chillers, and other capital equipment invested.

1.3 STRATEGIC LOCATION PLAN

In determining data center locations, the business drivers include market demands, market growth, emerging technology, undersea fiber-optic cable, Internet exchange points, electrical power, capital investments, and other factors. It is essential to have an orchestrated roadmap to build data centers around global locations. Thus, it is important to develop a strategic location plan that consists of a long-term data center plan from a global perspective and a short-term data center implementation plan. This strategic location plan starts from considering continents, countries, states, cities to finally the data center site.

Considerations for a macro long-term plan that is at continent and country levels include:

- Political and economic stability of the country
- Impacts from political economic pacts (e.g., EU, G8, OPEC, and APEC)
- Gross Domestic Products or relevant indicators
- Productivity and competitiveness
- Market demand and trend
- Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis
- Political, Economic, Social, and Technological (PET) analysis (PEST components)

Considerations for a midterm plan that is at province and city levels include:

- Natural hazards (e.g., earthquake, tsunami, hurricane, tornado, and volcano)
- Electricity sources with dual or multiple electrical grid services
- Electricity rate
- Fiber-optic infrastructure with multiple connectivity
- Public utilities (e.g., natural gas and water)
- Airport approaching corridor
- Labor markets (e.g., educated workforce and unemployment rate)

Considerations for a microterm plan within a city, which is at campus level, include:

- Site size, shape, accessibility, expandability, zoning, and code controls
- Tax incentives from city and state
- Topography, 100-year flood plan, and water table
- Quality of life (staff retention)
- Security and crime rate
- Proximity to airport and rail lines
- Proximity to chemical plant and refinery
- Proximity to electromagnetic field from high-voltage power lines
- Operational considerations

Other tools that could be used to formulate location plans include:

- Operations research
  - Network design and optimization
  - Regression analysis on market forecasting
- Lease versus buy analysis or build lease back
- Net present value
- Break-even analysis
- Sensitivity analysis and decision tree

As a reference, you might consider to compare your global location plan against data centers deployed by Google, Facebook, or Yahoo.

1.4 SUSTAINABLE DESIGN

Every business needs data centers to support changing environment such as new market demanding more capacity, new ICT products consuming higher power that requires rack-level cooling [17], and merge and requisition. Sustainable design is essential because data centers can consume 40–100 times more electricity compared to similar-size office spaces on a square foot basis. Data center design involves architectural, structural, mechanical, electrical, fire protection, security, and cabling systems.

1.4.1 Design Guidelines

Since a data center is heavily involved with electrical and mechanical equipments that cover 70–80% of data center capital costs (Fig. 1.4), oftentimes, a data center is considered an engineer-led project. Important factors for sustainable design encompass overall site planning, A/E design, energy efficiency best practices, redundancy, phased deployment, and so on. Building and site design could work with requirements as specified in the Leadership in Energy and Environment Design (LEED) program. The LEED program is a voluntary certification program that was developed by the U.S. Green Building Council (USGBC). Early on in the design process, it is essential to determine the rack floor plan and elevation plan of the building. The floor plate with large column spacing is best to accommodate the data center's ICT racks and cooling equipment. A building elevation plan must be evaluated carefully to cover needed space for mechanical (HVAC), electrical, structural, lighting, fire protection, and cabling systems. Properly designed column spacing and building elevation ensure appropriate capital investments and minimize operational expenses. Effective space planning will ensure maximum rack locations and achieve power density with efficient and effective power and cooling distribution [18].

International technical societies have developed many useful design guidelines. To develop data center design

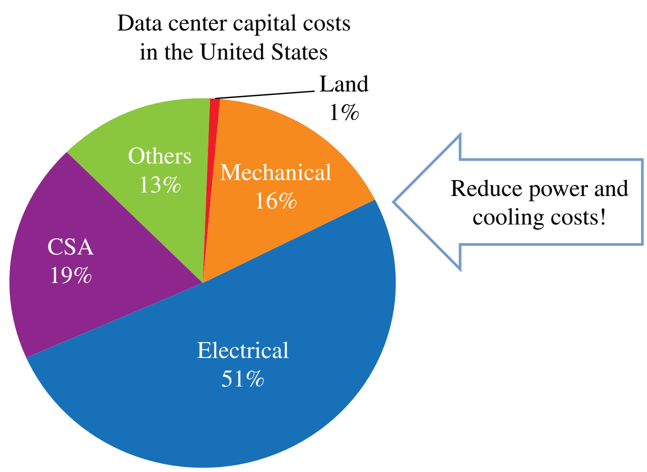


FIGURE 1.4 Focus on mechanical and electrical expenses to reduce cost significantly [16] (Courtesy of Microsoft Corporation).

requirements and specification, the following guidelines could be consulted:

- LEED Rating Systems[1]
- ANSI/ASHRAE/IES 90.1-2010: Energy Standard for Buildings
- ASHRAE TC 9.9 2011: Thermal Guideline for Data Processing Environments—Expanded Data Center Classes and Usage Guidance
- ASHRAE 2011: Gaseous and Particulate Contamination Guidelines for Data Center
- ANSI/BICSI 002-2011: Data Center Design and Implementation Best Practices
- ANSI/TIA-942-A (August 2012): Telecommunications Infrastructure Standard for Data Center
- Data Centre Code of Conduct Introduction Guide (EU)
- 2013 Best Practices Guidelines[2] (EU)
- Outline of Data Center Facility Standard[3] by Japan Data Center Council (JDCC)[4]
- *Code for Design of Information Technology and Communication Room* (GB50174-2008)

1.4.2 Reliability and Redundancy

"Redundancy" ensures higher reliability but it has profound impacts in initial investments and ongoing operating costs.

Uptime Institute® pioneered a tier certification program that structured data center redundancy and fault tolerance in a four-tiered scale. Different redundancies could be defined as follows:

- *N*: base requirement
- *N*+1 redundancy: provides one additional unit, module, path, or system to the minimum requirement
- *N*+2 redundancy: provides two additional units, modules, paths, or systems in addition to the minimum requirement
- 2*N* redundancy: provides two complete units, modules, paths, or systems for every one required for a base system
- 2(*N*+1) redundancy: provides two complete (*N*+1) units, modules, paths, or systems

Based on the aforementioned, a matrix table could be established using the following tier levels in relation to component redundancy categorized by telecommunication, architectural and structural, electrical, and mechanical:

Tier I Data Center: basic system
Tier II Data Center: redundant components
Tier III Data Center: concurrently maintainable
Tier IV Data Center: fault-tolerant

The Telecommunication Industry Association's TIA-942-A [19] contains tables that describe building and infrastructure redundancy in four levels. JDCC's "Outline of Data Center Facility Standard" is a well-organized matrix illustrating "Building, Security, Electric Equipment, Air Condition Equipment, Communication Equipment and Equipment Management" in relation to redundancy Tiers 1, 2, 3, and 4. It is worthwhile to highlight that the matrix also includes seismic design considerations with Probable Maximum Loss (PML) that relates to design redundancy.

The Chinese "National Standard" Code (GB 50174-2008) defines "Design of Information Technology and Communication Room" in A, B, and C tier levels with A being the most stringent.

Data center owners should work with A/E consultants to establish balance between desired reliability, redundancy, and total cost of ownership.

1.4.3 Computational Fluid Dynamics

Whereas data centers could be designed by applying best practices, the locations of systems (e.g., rack, air path, and CRAC) might not be in its optimum arrangement collectively. Computational Fluid Dynamics (CFD) technology has been used in semiconductor's cleanroom projects for decades to ensure uniform airflow inside a cleanroom. CFD offers a scientific analysis and solution to validate cooling capacity, rack layout, and location of cooling units. One can visualize airflow in hot and cold aisles for optimizing room design. During the operating stage, CFD could be used to emulate and manage airflow to ensure that air path does not recirculate, bypass, or create negative pressure flow. CFD could also be used to identify hot spots in rack space.

1.4.4 DCiM and PUE™

In conjunction with CFD technology, Data Center Infrastructure Management (DCiM) is used to control asset and capacity, change process, and measure and control power consumption, energy, and environment management.[5] The Energy Management system allows integrating information

[1] http://www.usgbc.org/leed/rating-systems
[2] European Commission, Directorate-General, Joint Research Centre, Institute for Energy and Transport, Renewable Energy Unit.
[3] http://www.jdcc.or.jp/english/facility.pdf
[4] http://www.jdcc.or.jp/english/council.pdf

[5] http://www.raritandcim.com/

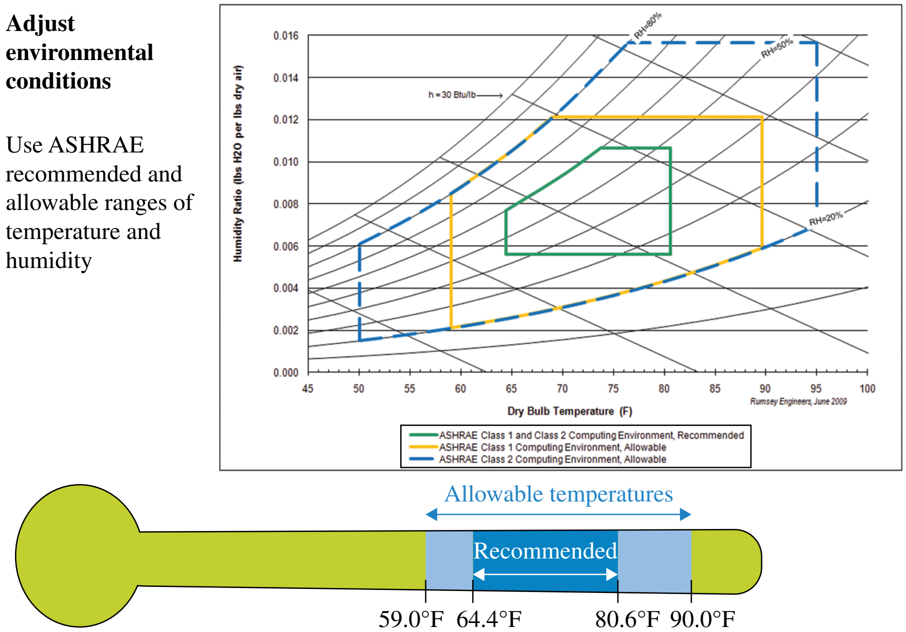


FIGURE 1.5 Adjust environmental conditions (FEMP First Thursday Seminars, U.S. Department of Energy).

such as from the Building Management System (BMS), utility meters, and UPS into actionable reports, such as accurate asset inventory, space/power/cooling capacity, and bill-back reports. A real-time dashboard display allows continuous monitoring of energy consumption and to take corrective actions.

Professors Robert Kaplan and David Norton once said: "If you can't measure it, you can't manage it." Power Usage Effectiveness (PUE™), among other accepted paradigms developed by the Green Grid, is a recognized metrics for monitoring and thus controlling your data center energy efficiency.

Incorporating both CFD and DCiM early on during design stage is imperative for successful design and ongoing data center operations. It will be extreme costly to install monitoring devices after construction of a data center.

1.5 BEST PRACTICES AND EMERGING TECHNOLOGIES

Although designing energy-efficient data centers is still evolving, many best practices could be applied whether you are designing a small server room or a large data center. The European Commission published a comprehensive "2013 Best Practices for the EU Code of Conduct on Data Centres." The U.S. Department of Energy's Federal Energy Management Program published "Best Practices Guide for Energy-Efficient Data Center Design." Both, and many other publications, could be referred to when preparing a data center design specification. Here is a short list of best practices and emerging technologies:

- Increase server inlet temperature (Fig. 1.5) and humidity adjustments [20]
- Hot- and cold-aisle configuration
- Hot and cold air containments
- Air management (to avoid bypass, hot and cold air mixing, and recirculation)
- Free cooling using air-side economizer or water-side economizer
- High-efficiency UPS
- Variable speed drives
- Rack-level direct liquid cooling
- Combined heat and power (CHP) in data centers (Fig. 1.6) [21]
- Fuel cell technology [22]
- Direct current power distribution

1.6 OPERATIONS MANAGEMENT AND DISASTER MANAGEMENT

Some of the best practices in operations management include applying ISO standards, air management, cable management, preventive and predictive maintenance, 5S, disaster management, and training.

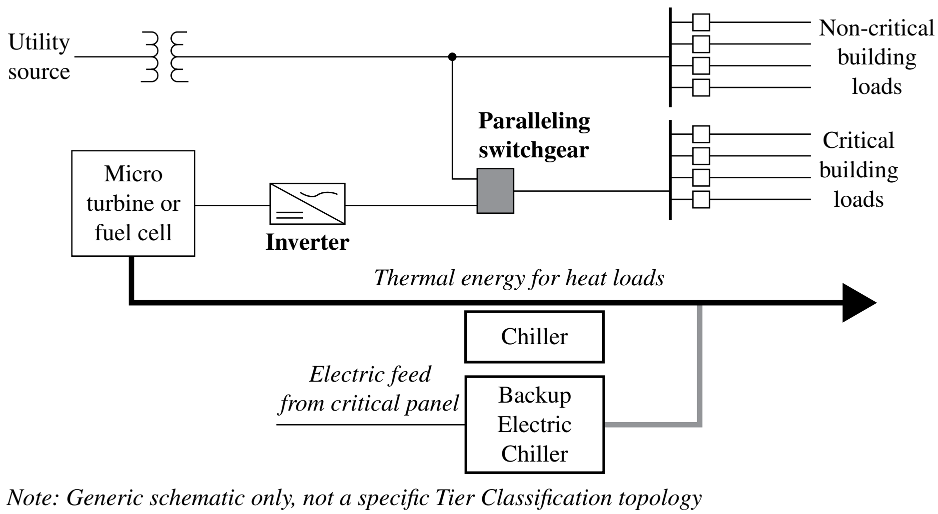


FIGURE 1.6 CHP System Layout for Data Center.

1.6.1 ISO Standards

To better manage your data centers, operations management adheres to international standards, so to "practice what you preach." Applicable ISO standards include the following:

- ISO 9000: Quality management
- ISO 14000: Environmental management
- OHSAS 18001: Occupation Health and Safety Management Standards
- ISO 26000: Social responsibility
- ISO 27001: Information security management
- ISO 50001: Energy management
- ISO 20121: Sustainable events

1.6.2 Computerized Maintenance Management Systems

Redundancy alone will not prevent failure and preserve reliability. Computerized maintenance management system (CMMS) is a proven tool, enhanced with mobile, QR/barcoding, or voice recognition capabilities, mainly used for managing and upkeeping data center facility equipment, scheduling maintenance work orders, controlling inventory, and purchasing service parts. ICT asset could be managed by DCiM as well as Enterprise Asset Management System. CMMS can be expanded and interfaced with DCiM, BMS, or Supervisory Control and Data Acquisition (SCADA) to monitor and improve Mean Time between Failure and Mean Time to Failure, both closely relating to dependability or reliability of a data center. Generally, CMMS encompasses the following modules:

- Asset management (Mechanical, Electrical, and Plumbing equipment)
- Equipment life cycle and cost management
- Spare parts inventory management
- Work order scheduling (man, machine, materials, method, and tools):
  - Preventive Maintenance (e.g., based on historical data and meter reading)
  - Predictive Maintenance (based on noise, vibration, temperature, particle count, pressure, and airflow)
  - Unplanned or emergency services
- Depository for Operations and Maintenance manual and maintenance/repair history

CMMS can earn points in LEED certification through preventive maintenance that oversees HVAC system more closely.

1.6.3 Cable Management

Cabling system may seem to be of little importance, but it makes a big impact and is long lasting, costly, and difficult to replace [23]. It should be planned, structured, and installed per network topology and cable distribution requirements as specified in TIA-942-A and ANSI/TIA/EIA-568 standards. The cable should be organized so that the connections are traceable for code compliance and other regulatory requirements. Poor cable management [24] could create electromagnetic interference due to the induction between cable and equipment electrical cables. To improve maintenance and serviceability, cabling should be placed in such a way that it could be disconnected to reach a piece of equipment for adjustments or changes. Pulling, stretching, or bending the radii of cables beyond specified ranges should be avoided. Ensure cable management "discipline" to avoid "out of control, leading to chaos [24]" in data centers.

1.6.4 The 5S Pillars [25]

5S is a lean method that organizations implement to optimize productivity through maintaining an orderly workplace.[6] 5S is a cyclical methodology including the following:

- Sort: eliminate unnecessary items from the workplace.
- Set in order: create a workplace so that items are easy to find and put away.
- Shine: thoroughly clean the work area.
- Standardize: create a consistent approach with which tasks and procedures are done.
- Sustain: make a habit to maintain the procedure.

1.6.5 Training and Certification

Planning and training play a vital role in energy-efficient design and the effective operation of data centers. The U.S. Department of Energy offers many useful training and tools.

The Federal Energy Management Program offers free interactive online "First Thursday Semin@rs" and "eTraining.[7]" Data center owners can use Data Center Energy Profiler (DC Pro) Software[8] to profile, evaluate, and identify potential areas for energy efficiency improvements. Data Center Energy Practitioner (DCEP) Program [26] offers data center practitioners with different certification programs.

1.7 BUSINESS CONTINUITY AND DISASTER RECOVERY

In addition to natural disasters, terrorist attack to the Internet's physical infrastructure is vulnerable and could be devastating. Also, statistics show that over 70% of all data centers was brought down by human errors such as improper executing procedures or maintenance. It is imperative to have detailed business continuity (BC) and disaster recovery (DR) plans well prepared and executed. BC at data centers should consider design beyond requirements per building codes and standards. The International Building Code (IBC) and other codes generally concern about life safety of occupants but with little regard to property or functional losses. To sustain data center operations after a natural disaster, the design of data center building structural and nonstructural components (mechanical equipment [27], electrical equipment [28], duct and pipe [29]) must be toughened considering BC.

Many lessons were learned on DR from two natural disasters: the Great East Japan Tsunami (March 2011) [30] and the eastern U.S. Superstorm *Sandy* (October 2012). "Many of Japan's data centers—apart from the rolling brownouts—were well prepared for the Japanese tsunami and earthquake. Being constructed in a zone known for high levels of seismic activity, most already had strong measures in place [31]."

Key lessons learned from the aforementioned natural disasters are highlighted as follows:

- Detailed crisis management procedure and communication command line.
- Conduct drills regularly by emergency response team using established procedures.
- Regularly maintain and test run standby generators and critical infrastructure in a data center.
- Have contract with multiple diesel oil suppliers to ensure diesel fuel deliveries.
- Fly in staff from nonaffected offices. Stock up food, drinking water, sleeping bags, etc.
- Have different communication mechanisms such as social networking, web, and satellite phones.
- Get required equipment on-site readily accessible (e.g., flashlight, portable generator, fuel and containers, hoses, and extension cords).
- Brace for the worst—preplan with your customers on communication during disaster and a controlled shutdown and DR plan.

Other lessons learned include using combined diesel fuel and natural gas generator, fuel cell technology, and submersed fuel pump and that "a cloud computing-like environment can be very useful [32]." "Too many risk response manuals will serve as a 'tranquilizer' for the organization. Instead, implement a risk management framework that can serve you well in preparing and responding to a disaster."

1.8 CONCLUSION

This chapter describes energy use that accelerates global warming and results in climate changes, flood, drought, and food shortage. Strategic planning of data centers applying best practices in design and operations was introduced. Rapidly increasing electricity demand by data centers for information processing and mobile communications outpaces improvements in energy efficiency. Lessons learned from natural disasters were addressed. Training plays a vital role in successful energy-efficient design and safe operations. By collective effort, we can apply best practices to radically accelerate speed of innovation (Fig. 1.7) to plan, design, build, and operate data centers efficiently and sustainably.

[6]"Lean Thinking and Methods," the U.S. Environmental Protection Agency.
[7]http://apps1.eere.energy.gov/femp/training/first_thursday_seminars.cfm
[8]http://www1.eere.energy.gov/manufacturing/datacenters/software.html

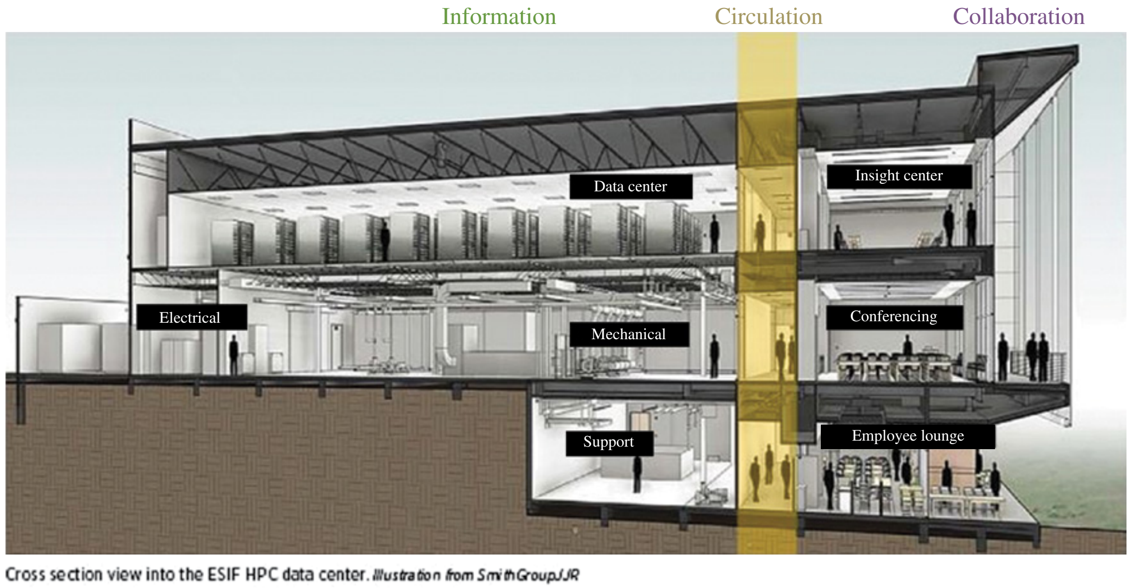


FIGURE 1.7 ESIF's high-performance computing data center—innovative cooling design with PUE at 1.06 [33].

REFERENCES

[1] IPCC. Summary for policymakers. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group l to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge/New York: Cambridge University Press; 2013.

[2] Turn down the heat: why a 4°C warmer world must be avoid. Washington, DC: The World Bank; November 18, 2012.

[3] Koomey J. Growth in data center electricity use 2005 to 2010. Analytics Press; August 2011.

[4] EMC/IDC (International Data Corporation). Extracting value from chaos; June 2011.

[5] Gartner. The top 10 strategic technology trends for 2012; November 2011.

[6] Ebbers M, Archibald M, Fonseca C, Griffel M, Para V, Searcy M. *Smart Data Centers, Achieving Greater Efficiency*. 2nd ed. IBM Redpaper; 2011.

[7] Best Practices Guide for Energy-Efficient Data Center Design. Federal Energy Management Program, U.S. Department of Energy; March 2011.

[8] Mell P, Grance T. The NIST definition of cloud computing. NIST, U.S. Department of Commerce; September 2011.

[9] Sigmon D. What is the difference between SDDC and cloud?. InFocus; August 2013.

[10] VMware. Delivering on the promise of the software-defined data center; 2013.

[11] Vision and roadmap: routing telecom and data centers toward efficient energy use. Sponsored by: Emerson Network Power, Silicon Valley Leadership Group, TIA, Yahoo Inc., U.S. Department of Energy; May 13, 2009.

[12] Porter ME. How competitive forces shape strategy. Harvard Bus Rev 1980;57(2):137–145.

[13] Geng H. Strategic planning process. Amica Association; 2012.

[14] Geng H. Data centers plan, design, construction and operations. Datacenter Dynamics Conference, Shanghai; September 2013.

[15] Bell MA. Use best practices to design data center facilities. Gartner Publication; April 22, 2005.

[16] Top 10 business practices for environmentally sustainable data centers. Microsoft; August 2012.

[17] Dunlap K, Rasmussen N. Choosing between room, row, and rack-based cooling for data centers. White Paper 130, Rev. 2, Schneider Electric Corporation.

[18] Rasmussen N, Torell W. Data center projects: establishing a floor plan. APC White Paper #144; 2007.

[19] Telecommunications Infrastructure Standard for Data Centers. Telecommunications Industry Association; August 2012.

[20] Server Inlet Temperature and Humidity Adjustments. Available at http://www.energystar.gov/index.cfm?c=power_mgt.datacenter_efficiency_inlet_temp. Accessed on May 6, 2014.

[21] Darrow K, Hedman B. Opportunities for combined heat and power in data centers. Arlington: ICF International, Oak Ridge National Laboratory; March 2009.

[22] 2010 hydrogen and fuel cell global commercialization & development update. International Partnership for Hydrogen and Fuel Cells in the Economy; November 2010.

[23] Best Practices Guides: Cabling the Data Center. Brocade; 2007.

[24] Apply proper cable management in IT racks—a guide for planning, deployment and growth. Emerson Network Power; 2012.

[25] Productivity Press Development Team. *5S for Operators: 5 Pillars of the Visual Workplace*. Portland: Productivity Press; 1996.

[26] DCEP program energy training-assessment process manual. LBNL and ANCIS Inc.; 2010.

[27] Installing seismic restraints for mechanical equipment. FEMA; December 2002.

[28] Installing seismic restraints for electrical equipment. FEMA; January 2004.

[29] Installing seismic restraints for duct and pipe. FEMA; January 2004.

[30] Yamanaka A, Kishimoto Z. The realities of disaster recovery: how the Japan Data Center Council is successfully operating in the aftermath of the earthquake. JDCC, Alta Terra Research; June 2011.

[31] Jones P. The after effect of the Japanese earthquake. Tokyo: Datacenter Dynamics; December 2012.

[32] Kajimoto M. One year later: lessons learned from the Japanese tsunami. ISACA; March 2012.

[33] High performance computing data center. National Renewable Energy Laboratory, U.S. Department of Energy; August 2012.

FURTHER READING

2011 Thermal Guidelines for Data Processing Environments. ASHRAE TC 9.9; 2011.

Al Gillen, et al., The software-defined datacenter: what it means to the CIO. IDC; July 2012.

A New Approach to Industrialized IT. HP Flexible Data Center; November 2012.

Annual Energy Outlook 2013 with Projections to 2040. U.S. Energy Information Administration; April 2013.

Avelar V, Azevedo D, French A. PUE™: a comprehensive examination of the metric. The Green Grid; 2012.

Brey T, Lembke P, et al. Case study: the ROI of cooling system energy efficiency upgrades. White Pager #39, the Green Grid; 2011.

Create a Project Plan-Cops, U.S. Dept. of Justice. Available at http://www.itl.nist.gov/div898/handbook/index.htm. Accessed on May 6, 2014.

Coles H, Han T, Price P, Gadgil A, Tschudi W. Assessing corrosion risk from outside-air cooling: are coupons a useful indicator? LBNL; March 2011.

eBay Data Center Retrofits: The Costs and Benefits of Ultrasonic Humidification and Variable Speed Drive. Energy Star Program, the U.S. EPA and DOE; March 2012. Available at http://www.energystar.gov/ia/products/power_mgt/downloads/Energy_Star_fact_sheet.pdf?0efd-83df. Accessed on May 6, 2014.

EU energy trends to 2030. European Commission; 2009.

European Code of Conduct on Data Centre Energy Efficiency, introductory guide for applicants 2013. European Commission; 2013.

Gartner IT Glossary. 2012. Available at http://www.gartner.com/itglossary/data-center/. Accessed on May 6, 2014.

Geary J. Who protects the internet? Popular Science; March 2009.

Gens F. Top 10 predictions, IDC predictions 2013: competing on the 3rd platform. IDC; November 2012.

Govindan S, Wang D, Chen L, Sivasubramaniam A, Urgaonkar B. Modeling and Analysis of Availability of Datacenter Power Infrastructure. Department of Computer Science and Engineering, The Pennsylvania State University, IBM Research Zurich. Technical Report. CSE 10-006.

Green Google. Available at http://www.google.com/green/. Accessed on May 6, 2014.

Hickins M. Companies test backup plans, and learn some lessons. Wall Street Journal, October 37, 2012.

Iyengar M, Schmidt RR. *Energy Consumption of Information Technology Data Centers*. Electronics Cooling Magazine, Publisher: ITEM Media, Plymouth Meeting, PA; December 2010.

Joshi Y, Kumar P. *Energy Efficient Thermal Management of Data Centers*. New York: Springer; 2012.

LaFontaine WR. Global technology outlook 2013. IBM Research; April 2013.

Newcombe L, et al. 2013 Best Practices for the EU Code of Conduct on Data Centers. European Commission; 2013.

Pitt Turner IV W, Brill K. Cost model: dollars per kW plus dollars per square foot of computer floor. White Paper, Uptime Institute; 2008.

Planning guide: getting started with big data. Intel; 2013.

Polar ice melt is accelerating. The Wall Street Journal, November 30, 2012.

Porter M. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press, Harvard University; 1980.

Prescription for server room growth: Design a Scalable Modular Data Center. IBM Global Services; August 2009.

Report to Congress on server and data center energy efficiency. U.S. Environmental Protection, Agency Energy Star Program; August 2007.

Rodgers TL. Critical facilities: reliability and availability. Facility Net; August 2013.

Salim M, Tozer R. Data Center Air Management Metrics-Practical Approach. Hewlett-Packard, EYP MCF.

Sawyer R. Calculating total power requirements for data centers. APC; 2004.

Trick MA. *Network Optimization*. Carnegie Mellon University; 1996.

U.S. Energy Information Administration. Available at http://www.eia.gov/todayinenergy/. Accessed on May 6, 2014.

VanDenBerg S. Cable pathways: a data center design guide and best practices. Data Center Knowledge, Industry Perspectives; October 2013.

Wenning T, MacDonald M. High performance computing data center metering protocol. Federal Energy Management Program, U.S. Department of Energy; 2010.

2

ENERGY AND SUSTAINABILITY IN DATA CENTERS

William J. Kosik
Hewlett-Packard Company, Chicago, IL, USA

2.1 INTRODUCTION

Flashback to 1999: *Forbes* published a seminal article coauthored by Peter Huber and Mark Mills. It had a wonderful tongue-in-cheek title: "Dig More Coal—the PCs Are Coming." The premise of the article was to challenge the idea that the Internet would actually reduce overall energy use in the United States, especially in sectors such as transportation, banking, and health care where electronic data storage, retrieval, and transaction processing were becoming integral to business operations. The opening paragraph, somewhat prophetic, reads as follows:

> Southern California Edison, meet Amazon.com. Somewhere in America, a lump of coal is burned every time a book is ordered on-line. The current fuel-economy rating: about 1 pound of coal to create, package, store and move 2 megabytes of data. The digital age, it turns out, is very energy-intensive. The Internet may someday save us bricks, mortar and catalog paper, but it is burning up an awful lot of fossil fuel in the process.

What Mills was trying to demonstrate is that even if you never have to drive to your bank to deposit a paycheck, or require delivery trucks to bring CDs to your house to acquire new music, a great deal of electricity is still being used by the *server* that processed your transaction or the *storage and networking gear* that is delivering your streaming media. While I am not going to detail out a life-cycle assessment counting kWh, carbon, or water, comparing the "old way" to the "new way," one thing is for sure: the Internet has created new services that do not replace anything at all, but are *completely new paradigms*. The energy use we are talking about here is completely *additive*.

Flash forward to now: One of these paradigms that come to mind is social networking. So if Mills and Huber wrote an article today, it would have to relate to how much coal is used to Tweet, friend someone in Facebook, or network with a professional group using LinkedIn. The good news here is that there are concerted efforts underway for some time by the data center industry to continue to look for ways to minimize the electricity required to power servers, storage, and networking gear, as well as to reduce the "overhead" energy used in cooling processes and power distribution systems. For example, data center owners and end users are demanding better server efficiency, airflow optimization, and using detailed building performance simulation techniques comparing "before and after" energy usage to justify higher initial spending to reduce ongoing operational costs.

The primary purpose of this chapter is to provide information and guidance on the drivers of energy use in data centers. It is a complex topic—the variables and skillsets involved in the optimization of energy use and minimization of environmental impacts are cross-disciplinary and include IT professionals, power and cooling engineers, builders, architects, finance and accounting professionals, and energy procurement teams. While these types of multidisciplinary teams are not unusual when tackling large, complex business challenges, planning, designing, and operating a new data center building is very intricate and requires a lot of care and attention. In addition, a data center has to run 8760 h/year nonstop including all scheduled maintenance, unscheduled breakdowns, and ensure that ultracritical business outcomes are delivered on time as promised. In summary, planning, design, implementation, and operations of a data center takes a considerable amount of effort and attention to detail. And after the data center is built

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

and operating, the energy cost of running the facility, if not optimized during the planning and design phases, will provide a legacy of inefficient operation and high electricity costs.

So to keep it simple, this chapter will provide some good information, tips, and resources for further reading that will help obviate (metaphorically) having to replace the engine on your car simply to reduce energy expenditures when the proper engine could have been installed in the first place. The good news is the industry as a whole is far more knowledgeable and interested in developing highly energy-efficient data centers (at least compared to a decade ago). With this said, how many more new paradigms that we haven't even thought of are going to surface in the next decade that could potentially eclipse all of the energy savings that we have achieved in the current decade? Only time will tell, but it is clear to me that we need to continue to push hard for nonstop innovation, or as another one of my favorite authors, Tom Peters, puts it, "Unless you walk out into the unknown, the odds of making a profound difference…are pretty low." So as the need for data centers continues to grow, each consuming as much electricity and water as a small town, it is imperative that we make this profound difference.

2.1.1 How Green Is Green?

I frequently get questions like, "Is going green the right thing to do, environmentally speaking? Or is it just an expensive trend? Or is there a business case for doing so (immediate energy savings, future energy savings, increased productivity, better disaster preparation, etc.)?" First, it is certainly the right thing to do. However, each individual, small business, or corporation will have different tolerance levels on the amount of "collateral goodness" they want to spread around. CIOs have shareholders and a board of directors to answer to so there must be a compelling business case for any green initiative. This is where the term *sustainable* can really be applied—sustainable from an environmental perspective but also from a business perspective. And the business perspective could include tactical upgrades to optimize energy use or it could include increasing market share by taking an aggressive stance on minimizing the impact on the environment—and letting the world know about it.

Certainly there are different shades of green here that need to be considered. When looking at specific greening activities for a data center for example, there is typically low-hanging fruit related to the power and cooling systems that will have paybacks (construction costs compared to reduced operational costs attributable to energy use) of 1 or 2 years. Some have very short paybacks because there are little or no capital costs involved. Examples of these are adjusting set points for temperature and humidity, minimizing raised floor leakage, optimizing control and sequencing of cooling equipment, and optimizing air management on the raised floor to eliminate hot spots (which reduces the need to subcool the air). Other upgrades, which are much more substantial in first costs, historically have shown paybacks closer to 5 years. These are upgrades that are done to not only increase energy efficiency but also to lengthen the life of the facility and increase reliability. So the first cost is attributable to things other than pure energy efficiency upgrades. These types of upgrades typically include replacement of central cooling plant components (chillers, pumps, cooling towers) as well as electrical distribution (UPS, power distribution units). These are typically more invasive and will require shutdowns unless the facility has been designed for concurrent operation during maintenance and upgrades. A thorough analysis, including first cost, energy cost, operational costs, and greenhouse gas emissions, is the only way to really judge the viability of different projects.

So when you're ready to go green in your data center, it is critical to take a holistic approach. As an example, when judging the environmental impact of a project, it is important to look at the entire life cycle, all the way from the extraction of the raw materials to the assembly, construction, shipping, use, and recycling/disposal. This can be a very complex analysis, but even if it is looked at from a cursory standpoint, it will better inform the decision-making process. The same is true for understanding water and land use and how the people that will be a part of final product are impacted. Similarly, the IT gear should also be included in this analysis. Certainly it is not likely that servers will be replaced simply to reduce energy costs, but it is possible to combine IT equipment retirement with energy efficiency programs. The newer equipment will likely have more efficient power supplies, more robust power management techniques, resulting in overall lower energy consumption. The newer IT gear will reduce the cooling load and, depending on the data center layout, will improve airflow and reduce air management headaches. Working together, the facilities and IT organizations can certainly make an impact in reducing energy use in the data center that wouldn't be able to be achieved by either group working alone (Fig. 2.1).

2.1.2 Environmental Impact

Bear in mind that a typical enterprise data center consumes *40 times, or more*, as much energy as a similarly sized office building. This can have a major impact on a company's overall energy use, operational costs, and carbon footprint. As a further complication, not all IT and facilities leaders are in a position to adequately ensure optimal energy efficiency, given their level of sophistication, experience, and budget availability for energy efficiency programs. Within these programs, there are multiple layers that can be applied, including encouraging employees to work from home, recycling programs, rigorous building energy and water management, vehicle fleet optimization, and carbon reporting. So where is the best place to begin?

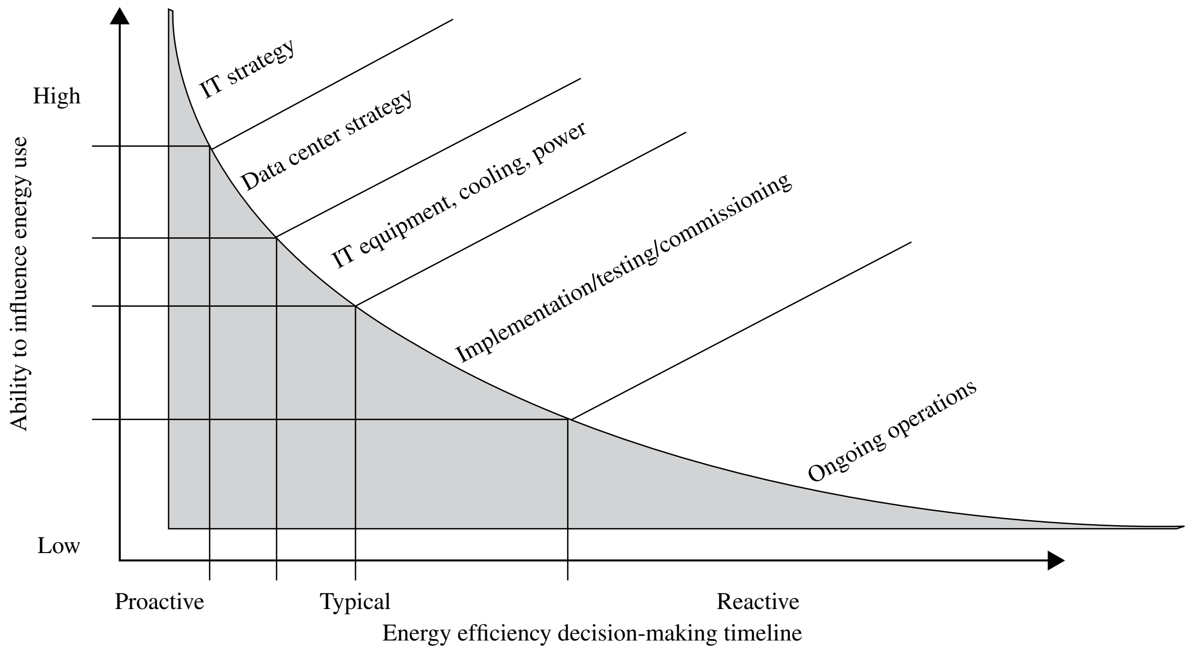


FIGURE 2.1 Data center planning timeline (HP image).

One example, Executive Order (EO) 13514, Federal Leadership in Environmental, Energy, and Economic Performance, signed by President Obama on October 5, 2009, outlines a mandate for reducing energy consumption, water use, and greenhouse gas emissions in U.S. Federal Facilities. Although the EO is written specifically for U.S. federal agencies, the broader data center industry is also entering the next era of energy and resource efficiency: strongly encouraged or compulsory reductions in resource use and greenhouse gas emissions. And while the EO presents requirements for reductions for items other than buildings (vehicles, electricity generation, etc.), the majority of the EO is geared toward the built environment. Related to data centers specifically, and the impact that technology use has on the environment, there is a dedicated section in the EO on electronics and data processing facilities. An excerpt from this section states: "… [agencies should] promote electronics stewardship, in particular by implementing best management practices for energy-efficient management of servers and Federal data centers."

Although most industry sectors in the United States continue to be immune to compulsory carbon reporting, we need to think about the future. Who knows when (or if) mandatory carbon emission reporting will start in the private sector, but the fact is that these programs are gaining momentum; and in the meantime, the CIO plays a major role in helping to hit the corporate targets, proactively getting ahead of the curve. Organizations such as the Carbon Disclosure Project (CDP) are attempting a paradigm shift in how organizations go about managing and reporting carbon emissions. (Most large corporations publish annual reports that address energy use, carbon emissions, and other sustainability goals.) No matter what program eventually becomes dominant, one carbon reporting category will have a big impact on the CIO: the amount of electricity used in powering the data center. Scope 2 Emissions, as they are known, are attributable to the generation of purchased electricity consumed by the company. And for many companies, purchased electricity represents one of the largest sources of GHG emissions (and the most significant opportunity to reduce these emissions).

When using the EO (or other greenhouse gas accounting and reporting protocols) for analyzing the carbon footprint of an data center, the entire electrical power production chain that runs from the generation plant to the building has to be considered. The utility that supplies energy in the form of electricity and natural gas will not only impact the operating cost of the facility but also drive the amount of CO_{2eq} that is released into the atmosphere. When evaluating a comprehensive energy and sustainability plan, it is critical to understand the source of energy (fossil fuel, coal, oil, natural gas, wind, solar, hydro, etc.) and the efficiency of the electricity generation to develop an all-inclusive view of how the facility impacts the environment. Each technology uses different types and amounts of fuel, and each power producer uses varying types of renewable power generation technology such as wind and solar. When sites for a new data are being evaluated, in addition to cost of electricity, the data on CO_2 emissions need be taken into consideration. The type of power generation will also dictate the amount of water used in the thermomechanical process of making electricity. Most likely, we will have no control over the efficiency of the power generation plant, but understanding key statistics on the plant will inform decisions on data center planning.

To help through this maze of issues, the EO has a very clear mandate on how to organize the thinking behind reporting and reducing CO_2 emissions by using the following framework:

- Accountability and Transparency—develop a clear strategic plan, governance, and a rating protocol
- Strategic Sustainability Performance Planning—outline goals, identify policies and procedures
- Greenhouse Gas Management—use low-emission vehicles, reduce energy use in buildings, and use on-site energy sources using renewables
- Sustainable Buildings and Communities—implement strategies for developing high-performance buildings looking at new construction, operation, and retrofits
- Water Efficiency—reduce potable water use by developing water reduction of faucets, showers, toilets, irrigation, and water use in cooling systems
- Electronic Products and Services—use Energy Star products, employ high-efficiency power supplies, non-toxic or less toxic than the alternate, and "implement best management practices for the energy-efficient management of servers and Federal data centers"
- Fleet and Transportation Management—include fleet and transportation management during greenhouse gas inventory and mitigation processes
- Pollution Prevention and Waste Reduction—minimize the generation of waste, minimize paper use, use paper made from recycled content, decrease use of chemicals, and divert 50% of nonhazardous solid waste

Clearly, some of these items are outside of the purview of the CIO, but many (directly or indirectly) impact data center planning, construction, and operations. Even if there is no formal program in place for energy and greenhouse gas reduction, this framework can be used as a starting point to outline a strategy.

All of this plays in front of a backdrop of data center electricity use that has exceeded predictions, potentially understated by as much as 70%. If one were to consider data center electricity use as the electricity use of a country, the cloud/Internet data center and telecommunications network would rank fifth in the world in electricity consumption. And, based on current projections, the future demand for electricity could more than triple to 1973 bn kWh, which is greater than the combined electricity use of France, Germany, Canada, and Brazil. Data center energy use and greenhouse gas emissions are very much in the spotlight.

2.2 FLEXIBLE FACILITIES—MODULARITY IN DATA CENTERS

As a new data center goes live, there is typically a period of time where the IT equipment is ramping up to its full potential. This will most often come in the form of empty raised floor space that is waiting for build-out with servers, networking, and storage gear. Even after the IT gear is installed, there will be time period before the utilization increases, drives up the power consumption, and intensifies heat dissipation of the IT gear well beyond minimum ratings. In some data centers, this might take a few months, and with others, even longer. In fact, most data centers contain IT equipment that, by design, will never hit 50% of its computing ability (this is done due to capacity and redundancy considerations). This exemplifies why a data center facility needs to be planned and designed in a modular fashion with malleability and the capability to react to shifts, expansions, and contractions in power use as the business needs of the organization drive the IT equipment requirements.

So what does a flexible data center look like? The needs of the end user will drive the specific type of design approach, but all approaches will have similar characteristics that will help in achieving the optimization goals of the user:

1. Container—This is typically what one might think of when discussing modular data centers. Containerized data centers were first introduced using standard 20- and 40-ft shipping containers. Newer designs now use custom-built containers with insulated walls and other features that are better suited for housing computing equipment. Since the containers will need central power and cooling systems, the containers will typically be grouped and fed from a central source. Expansion is accomplished by installing additional containers along with the required additional sources of power and cooling. Cutting-edge containerized data centers now have an "eco" option and can cool the IT equipment without using mechanical cooling. The climate and the thermal requirements of the IT equipment will drive applicability of this option.
2. Industrialized data center—This type of data center is a hybrid model of a traditional brick-and-mortar data center and the containerized data center. The data center is built in increments like the container, but the process allows for a greater degree of customization of power and cooling system choices and building layout. The modules are connected to a central spine containing "people spaces," while the power and cooling equipment is located adjacent to the data center modules. Expansion is accomplished by placing additional modules like building blocks, including the required power and cooling sources.

3. Traditional data center—Modular planning and design philosophies can also be applied to traditional brick-and-mortar facilities. However, to achieve effective modularity, tactics are required that diverge from the traditional design procedures of the past three decades. The entire shell of the building must accommodate space for future data center growth. The infrastructure area needs to be carefully planned to ensure sufficient space for future installation of power and cooling equipment. Also, the central plant will need to continue to operate and support the IT loads during expansion. If it is not desirable to expand within the confines of a live data center, another method is to leave space on the site for future expansion of a new data center module. This allows for an isolated construction process with tie-ins to the existing data center kept to a minimum.

2.2.1 Optimizing the Modular Design of Flexible Facilities

Within the context of needing additional power and cooling equipment to increase reliability, and the need to increase, decrease, or shift power for the IT equipment, applying a modularized approach can also reduce energy consumption in the data center. Using a conventional monolithic approach for power and cooling systems yields a smaller quantity of equipment that is larger in size (compared to a modular design approach). For smaller, less complex data centers, this approach is entirely acceptable. However, for large data centers with multiple data halls, possibly having different reliability requirements, a monolithic approach could produce significant difficulties in optimizing the reliability, scalability, and efficiency of the data center.

To demonstrate this idea, an analysis is done on a data center that is designed to be expanded from the day-one build of one data hall to a total of three data halls. To achieve concurrent maintainability, the power and cooling systems will be designed to an $N + 2$ topology. To optimize the system design and equipment selection, the operating efficiencies of the electrical distribution system and the chiller equipment are required. At a minimum, the operating efficiencies should be calculated at four points: 25, 50, 75, and 100% of total operating capacity. The following parameters are to be used in the analysis:

1. Electrical/UPS System—For the purposes of the analysis, a double-conversion UPS was used. The unloading curves were generated using a three parameter analysis model and capacities defined in accordance with the European Commission "Code of Conduct on Energy Efficiency and Quality of AC Uninterruptible Power Systems (UPS)." The system was analyzed at 25, 50, 75, and 100% of total IT load.
2. Chillers—Water-cooled chillers were modeled using the ASHRAE minimum energy requirements (for kW/ton) and a bi-quadratic-in- ratio-and-DT equation for modeling the compressor power consumption. The system was analyzed at 25, 50, 75, and 100% of total IT load.

2.2.2 Analysis Approach

The goal of the analysis is to build a mathematical model defining the relationship between the electrical losses at the four loading points, comparing two system types. This same approach is used to determine the chiller energy consumption. The following two system types are the basis for the analysis:

1. Monolithic design—The approach used in this design assumes 100% of the IT electrical requirements are covered by one monolithic system. Also, it is assumed that the monolithic system has the ability to modulate (power output or cooling capacity) to match the four loading points.
2. Modular design—This approach consists of providing four equal-sized units that correspond to the four loading points.

(It is important to understand that this analysis demonstrates how to go about developing a numerical relationship between energy efficiency of a monolithic versus a modular system type. There are many additional variables that will change the output and may have a significant effect on the comparison of the two system types.)

For the electrical system (Fig. 2.2a), the efficiency losses of a monolithic system were calculated at the four loading points. The resulting data points were then compared to the efficiency losses of four modular systems, each loaded to one-quarter of the IT load (mimicking how the power requirements increase over time). Using the modular system efficiency loss as the denominator, and the efficiency losses of the monolithic system as the numerator, a multiplier was developed.

For the chillers (Fig. 2.2b), the same approach is taken, with the exception of using chiller compressor power as the indicator. A monolithic chiller system was modeled at the four loading points in order to determine the peak power at each point. Then four modular chiller systems were modeled, each at one-quarter of the IT load. Using the modular system efficiency loss as the denominator, and the efficiency losses of the monolithic system as the numerator, a multiplier was developed. The electrical and chiller system multipliers can be used as an indicator during the process of optimizing energy use, expandability, first cost, and reliability.

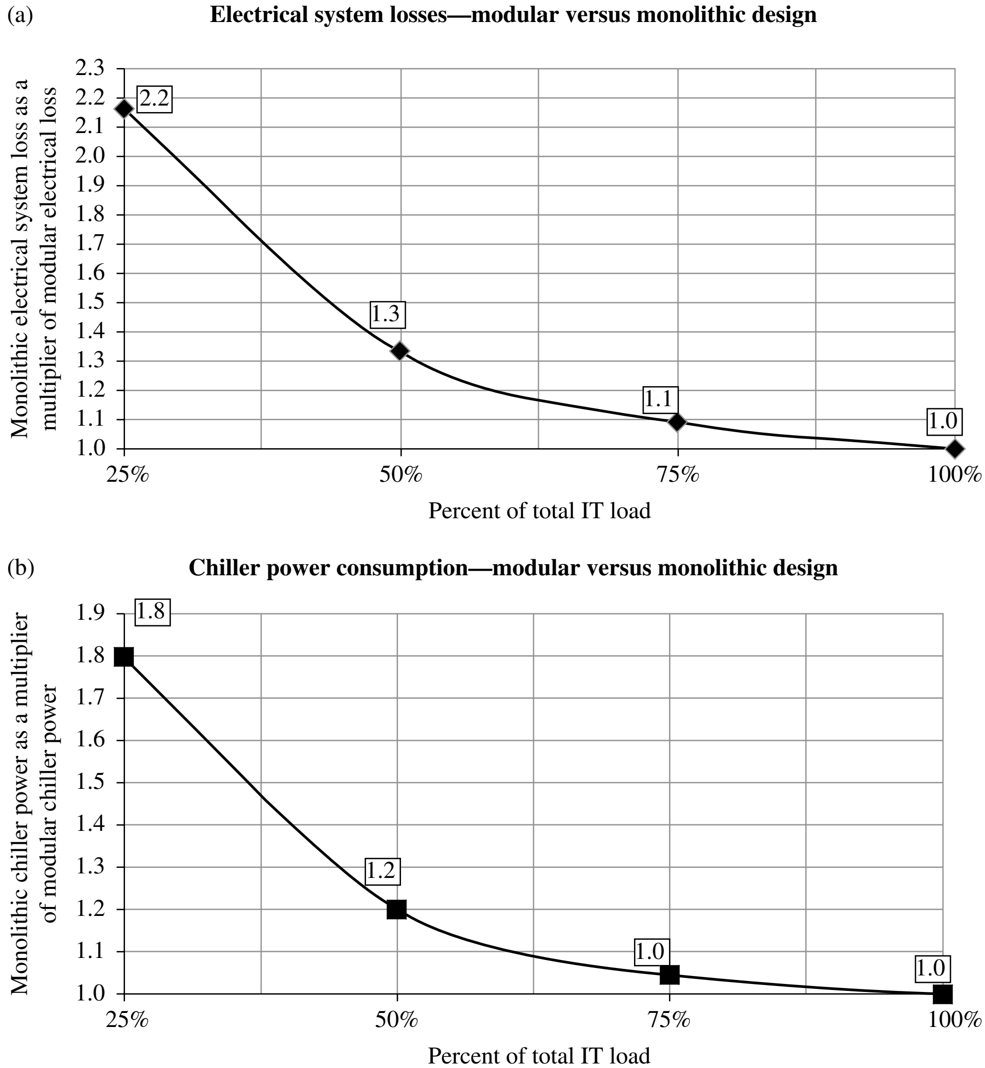


FIGURE 2.2 (a) and (b) IT load has a significant effect on electrical and cooling system efficiency.

2.2.3 Cooling a Flexible Facility

Data center users may have much tighter tolerance requirements and, for a number of reasons, need to stay in the lower temperature ranges. Analysis has shown that the normal air conditioning systems using traditional temperature and relative humidity set points will generally keep the upper end of the humidity in a reasonable range; the lower end becomes problematic, especially in mild, dry climates where there is great potential in minimizing the amount of hours that mechanical cooling is required. (When expressing moisture-level information, it is recommended to use humidity ratio or dew-point temperature since these do not change relative to the dry-bulb temperature. Relative humidity will change as the dry-bulb temperature changes.)

Energy consumption in data centers is affected by many factors such as cooling system type, UPS equipment, and IT load. Determining the impact on energy use from the climate is a nontrivial exercise requiring a considerably more granular analysis technique. Using sophisticated energy modeling tools linked with multivariate analysis techniques provides the required information for geovisualizing data center energy consumption. This is extremely useful in early concept development of a new data center giving the user powerful tools to predict approximate energy use simply by geographic siting.

As the data center design and construction industry continues to evolve and new equipment and techniques that take advantage of local climatic conditions are developed, the divergence in the PUE values will widen. It will be important to take this into consideration when assessing energy efficiency of data centers across a large geographic region so that facilities in less forgiving climates are not directly compared to facilities that are located in climates more conducive to using energy reduction strategies. Conversely, facilities that are in the cooler climate regions should be held to a higher standard in attempting to reduce annual energy consumption.

2.3 WATER USE

Water use in a data center is a very important, and a typically understated environmental challenge. The greatest amount of water consumed in a data center is not the potable water used for drinking, irrigation, cleaning, or toilet flushing; it is the cooling system, namely evaporative cooling towers and other evaporative equipment, and to a lesser extent, humidification. The water gets consumed by direct evaporation into the atmosphere, by unintended water "drift" that occurs from wind carryover, and from replacing the water used for evaporation to maintain proper cleanliness levels in the water.

In addition to the water consumption that occurs at the data center (site water use), a substantial amount of water is used at the electricity generation facility in the thermoelectrical process of making power (source water use). While discussing specific issues related to water use of electrical power plants is not in the scope of this chapter, it certainly must be understood by the data center decision-maker, especially during the site selection of a new data center. The water use of a thermal power plant is analogous to CO_2 emissions; that is, not much related to power plant efficiency can be influenced by the data center owner. Knowing that the environmental footprint of the data center will be judged by elements that extend outside of the boundaries of the data center, it is vital that decisions be made with the proper data. Different types of electrical generation processes (e.g., nuclear, coal, oil, natural gas, hydroelectric) and how the cooling water is handled (recirculated or run once through) will dictate how much water is ultimately used. For the purposes of the examples shown here, averages are used to calculate the water use in gallons/MWh. (Water use discussed in this writing refers to the water used in the operation of cooling and humidification systems only. Data come from NREL report NREL/TP-550-33905, "Consumptive Water Use for U.S. Power Production," December 2003. It is advisable to conduct analyses on potable water consumption for drinking, toilet/urinal flushing, irrigation, etc.)

For a data center that is air-cooled (DX condensing units, dry-coolers, and air-cooled chillers), water consumption is limited to humidification. For indirect economization that uses evaporative cooling, the water use will now include water that is sprayed into the outside airstream and on the heat exchanger to lower the dry-bulb temperature of the air passing through the coil. Evaporative cooling can also be used by spraying water directly into the airstream of the air-handling unit (direct evaporative cooling). If the data center has water-cooled HVAC equipment (i.e., water-cooled chillers or water-cooled computer room air conditioners), a cooling tower comes into the picture. Using an evaporation technique, the water that flows through the cooling tower is cooled down so it can be returned to the primary cooling equipment to take on the heat of the compressor. The different cooling systems can be analyzed to determine the water use for both the source and site. (The climate will have a large influence on the amount of energy consumed, and the amount of water consumed, so the following analysis is meant to be an example of what could be expected, not absolute numbers.)

It becomes apparent that in some cases even though the water use at the site increases, the water used at the source (power plant) is decreased significantly (Table 2.1).

It is not advisable to make generalized recommendations on the best way to reduce water at the site and the source, while minimizing the use of electricity—there are many variables that will contribute to the ultimate water consumption. Also the local availability of water and limitations on the amount of off-site water treatment available (the water has to go somewhere after it is used) will play into the decision and may require a less energy-efficient system in order to avoid using water on-site.

2.4 PROPER OPERATING TEMPERATURE AND HUMIDITY

The power and cooling distribution systems in a data center facility are "end of the pipe" as compared to the technology areas. The impact of the required environmental conditions (temperature and humidity) in the technology areas are non-trivial and will have a large impact on the overall energy use

TABLE 2.1 Different data center cooling systems will have different electricity and water consumption

Cooling system	Economization technique	Site/source annual HVAC energy (kWh)	Site/source annual HVAC water use (gal)
Air-cooled Dx	None	11,975,000	5,624,000
Air-cooled Dx	Indirect evaporative cooling	7,548,000	4,566,000
Air-cooled Dx	Indirect outside air	7,669,000	3,602,000
Water-cooled chillers	Water economizer	8,673,000	29,128,000
Water-cooled chillers	Direct outside air	5,532,000	2,598,000
Air-cooled chillers	Direct outside air	6,145,000	2,886,000

of the data center. Assessing the energy impact of a data center must include a close look at the thermal and power conditions in the technology area. In an existing data center, taking dry-bulb and dew-point temperature readings in multiple locations within the data center, as well as the supply and return temperatures in the air-handling system will provide the data necessary for energy analysis and subsequent recommendations.

Traditionally, most computer servers, storage devices, networking gear, etc. come with an operating manual stating environmental conditions of 20–80% noncondensing relative humidity (RH), and a recommended operation range of 40–55% RH. What is the difference between maximum and recommended? It has to do with prolonging the life of the equipment and avoiding failures due to electrostatic discharge (ESD) and corrosion failure that can come from out-of-range humidity levels in the facility. However, there is little, widely accepted data on what the projected service life reduction would be based on varying humidity levels. (ASHRAE's latest document on the subject, "2011 Thermal Guidelines for Data Processing Environments—Expanded Data Center Classes and Usage Guidance," contains very useful information related to failure rates as a function of ambient temperature, but they are meant to be used as generalized guidelines only.) In conjunction with this, the use of outside air for cooling will reduce the power consumption of the cooling system, but with outside air comes dust, dirt, and wide swings in moisture content during the course of a year. These particles can accumulate on electronic components, resulting in electrical short circuits. Also, accumulation of particulate matter can alter airflow paths inside the IT equipment and adversely affect thermal performance.

These data are necessary when looking at the first cost of the computer equipment as compared to the ongoing operating expense of operating a very tightly controlled facility. Placing computers in an environment that will certainly cause unexpected failures is not acceptable. However, if the computers are envisioned to have a 3-year in-service life and it is known that relaxing stringent internal temperature and moisture requirements will not cause a reduction in this service life, a data center owner may opt to save ongoing operating expense stemming from strict control of temperature and humidity levels. In order to use this type of approach, the interdependency of factors related to thermomechanical, electromagnetic compatibility (EMC), vibration, humidity, and temperature will need to be better understood. The rates of change of each of these factors, not just the steady-state conditions, will also have an impact on the failure mode. Finally, a majority of failures occur at "interface points" and not necessarily of a component itself. Translated, this means contact points such as soldering often cause failures. So it becomes quite the difficult task for a computer manufacturer to accurately predict distinct failure mechanisms since the computer itself is made up of many subsystems developed and tested by other manufacturers.

It is important to know that the recommended conditions of the air are at the *inlet* to the computer. There are a number of legacy data centers (and many still in design) that produce air much colder than what is required by the computers. Also, the air will most often be saturated (cooled to the same value as the dew point of the air) and will require the addition of moisture in the form of humidification in order to get it back to the required conditions. This cycle is very energy-intensive and does nothing to improve the environment conditions that the computers operate in. (In defense of data centers that house legacy IT equipment, the air is often very cold due to inadequate airflow to the IT equipment, which causes hot spots that need to be overcome with the extra-cold air.)

The results of using relative humidity as a metric in data center design can be misleading. Relative humidity (as the name implies) changes as the dry-bulb temperature of the air changes. If the upper and lower limits of temperature and relative humidity are plotted on a psychrometric chart, the dew-point temperatures range from approximately 43 to –59°F and the humidity ratios range from approximately 40 to 45–83 grains/pound. It is important to establish precise criteria on not only the temperature but also the dew-point temperature or humidity ratio of the air at the inlet of the computer. This would eliminate any confusion of what relative humidity value to use at which temperature. This may be complicated by the fact that most cooling and humidification equipment is controlled by relative humidity, and most operators have a better feel for relative humidity versus grains/pound as an operating parameter. Changes in how equipment is specified and controlled will be needed to fully use dew point or humidity ratio as a means for measurement and control.

What impact does all of this have on the operations of a data center? The main impact comes in the form of increased energy use, equipment cycling, and quite often, simultaneous cooling/dehumidification and reheating/humidification. Discharging air at 55°F from the coils in an air-handling unit is common practice in HVAC industry, especially in legacy data centers. Why? The answer is because typical room conditions for comfort cooling during the summer months are generally around 75°F and 50% RH. The dew point at these conditions is 55°F, so the air will be delivered to the conditioned space at 55°F. The air warms up (typically 20°F) due to the sensible heat load in the conditioned space and is returned to the air-handling unit. It will then be mixed with warmer, more humid outside air and then it is sent back to flow over the cooling coil. The air is then cooled and dried to a comfortable level for human occupants and supplied back to the conditioned space. While this works pretty well for office buildings, this design tactic does not transfer to data center design.

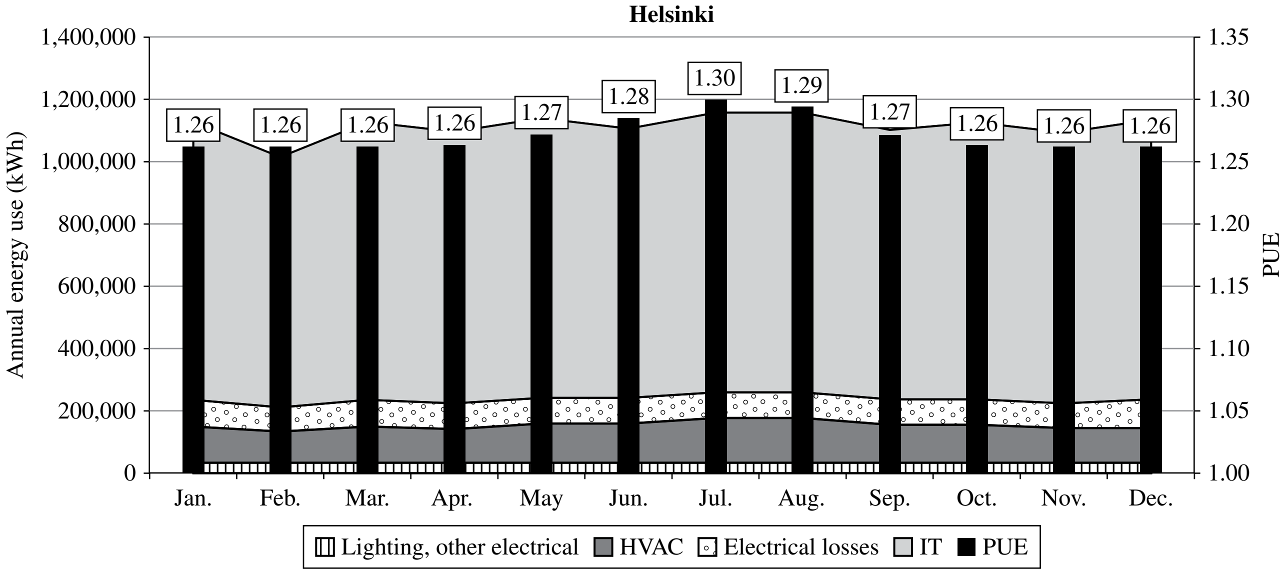


FIGURE 2.3 Monthly data center energy use and PUE for Helsinki, Finland.

Using this same process description for an efficient data center cooling application, it would be modified as follows: Since the air being supplied to the computer equipment needs to be (as an example) 78°F and 40% RH, the air being delivered to the conditioned space would be able to range from 73 to 75°F, accounting for safety margins due to unexpected mixing of air resulting from improper air management techniques. (The air temperature could be higher with strict airflow management using enclosed cold aisles or cabinets that have provisions for internal thermal management.) The air warms up (typically 20–40°F) due to the sensible heat load in the conditioned space and is returned to the air-handling unit. (Although the discharge temperature of the computer is not of concern to the computer's performance, high discharge temperatures need to be carefully analyzed to prevent thermal runaway during a loss of cooling as well as the effects of the high temperatures on the data center operators when working behind the equipment.) It will then be mixed with warmer, more humid outside air and then it is sent back to flow over the cooling coil (or there is a separate air-handling unit for supplying outside air). The air is then cooled down and returned to the conditioned space.

What is the difference in these two examples? All else being equal, the total air-conditioning load in the two examples will be the same. However, the power used by the central cooling equipment in the first case will be close to 50% greater than that of the second. This is due to the fact that much more energy is needed to produce 55°F air versus 75°F air (Fig. 2.3). Also, if higher supply air temperatures are used, the hours for using outdoor air for either air economizer or water economizer can be extended significantly. This includes the use of more humid air that would normally be below the dew point of the coil using 55°F discharge air. Similarly, if the relative humidity or humidity ratio requirements were lowered, in cool and dry climates that are ideal for using outside air for cooling, more hours of the year could be used to reduce the load on the central cooling system without having to add moisture to the airstream. Careful analysis and implementation of the temperature and humidity levels in the data center are critical to minimize energy consumption of the cooling systems.

2.5 AVOIDING COMMON PLANNING ERRORS

When constructing a new or retrofitting an existing data center facility, there is a window of opportunity at the beginning of the project to make decisions that can impact long-term energy use, either positively or negatively. Since the goal is to achieve a positive outcome, there are some very effective analysis techniques available to gain an understanding of the best optimization strategies, ensuring you're leaving a legacy of energy efficiency. In the early design phases of a data center build or upgrade, design concepts for cooling equipment and systems are not yet finalized; this is the perfect time to analyze, challenge, and refine system design requirements to minimize energy consumption attributable to cooling.

Energy is not the only criterion that will influence the final design scheme, and other conditions will affect energy usage in the data center: location, reliability level, system topology, and equipment type, among others. There is danger in being myopic when considering design alternatives. Remember cooling systems by design are dynamic and, based on the state of other systems, will continuously adjust and course-correct to maintain the proper indoor environment. Having a full understanding of the interplay that exists between seemingly unrelated factors will enable a decision-making process that is accurate and defendable. As an example, there are a number of scenarios that, if not properly analyzed and understood, could create inefficiencies, possibly significant.

- Scenario 1—Location of facility negatively impacts energy use
- Scenario 2—Cooling system mismatched with location
- Scenario 3—Data center's temperature is colder than recommended minimum
- Scenario 4—Partial loading of servers not considered in cooling system efficiency
- Scenario 5—Lack of understanding of how IT equipment energy is impacted by the cooling system

2.5.1 Scenario 1—Impacts of Climate on Energy Use

Climate use is just one of dozens of parameters that impacts energy use in the data center. Also considering the cost of electricity and type of fuel used in generating electricity, a thorough analysis will provide a much more granular view of both environmental impacts and long-term energy costs. Without this analysis, there is a risk of mismatching the cooling strategy to the local climate. It is true there are certain cooling systems that show very little sensitivity in energy use to different climates; these are primarily ones that don't use an economization cycle. The good news is that there are several cooling strategies that will perform much better in some climates than others, and there are some that perform well in many climates. A good demonstration of how climate impacts energy use comes by estimating data center energy use for the same hypothetical data center with the same power and efficiency parameters located in very different climates (Figs. 2.3 and 2.4).

2.5.2 Scenario 2—Electrical System Topology is not Considered when Establishing Preliminary PUE

Electrical system losses attributable to transformation and distribution could be equal to all of the energy consumed by the cooling system fans and pumps. This is a nontrivial number; therefore, it is vital to include the impact that the electrical system has on the PUE. Generally, the higher the reliability level, the greater the system losses simply due to the fact that UPS and PDUs will run at low loads. For facilities requiring very high availability, reliability will almost certainly trump energy efficiency—but it will come at a high cost. This is why we see the market responding with new UPS technology and innovative design solutions that enable very high efficiencies even at low loads. The overall data center PUE is affected by the type of electrical system design and the loading on the UPS system (Fig. 2.5a and b).

2.5.3 Scenario 3—Data Center Temperature is Colder than Recommended Minimum

The second law of thermodynamics tells us that heat cannot spontaneously flow from a colder area to a hotter one; work is required to achieve this. It also holds true that the colder the data center, the more the work required. So the colder the data center, the more the energy the cooling system uses to do its job. Conversely, the warmer the data center, the less the energy consumed (see caveat in the following) (Fig. 2.6). But this is

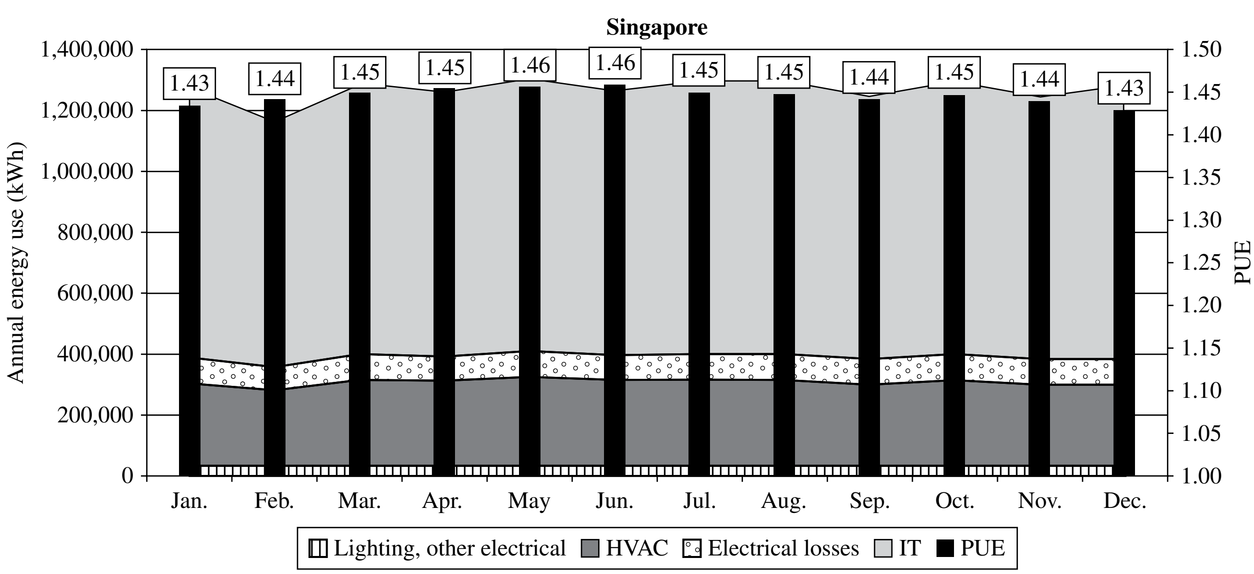


FIGURE 2.4 Monthly data center energy use and PUE for Singapore.

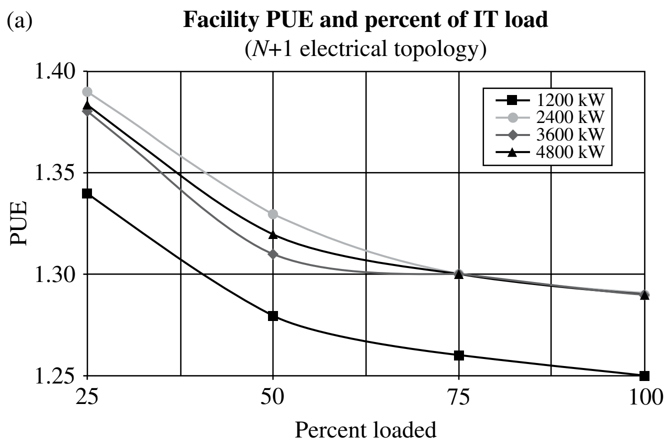


FIGURE 2.5 (a) and (b) Electrical system topology and percent of total IT load will impact overall data center PUE. In this example, a scalable electrical system starting at 1200 kW and growing to 4800 kW is analyzed. The efficiencies vary by total electrical load as well as percent of installed IT load.

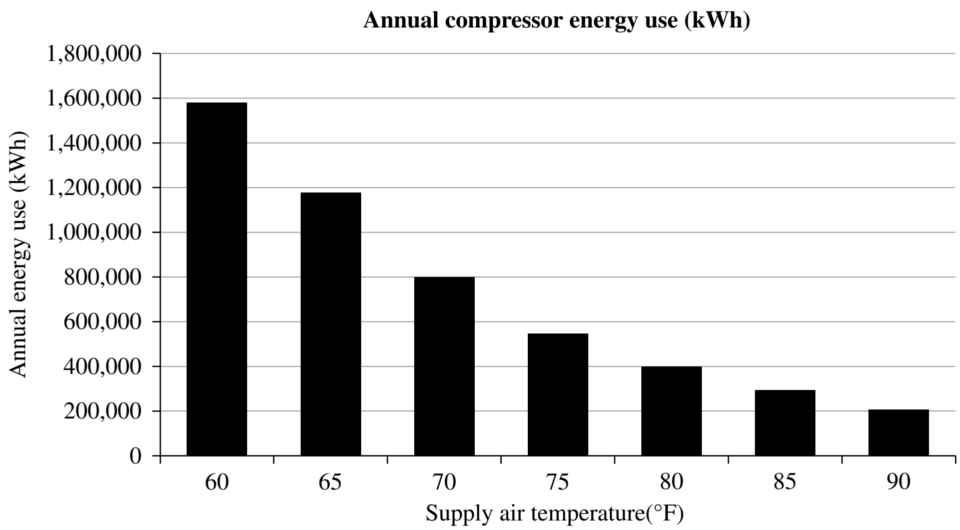


FIGURE 2.6 As supply air temperature increases, power for air conditioning compressors decreases.

just the half of it—the warmer the set point in the data center, the greater the amount of time the economizer will run. This means the energy-hungry compressorized cooling equipment will run at reduced capacity or not at all during times of economization. Now for the caveat: there will be a point when the cooling fans in the servers will consume more energy at warmer inlet temperatures, reducing or even negating any savings from reducing energy use in the cooling equipment. This will vary heavily based on the actual piece of hardware, so consult your IT vendor (and see Scenario 5).

2.5.4 Scenario 4—Partial Loading of Servers not Considered in Cooling System Efficiency

A PUE of a well-designed facility humming along at 100% load can look really great. Turn the IT load way down (simulating what happens at move-in of the facility or when the workload fluctuates) and then things suddenly don't look so good. The definition of PUE describes how efficiently a given IT load is supported by the facility's cooling and power systems. The facility will always have base-level energy consumption (people, lighting, other power, etc.) even if the IT equipment is running at very low levels. Plug these conditions into the formula for PUE and what do you get? A metrics nightmare. PUEs will easily exceed 10.0 at extremely low IT loads and will still be 5.0 or more at 10%. Not until 20–30% IT loads will the PUE start resembling a number we can be proud of. So the lesson here is to be careful when predicting PUE values and always make sure the PUE number and the time frame when that PUE can be achieved is presented (Fig. 2.7).

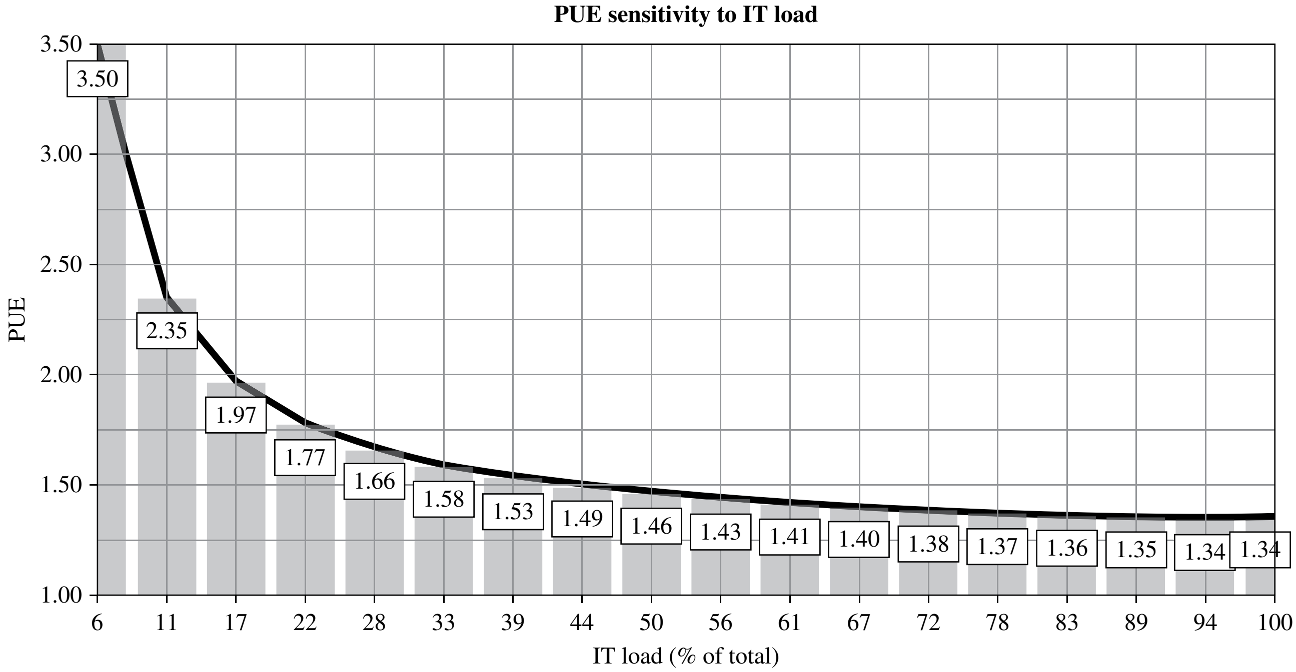


FIGURE 2.7 At very low IT loads, PUE can be very high. This is common when the facility first opens and the IT equipment is not fully installed.

2.5.5 Scenario 5—Lack of Understanding of how IT Equipment Energy is Impacted by the Cooling System

The ASHRAE TC9.9 thermal guidelines for data centers present expanded environmental criteria depending on the server class that is being considered for the data center. Since there are various types of IT servers, storage, and networking equipment manufactured by many different vendors, the details are important here. With regard to IT equipment energy use, there is a point at the lower end of the range (typically 65°F) at which the energy use of a server will level out and use the same amount of energy no matter how cold the data center temperature gets. Then there is a wide band where the temperature can fluctuate with little impact on server energy use (but a big impact on cooling system energy use—see Scenario 4). This band is typically 65–80°F, where most data centers currently operate. Above 80°F, things start to get interesting. Generally, server fan energy consumption will start to increase beyond 80°F (sometime 75°F) and will start to become a significant part of the overall IT power consumption (as compared to the server's minimum energy consumption). There comes a point where we start to see diminishing returns from the increased ambient temperature when the fan energy begins to outweigh cooling system savings. The good news is that IT equipment manufacturers are responding to this by designing servers that can tolerate higher temperatures, no longer inhibiting high temperature data center design (Fig. 2.8).

Planning, designing, building, and operating a data center requires a lot of cooperation amongst the various constituents on the project team. Data centers have lots of moving parts and pieces, both literally and figuratively. Responding to this requires a dynamic decision-making process that is fed with the best information available, so the project can continue to move forward. The key element is linking the IT and power and cooling domains so there is an ongoing dialog about optimizing not one domain or the other, but all simultaneously.

2.6 COOLING SYSTEM CONCEPTS

In a data center, the HVAC system energy consumption is dependent on three main factors: outdoor conditions (temperature and humidity), the use of economization strategies, and the primary type of cooling consider the following:

1. The HVAC energy consumption is closely related to the outdoor temperature and humidity levels. In simple terms, the HVAC equipment takes the heat from the data center and transfers it outdoors. The higher the outdoor air temperature (and the higher the humidity level is for water-cooled systems), more work is required of the compressors to lower the air temperature back down to the required levels in the data center.

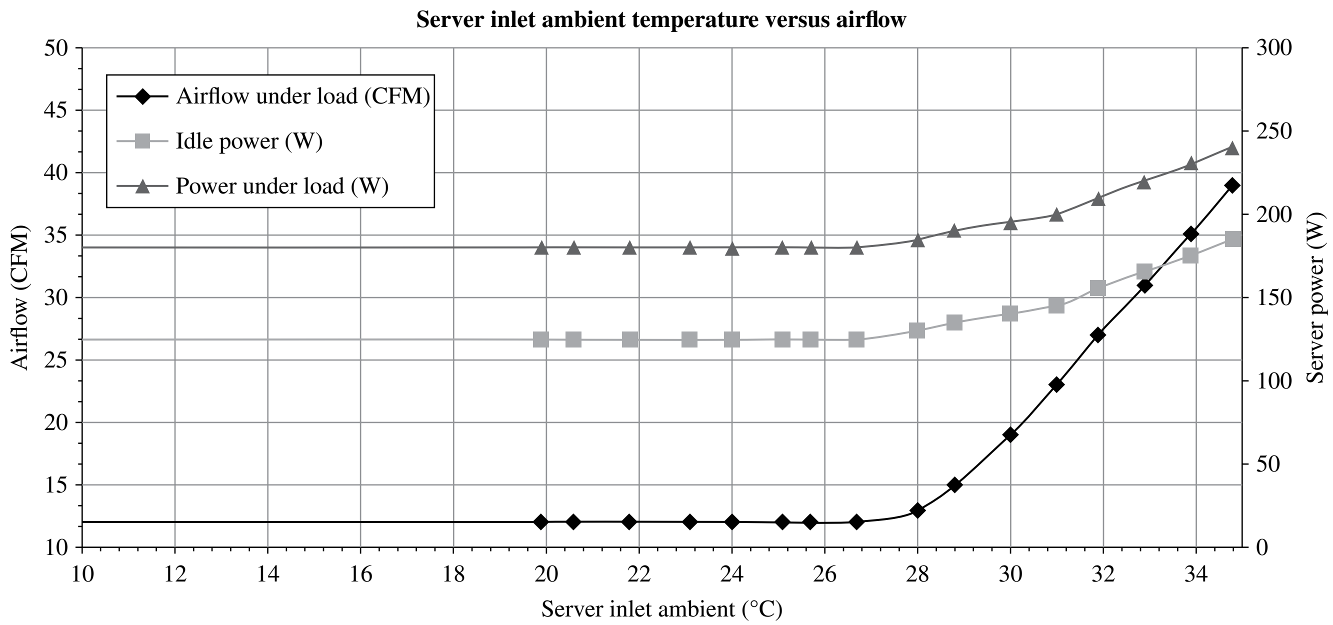


FIGURE 2.8 As server inlet temperatures increase, the overall server power will increase.

2. Economization for HVAC systems is a process in which the outdoor conditions allow for reduced compressor power (or even allowing for complete shutdown of the compressors). This is achieved by supplying outdoor air directly to the data center (direct air economizer) or, as in water-cooled systems, cooling the water and then using the cool water in place of chilled water that would normally be created using compressors.
3. Different HVAC system types have different levels of energy consumption. And the different types of systems will perform differently in different climates. As an example, in hot and dry climates water-cooled equipment generally consumes less energy than air-cooled systems. Conversely, in cooler climates that have higher moisture levels, air-cooled equipment will use less energy. The maintenance and operation of the systems will also impact energy (possibly the greatest impact). Related to the cooling system type, the supply air temperature and allowable humidity levels in the data center will have an influence on the annual energy consumption.

2.6.1 Major Cooling System Equipment Types

- Central cooling plants—Broadly speaking, cooling systems will connect to a central cooling plant that generates chilled water or condenser water for use in the remote air-handling units or CRAHs. The decision to use a central plant can be made for many different reasons: facility size, growth plans, efficiency reliability, and redundancy, among others. Generally, a central plant consists of primary equipment such as chillers and cooling towers, piping, pumps, heat exchangers, and water treatment systems. Typically, central plants are used for large data centers and have the capability for future expansion.
- Water-cooled plant equipment—Chilled water plants include chillers (either air- or water-cooled) and cooling towers (if water-cooled). These types of cooling plants are complex in design and operation but can yield superior energy efficiency. Some of the current, highly efficient water-cooled chillers offer power usage that can be 50% less than legacy models.
- Air-cooled plant equipment—Similar to the water-cooled chiller plant, the air-cooled chiller plant can be complex, yet efficient. Depending on the climate, the chiller will use more energy annually than a comparably sized water-cooled chiller. To minimize this, manufacturers offer economizer modules built into the chiller that uses the cold outside air to extract heat from the chilled water without using compressors. Dry coolers or evaporative coolers are also used to precool the return water back to the chiller.
- Direct expansion (DX) equipment—DX systems have the least amount of moving parts since both the condenser and evaporator use air as the heat transfer medium not water. This reduces the complexity but it also can reduce the efficiency. A variation on this system is to water cool the condenser, which improves the efficiency. (Water-cooled CRAC units fall into this category.)

- Evaporative Cooling Systems—Evaporative cooling uses the principle that when air is exposed to water spray, the dry-bulb temperature of the air will be reduced to a level that is close to the wet-bulb temperature of the air. The difference between the air's dry bulb and wet bulb is known as the wet-bulb depression. In climates that are dry, evaporative cooling works well, because the wet-bulb depression is large enabling the evaporative process to lower the dry-bulb temperature significantly. Evaporative cooling can be used in conjunction with any of the cooling techniques outlined earlier.
- Water Economization—Water can be used for many purposes in cooling the data center. It can be chilled via a vapor-compression cycle and sent out to the terminal cooling equipment. It can also be cooled using an atmospheric cooling tower using the same principles of evaporation and used to cool compressors or, if it is cold enough, it can be sent directly to the terminal cooling devices. The goal of a water economization strategy is to use mechanical cooling as little as possible and rely on the outdoor air conditions to cool the water to what is required to generate the required supply air temperature. When the system is in economizer mode, only air-handling unit fans, chilled water pumps, and condenser water pumps will run. The energy required to run these pieces of equipment should be examined carefully to ensure the savings of using water economizer will not be diminished by excessively high motor energy consumption. Data centers that use water-cooled servers, such as in a high-performance computing facility, can use much warmer water due to the server's ability to maintain internal temperatures using water that is at a much higher temperature than what is typically seen.
- Direct Economization—Direct economization typically means the use of outside air directly without the use of heat exchangers. Direct outside air economizer systems will mix the outdoor air with the return air to maintain the required supply air temperature. At outdoor air temperatures that range from that of the supply air temperature to that of the return air temperature, partial economization is achievable, but supplemental mechanical cooling is necessary. Evaporative cooling can be used at this point to extend the ability to use outside air by reducing the dry-bulb temperature, especially in drier climates. Once the supply air temperature can no longer be maintained, mechanical cooling will start and begin to cool the load. After the outdoor dry-bulb and moisture levels reach acceptable limits, the supplemental cooling equipment will stop and the outdoor air dampers will open to maintain the temperature. For many climates, it is possible to run direct air economization year round with little or no supplemental cooling. There are climates where the outdoor dry-bulb temperature is suitable for economization but the outdoor moisture level is too high. In this case, a control strategy must be in place to take advantage of the acceptable dry-bulb temperature without risking condensation in the data center or unintentionally incurring higher energy costs.
- Indirect Economization—Indirect economization is used when it is not advantageous to use air directly from the outdoors for economization. Indirect economization uses the same control principles as the direct outdoor air systems. In direct systems, the outdoor air is used to cool the return air by physically mixing the two airstreams. When indirect economization is used, the outdoor air is used to cool down a heat exchanger on one side that indirectly cools the return air on the other side with no contact of the two airstreams. In indirect evaporative systems, water is sprayed on a portion of the heat exchanger where the outdoor air runs through. The evaporative effect lowers the temperature of the heat exchanger, thereby reducing the temperature of the outdoor air. These systems are very effective in a number of climates, even humid climates. Since an indirect heat exchanger is used, a fan is required to draw the outside air across the heat exchanger, sometimes known as a scavenger fan. This fan motor power is not trivial and needs to be accounted for in estimating energy use.
- Economization Options—There are several different approaches and technology available when designing an economization system. For indirect economizer designs, heat exchanger technology varies widely.

    1. It can consist of a rotary heat exchanger, also known as a heat wheel, which uses thermal mass to cool down the return air by using outdoor air.
    2. Another approach is to use a cross-flow heat exchanger.
    3. Heat pipe technology can also be incorporated in an indirect economization strategy.

Within these options, there are several suboptions driven by the specific application that ultimately will inform the design strategy for the entire cooling system.

2.7 BUILDING ENVELOPE AND ENERGY USE

Buildings leak air. Sometimes, this leakage can actually produce favorable results, but most often not. No matter what, this leakage will have a significant impact on indoor temperature and humidity and must be accounted for in the design process.

Engineers who design HVAC systems for data centers generally understand that computers require an environment where temperature and humidity is maintained in accordance with some combination of the computer manufacturers' recommendations, ASHRAE guidelines, and the Telcordia NEBS requirements. Modern data center facilities will typically be designed to provide air to the inlet of the computer according to user requirements and ASHRAE guidelines. Since maintaining these temperature and humidity tolerances for 8760 h/year is very energy-intensive, much attention and research is currently aimed at HVAC system control strategies and system efficiencies to reduce energy usage. One area that is not being fully addressed is how the *building* that houses the computers affects the temperature, humidity, and energy use. In order to address what role the building plays, the following questions need to be answered:

1. Does the amount of leakage across the building envelope correlate to indoor humidity levels and energy use?
2. How does the climate where the data center is located affect the indoor temperature and humidity levels? Are certain climates more favorable for using outside air economizer without using humidification to add moisture to the air during the times of the year when outdoor air is dry?
3. Will widening the humidity tolerances required by the computers actually produce worthwhile energy savings?

2.7.1 Building Envelope Effects

The building envelope is made up of the roof, exterior walls, floors, and underground walls in contact with the earth, windows, and doors. Many data center facilities have minimal amounts of windows and doors, so the remaining elements of roof, walls, and floor are the primary elements for consideration. The elements have different parameters to be considered in the analysis: thermal resistance (insulation), thermal mass (heavy construction such as concrete versus light-weight steel), air-tightness, and moisture permeability.

When a large data center is running a full capacity, the effects of the building envelope on energy use (as a percent of the total) are relatively minimal. However, since many data center facilities never reach their full build-out potential, or if they do it is over a longer period of time, defining the requirements of the building envelope need to be an integral part of the design process.

When analyzed over time as the facility load increases, the envelope losses start out as a significant component of the overall cooling load, but decrease over time as the computer load becomes a greater portion of the load (Table 2.2).

TABLE 2.2 Example of how building envelope cooling changes as a percent of total cooling load

Percent of computer equipment running (%)	Envelope losses as a percent of total cooling requirements (%)
20	8.2
40	4.1
60	2.8
80	2.1
100	1.7

The ASHRAE Energy Standard 90.1 has very specific information on different building envelope alternatives that can be used to meet the minimum energy performance requirements. Additionally, the ASHRAE publication *Advanced Energy Design Guide for Small Office Buildings* also goes into great detail on the most effective strategies for building envelope design by climatic zone. Finally, another good source of engineering data is the CIBSE Guide A on Environmental Design.

2.7.2 Building Envelope Leakage

Building leakage will impact the internal temperature and relative humidity by outside air infiltration and moisture migration. Depending on the climate, building leakage can negatively impact both the energy use of the facility as well as the indoor moisture content of the air. But, depending on the climate, building leakage may actually *reduce* energy consumption by providing supplemental cooling. This, however, is not a recommended way of using outside air to reduce the mechanical cooling load.

Based on a number of studies from NIST, CIBSE, and ASHRAE investigating leakage in building envelope components, it is clear that oftentimes building leakage is underestimated by a significant amount. Also, there is not a consistent standard on which to base building air leakage. For example:

- CIBSE TM-23, Testing Buildings for Air Leakage and the Air Tightness Testing and Measurement Association (ATTMA) TS1 recommend building air leakage rates from 0.11 to 0.33 CFM/ft^2.
- Data from Chapter 27, "Ventilation and Air Infiltration" from ASHRAE Fundamentals show rates of 0.10, 0.30, and 0.60 CFM/ft^2 for tight, average, and leaky building envelopes.
- The NIST report of over 300 existing U.S., Canadian, and U.K. buildings showed leakage rates ranging from 0.47 to 2.7 CFM/ft^2 of above-grade building envelope area.
- The ASHRAE Humidity Control Design Guide indicates typical commercial buildings have leakage rates of 0.33–2 air changes per hour and buildings constructed in the 1980s and 1990s are not significantly tighter than those constructed in the 1950s, 1960s, and 1970s.

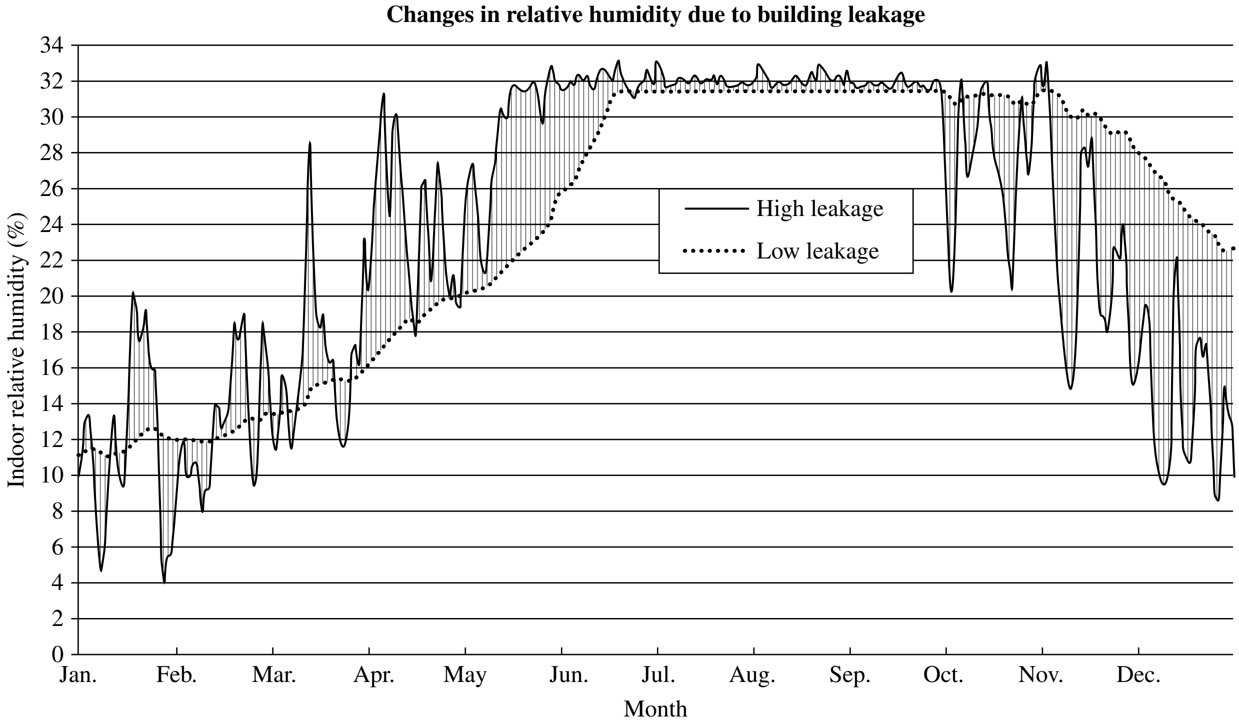


FIGURE 2.9 Internal humidity levels will correspond to outdoor moisture levels based on the amount of building leakage.

So to what extent should the design engineer be concerned about building leakage? Using hourly simulation of a data center facility and varying the parameter of envelope leakage, it is possible to develop profiles of indoor relative humidity and air change rate.

2.7.3 Using Building Performance Simulation for Estimating Energy Use

Typical analysis techniques look at peak demands or steady-state conditions that are just representative "snapshots" of data center performance. These analysis techniques, while very important for certain aspects of data center design such as equipment sizing, do not tell the engineer anything about the dynamics of indoor temperature and humidity—some of the most crucial elements of successful data center operation. However, using an hourly (and subhourly) simulation tool such as EnergyPlus (developed by the U.S. Department of Energy) will yield results that will provide the engineer rich detail to be analyzed, which can inform solutions to optimize energy use. As an example of this, using building performance simulation techniques for data center facilities yields marked differences in indoor relative humidity and air change rates when comparing different building envelope leakage rates (Fig. 2.9). Since it is not possible to develop full-scale mock-ups to test building envelope integrity, the simulation process is an invaluable tool to analyze the impact to indoor moisture content based on envelope leakage. Based on research, the following conclusions can be drawn:

- There is a high correlation between leakage rates and fluctuations in indoor relative humidity—the greater the leakage rates, the greater the fluctuations.
- There is a high correlation between leakage rates and indoor relative humidity in the winter months—the greater the leakage rates, the lower the indoor relative humidity.
- There is low correlation between leakage rates and indoor relative humidity in the summer months—the indoor relative humidity levels remain relatively unchanged even at greater leakage rates.
- There is a high correlation between building leakage rates and air change rate—the greater the leakage rates, the greater the number of air changes due to infiltration.

2.8 AIR MANAGEMENT AND CONTAINMENT STRATEGIES

Proper airflow management creates cascading efficiency through many elements in the data center. If done correctly, it will significantly reduce problems related to re-entrainment of hot air into the cold aisle, which is often the culprit of hot spots and thermal overload. Air containment will also create a microenvironment with uniform temperature gradients, enabling predictable conditions at the air inlets to the servers. These conditions ultimately allow for the use of increased server cooling air temperatures, which reduces the energy needed to cool the air. It also allows for an expanded window of operation for economizer use.

There are many remedial yet effective approaches to improve cooling effectiveness and air distribution in existing data centers. These include rearrangement of solid and perforated floor tiles, sealing openings in the raised floor, installing air dam baffles in IT cabinets to prevent air bypassing the IT gear, and other more extensive retrofits that result in pressurizing the raised floor more uniformly to ensure the air gets to where it is needed.

But arguably, the most effective air management technique is the use of physical barriers to contain the air and efficiently direct it to where it will be most effective. There are several approaches that give the end user options to choose from that meet the project requirements.

2.8.1 Passive Chimneys Mounted on IT Cabinets

These devices are the simplest and lowest cost of the options, and have no moving parts. Depending on the IT cabinet configuration, the chimney is mounted on the top and discharges into the ceiling plenum. There are specific requirements for the cabinet, and it may not be possible to retrofit on all cabinets. Also, the chimney diameter will limit the amount of airflow from the servers, so it might be problematic to install them on higher-density cabinets.

Fan-Powered Chimneys Mounted on IT Cabinets—These devices are materially the same as the passive chimneys; but as the name implies, the air movement is assisted by a fan. The fan ensures a more positive discharge into the ceiling plenum, but it is also a point of failure and increases the cost of the installation and energy use. UPS power is required if continuous operation is needed during a power failure. Though the fan-assist allows for more airflow through the chimney, it still will have limits on the amount of air that can flow through it.

2.8.2 Hot-Aisle Containment

The tried-and-true hot-aisle/cold-aisle arrangement used in laying out the IT cabinets was primarily developed to compartmentalize the hot and cold air. Certainly, it provided benefits compared to layouts where IT equipment discharged hot air right into the air inlet of adjacent equipment. (Unfortunately, this circumstance still exists in many data centers with legacy equipment.) Hot-aisle containment takes the hot-aisle/cold-aisle strategy and builds upon it substantially. The air in the hot aisle is contained using a physical barrier that can range from the installation of a curtain system that is mounted at the ceiling level and terminates at the top of the IT cabinets. Other more expensive techniques used solid walls and doors that create a hot chamber that completely contains the hot air. This system is generally more applicable for new installations. The hot air is discharged into the ceiling plenum from the contained hot aisle. Since the hot air is now concentrated into a small space, worker safety needs to be considered since the temperatures can get quite high.

2.8.3 Cold-Aisle Containment

While the cold-aisle containment may appear to be simply a reverse of the hot-aisle containment, it can tend to be much more complicated in its operation. The cold-aisle containment system can also be constructed from a curtain system or solid walls and doors. The difference between this and the hot-aisle containment comes from the ability to manage airflow to the computers in a more granular way. When constructed out of solid components, the room can act as a pressurization chamber, which will maintain the proper amount of air that is required by the servers by monitoring and adjusting the pressure. The air-handing units serving the data center are given instructions to increase or decrease air volume in order to keep the pressure in the cold aisle at a preset level. As the server fans speed up, more air is delivered; when they slow down, less is delivered. This type of containment has several benefits beyond traditional airflow management.

2.8.4 Self-Contained In-Row Cooling

To tackle air management problems on an individual level, self-contained in-row cooling units are a good solution. These come in many varieties such as chilled water-cooled, air-cooled DX, low-pressure pumped refrigerant and even CO_2-cooled. These are best applied when there is a small grouping of high-density, high heat generating servers that are creating difficulties for the balance of the data center.

2.8.5 Water-Cooled Computers

Once the staple of data centers of yore, sectors like academic and research high-performance computing continue to use water-cooled computers. This is typically not thought of as an airflow management strategy, but it is. The water cooling keeps the airflow through the computer to a minimum (the components not water-cooled still need airflow for heat dissipation). Typically, a water-cooled cabinet will reject 10–30% of the total cabinet capacity to the air—a nontrivial number when the IT cabinet houses 50–80 kW computers. Water-cooling similarly allows for uniform cabinet spacing without creating hot spots. Certainly not a mainstream tactic to be used for enhancing airflow management, it is important to be aware of the capabilities for future applicability.

2.8.6 Immersion Cooling

Immersion cooling is a technique that submerges the servers in a large container of mineral oil. The servers require some modification; but by using this type of strategy, fans are eliminated from the computers. The oil is circulated through the container around the servers and is typically pumped to heat exchanger that is tied to outdoor heat rejection equipment.

If a data center owner is considering the use of using elevated supply air temperatures, some type of containment will be necessary as the margin for error (unintentional air mixing) gets smaller as the supply air temperature increases. As the use of physical air containment becomes more practical and affordable, implementing these types of energy efficiency strategies will become more feasible.

2.9 ELECTRICAL SYSTEM EFFICIENCY

In data centers, reliability and maintainability of the electrical and cooling systems are fundamental design requirements to enable successful operation of the IT and cooling systems. It is possible to achieve the reliability goals and optimize energy efficiency at the same time, but it requires close collaboration amongst the IT and facility teams to make it happen.

The electrical distribution system in a data center encompasses numerous equipment and subsystems that begin at the utility and building transformers, switchgear, UPS, PDUs, RPPs, and power supplies, ultimately powering the fans and internal components of the IT equipment. All of these components will have a degree of inefficiency resulting in a conversion of the electricity into heat (energy loss). Some of these components have a linear response to the percent of total load they are designed to handle; others will demonstrate a very nonlinear behavior. Response to partial load conditions is important to understand when estimating overall energy consumption in a data center with varying IT loads. Also, while multiple concurrently energized power distribution paths can increase the availability (reliability) of the IT operations, this type of topology can decrease the efficiency of the overall system, especially at partial IT loads.

In order to illustrate the impacts of electrical system efficiency, it is important to understand the primary factors that have the biggest influence on overall electrical system performance:

1. UPS module and overall electrical distribution system efficiency
2. Part load efficiencies
3. System modularity
4. System topology (reliability)
5. Impact on cooling load

There are many different types of UPS technologies, each being suited for a particular use. Some perform better at lower loads, where others are used almost exclusively for very large IT loads. The final selection of the UPS technology is really dependent on the specific case. With this said, it is important to know that different UPS sizes and circuit types have different efficiency curves—it is certainly not a "one-size-fits-all" proposition. Each UPS type will perform differently at part-load conditions, so analysis at 100, 75, 50, 25, and 0% loading is necessary to gain a complete picture of UPS and electrical system efficiency (Fig. 2.10). At lower part-load values, the higher reliability systems (generally) will have higher overall electrical system losses as compared to a lower reliability system. As the percent load approaches unity, the gap narrows between the two systems. The absolute losses of the high-reliability system will be 50% greater at 25% load than the regular system, but this margin drops to 23% at 100% load. When estimating annual energy consumption of a data center, it is advisable to include a schedule for the IT load that is based on the actual operational schedule of the IT equipment, thus providing a more accurate estimate of energy consumption. This schedule would contain the predicted weekly or daily operation, including operational hours and percent loading at each hour, of the computers (based on historic workload data), but more importantly the long-term ramp-up of the power requirements for the computers. With this type of information, planning and analysis for the overall annual energy consumption will be more precise.

In addition to the UPS equipment efficiency, the modularity of the electrical system will have a large impact on the efficiency of the overall system. UPS modules are typically designed as systems, where the systems consist of multiple modules. So within the system, there could be redundant UPS modules or there might be redundancy in the systems themselves. The ultimate topology design is primarily driven by the owner's reliability, expandability, and cost requirements. The greater the number of UPS modules, the smaller the portion of the overall load that will be handled by each module. The effects of this become pronounced in high-reliability systems at low loads where it is possible to have a single UPS module working at 25% (or lower) of its rated capacity.

Ultimately, when all of the UPS modules, systems, and other electrical equipment are pieced together to create a unified electrical distribution system that is designed to meet certain reliability and availability requirements, efficiency values at the various loading percentages are developed for the entire system. The entire system now includes all power distribution upstream and downstream of the UPS equipment. In addition to the loss incurred by the UPS equipment, losses from transformers, generators, switchgear, power distribution units (with and without static transfer switches), and distribution wiring must be accounted for. When all of these

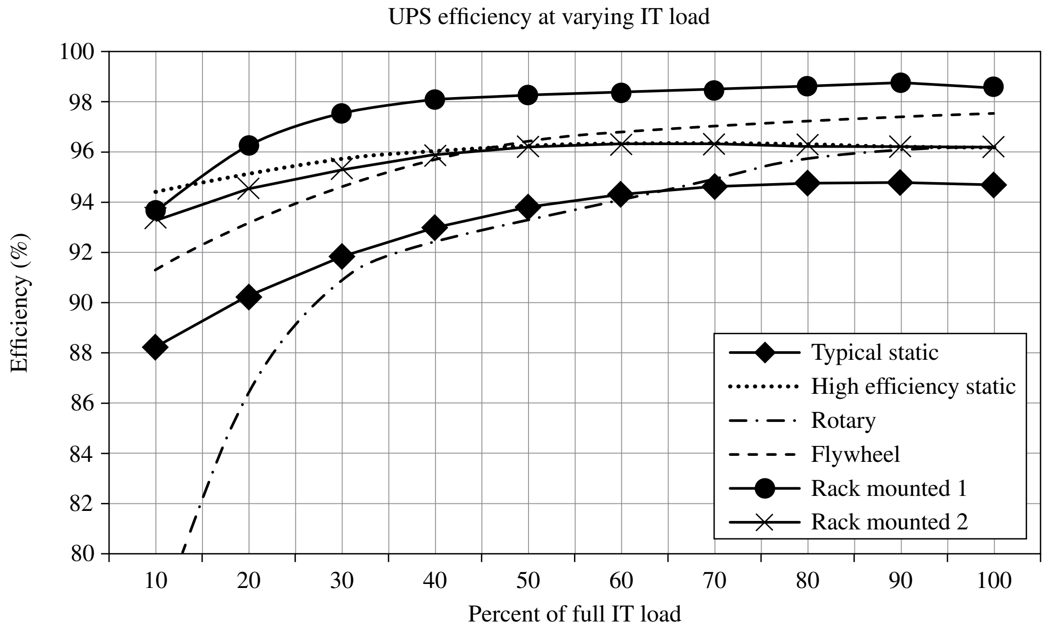


FIGURE 2.10 Example of manufacturers' data on UPS part load performance.

components are analyzed in different system topologies, loss curves can be generated so the efficiency levels can be compared to the reliability of the system, assisting in the decision-making process. There generally is an inverse correlation between the anticipated reliability level of the electrical system and the efficiency; in general, the higher the reliability, the lower the efficiency.

Ultimately, when evaluating data center energy efficiency, it is the overall energy consumption and PUE that matters. So while oftentimes the UPS system concepts are studied in isolation, this process should create an electrical system silo. Integration is key here. Early in the design process a timeline of the anticipated IT load growth needs to be developed in order to properly design the power and cooling systems, from a modular growth point of view. As mentioned earlier, the annual PUE based on this load growth can be calculated using energy modeling techniques. If modeled properly, the part-load efficiencies for the electrical system (and the cooling system) will determine the amount of energy that is ultimately used for powering the computers, and the amount dissipated as heat. Keep in mind that the UPS is just a part of the equation that is driving the PUE values high; the PUE is burdened with overhead items that are required to operate the data center (lighting, administrative space, back of the house power).

Since the losses from the electrical systems ultimately result in additional cooling load (except for equipment located outdoors or in nonconditioned spaces), the mechanical engineer will need to use this data in sizing the cooling equipment and evaluating annual energy consumption. The efficiency of the cooling equipment will determine the amount of energy required to cool the electrical losses. It is essential to include cooling system energy usage resulting from electrical losses in any life cycle studies for UPS and other electrical system components. It is possible that lower-cost, lower-efficiency UPS equipment will have a higher life cycle cost from the cooling energy required, even though the capital cost may be significantly less than a high-efficiency system. In addition to the energy that is "lost," the additional cooling load resulting from the loss will negatively impact the annual energy use and PUE for the facility. The inefficiencies of the electrical system have a twofold effect on energy consumption.

Reliability and availability in the data center are of paramount importance for the center's operator. Fortunately, in recent years, the industry has responded well with myriad new products and services to help increase energy efficiency, reduce costs, and improve reliability. When planning a new data center or considering a retrofit to an existing one, the combined effect of all of the different disciplines collaborating in the overall planning and strategy for the power, cooling, and IT systems will produce plans that will yield high levels of efficiency and reliability. And using the right kind of tools and analysis techniques are an essential part of accomplishing this.

2.10 ENERGY USE OF IT EQUIPMENT

Since the EPA's 2007 EPA Report to Congress on Server and Data Center Energy Efficiency has been released, the IT industry has responded by developing benchmarks and transparent testing methodology for IT equipment and data center power use. An important aspect of this was to develop benchmarks and reporting methodology that is manufacturer-agnostic and that provides clear and understandable data to be used as part of a decision-making process. This is an ongoing process since new equipment is frequently released requiring new testing and reporting.

2.10.1 The EPA ENERGY STAR Specification

The EPA's Enterprise Servers Specification Version 2.0 for computer servers is the definitive process for determining the power use of different server types manufactured by different vendors. The procedures in the documentation ensure a uniform approach in testing and reporting of IT equipment, including consistent specifications for ambient temperature and humidity conditions during the testing. Another key element of this process is reporting of power use at varying load points, including idle and full load. The manufacturer has to verify that server processor power management protocols, enabled by default by the BIOS, are in place to make sure that power consumption in times of low utilization is reduced using methods such as reducing voltage and/or frequency, or reducing processor or core power states when not in use. Additionally, to qualify for ENERGY STAR, the computer must have a preinstalled supervisor system that includes power management. This too must be enabled by default by the BIOS.

The importance of this testing and reporting becomes obvious when reviewing some of the initial data submitted to the EPA as a part of the ENERGY STAR program. One of the vital criteria for server energy efficiency is the power measured in watts in an idle state and in a full power mode.

2.10.2 SPECpower_ssj2008

The Standard Performance Evaluation Corporation (SPEC) has designed SPECpower_ssj2008 as both a comparison benchmark and a methodology to increase efficiency of server-class computer equipment. For the purposes of this discussion, the metric used to demonstrate efficiency is the difference between the server high-energy use state and the idle state. As the difference increases between the high and low energy use states, the server uses energy more efficiently at low loads (Fig. 2.11).

Reviewing the data we see that the ratio of the minimum to maximum power states has decreased from over 60% to just under 30%. This means that at a data center level, if all of the servers were in an idle state, in 2007 the running IT load would be 60% of the total IT load, while in 2013, it would be under 30%. This trickles down to the cooling and power systems consuming even more energy. Clearly, this is a case for employing aggressive power management strategies in existing equipment and evaluating server equipment energy efficiency when planning an IT refresh.

(The supercomputing community has developed a standardized ranking technique, since the processing ability of these types of computers is different than that of enterprise servers than run applications using greatly different amounts of processing power. The metric that is used is megaFLOPS/watt, which is obtained by running a very prescriptive test using a standardized software package (HPL). This allows for a very fair head-to-head energy efficiency comparison of different computing platforms.)

Studies have shown that the average enterprise server will typically have a utilization of 20% or less. The principal method to reduce server energy consumption starts with using more effective equipment, which uses efficient power supplies and supports more powerful processor and memory.

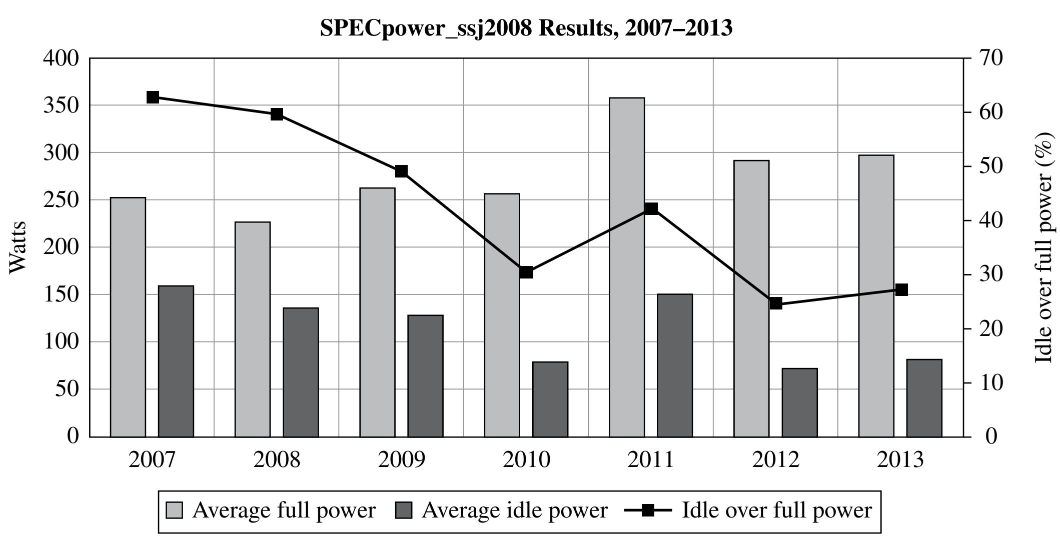


FIGURE 2.11 Since 2007, the idle over full-power ratio has trended downward, indicating an increase in server efficiency.

Second, reducing (physically or virtually) the number of servers that are required to run a given workload will reduce the overall power demand. Coupling these two approaches together with a robust power management protocol will ensure that when the servers are in operation they run as efficiently as possible.

It is important to understand the potential energy reduction from using virtualization and power management strategies. In this example, a 1000 kW data center with an average of 20% utilization was modeled with 100% of the IT load attributable to compute servers. Applying power management to 20% of the servers will result in a 10% reduction in annual energy attributable to the servers. Virtualizing the remaining servers with a 4:1 ratio will reduce the energy another 4% to a total of 14%. Increasing the utilization of the physical servers from 20 to 40% will result in a final total annual energy reduction of 26% from the base. These might be considered modest changes in utilization and virtualization; but at 10 cents/kWh, these changes would save over $130,000/year. And this is only for the electricity for the servers, not the cooling energy and electrical system losses (Table 2.3).

Average of all servers measured—average utilization = 7.9%
Busiest server measured—average utilization = 16.9%

Figuring in the cooling and power energy consumed in Scenario 1, the cooling and electrical losses will be reduced from 1,747,000 to 1,573,000, 174,000 kWh/year. Further reduction for Scenario 2 brings the total down to 1,278,000, an additional 295,000 kWh annually. For Scenario 3, the total annual energy for the power and cooling systems is further reduced to 789,000 kWh, 489,000 kWh additional reductions (Figs. 2.12, 2.13, and 2.14).

Another aspect demonstrating the interdependency between the IT equipment, and the power and cooling systems

TABLE 2.3 Analysis showing the impact on energy use from using power management, virtualization, and increased utilization

	Server energy (kWh)	Power and cooling energy (kWh)	Total annual energy consumption (kWh)	Reduction from base case (%)	Annual electricity expense reduction (based on $0.10/kWh)
Base case	5,452,000	1,747,000	7,198,000	Base	Base
Scenario 1—Power management	4,907,000	1,573,000	6,479,000	10%	$71,000
Scenario 2—Virtualization	3,987,000	1,278,000	5,265,000	27%	$121,000
Scenario 3—Increased utilization	2,464,000	789,000	3,253,000	55%	$201,000

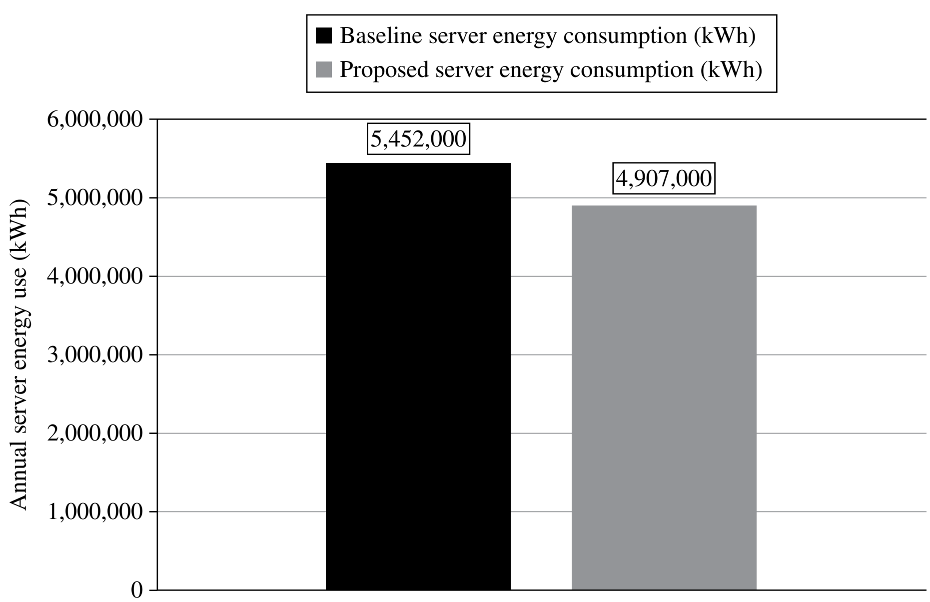


FIGURE 2.12 Energy use reduction by implementing server power management strategies.

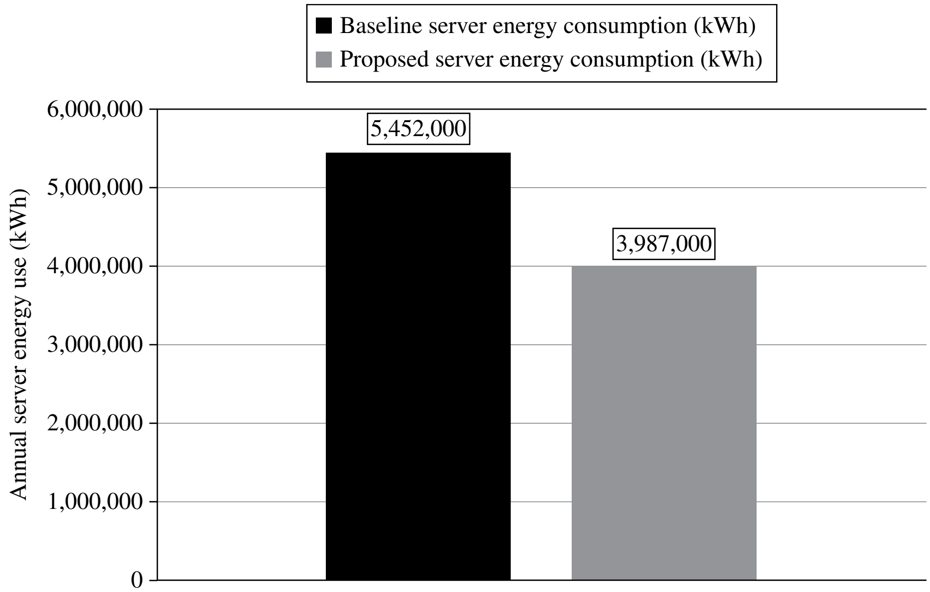


FIGURE 2.13 Energy use reduction by implementing server power management strategies and virtualizing servers.

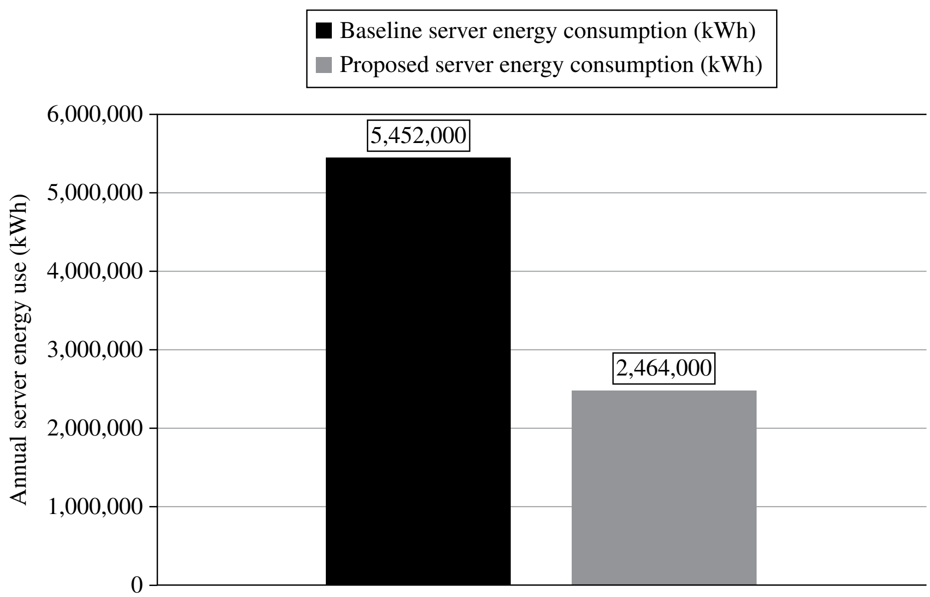


FIGURE 2.14 Energy use reduction by implementing server power management strategies, virtualizing servers, and increasing utilization.

is the temperature of the air delivered to the computers for cooling. A basic design tenet is to design for the highest internal air temperature allowable that will still safely cool the computer equipment and not cause the computers' internal fans to run at excessive speeds. The ASHRAE temperature and humidity guidelines for data centers recommend an upper dry-bulb limit of 80°F for the air used to cool the computers. If this temperature is used (and even higher temperatures in the near future), the hours for economization will be increased and when vapor compression (mechanical) cooling is used, the elevated temperatures will result in lower compressor power. However, the onboard fans in the servers will typically begin to increase speed and draw in more air to maintain the temperature of the server's internal components at an acceptable level (Fig. 2.15). The good news here is that server manufacturers are designing new servers to handle elevated temperatures without increasing airflow and server fan power. It is important to balance the additional energy from the server fans with the increase in efficiency for the cooling system. The climate in which the data center is located will drive the number of hours that are useful for using the air economizer, which will

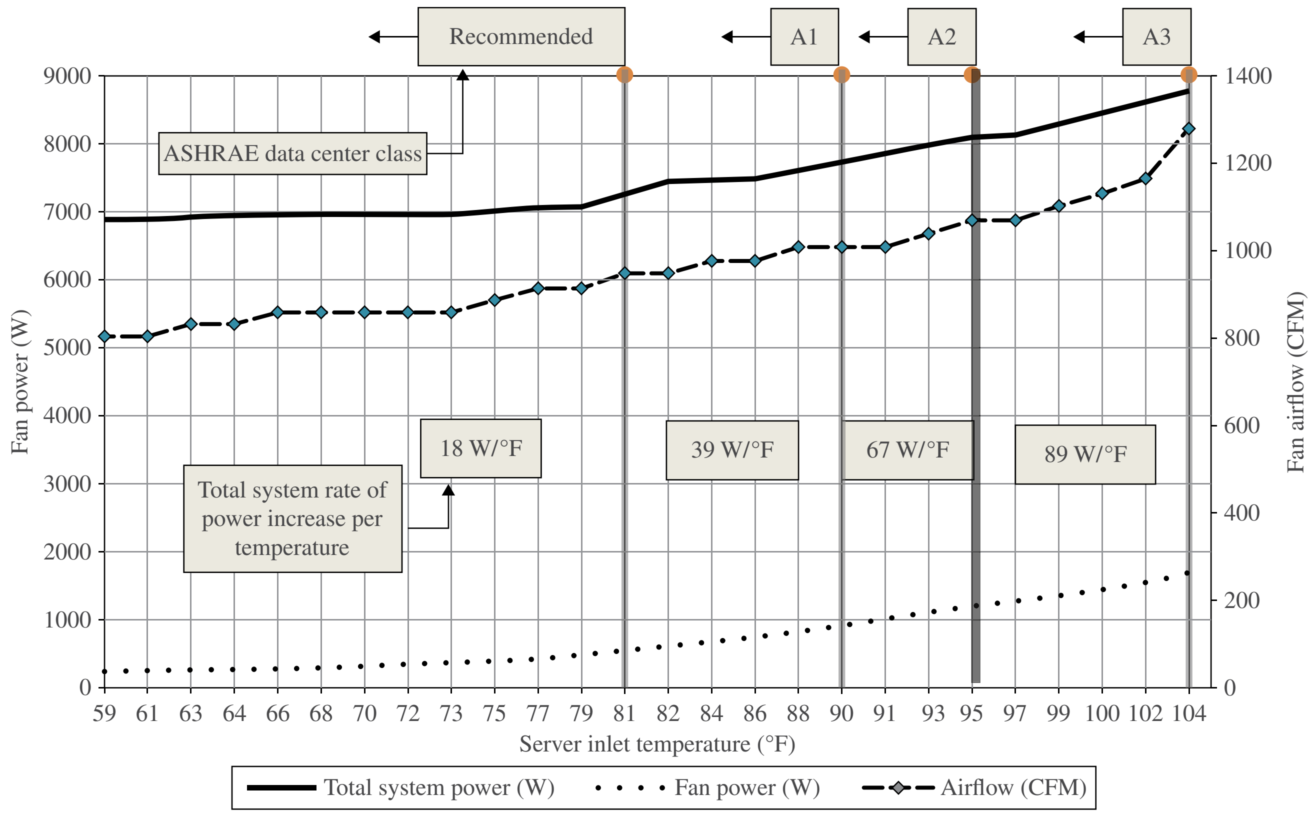


FIGURE 2.15 Increasing supply air temperature to the server by moving from ASHRAE environmental class "Recommended" to "A3" will increase not only server airflow but also the rate of change of power increase per °F. Finding the right balance point of cooling system energy efficiency and server energy use must happen on a case-by-case basis.

determine if the increase in temperature will result in a net energy savings.

2.10.3 IT and Facilities Working Together to Reduce Energy Use

Given the multifaceted interdependencies between IT and facilities, it is imperative that close communication and coordination start early in any project. When this happens, the organization gains an opportunity to investigate how the facility, power, and cooling systems will affect the servers and other IT equipment from a reliability and energy-use standpoint. Energy efficiency demands a holistic approach and incorporating energy use as one of the metrics when developing the overall IT strategy will result in a significant positive impact in the subsequent planning phases of any IT enterprise project.

If we imagine the data center as the landlord and the IT equipment as the primary tenant, it is essential that there is an ongoing dialog to understand the requirements of the tenant and the capabilities of the landlord. This interface arguably presents the greatest opportunities for overall energy use optimization in the data center. From a thermal standpoint, the computer's main mission is to keep its internal components at a prescribed maximum temperature to minimize risk of thermal shutdown, reduce electrical leakage, and, in extreme cases, mitigate any chances of physical damage to the equipment. The good news is that thermal engineers for the IT equipment understand the effects of wide temperature and humidity swings on the equipment and on the corresponding energy consumption. From a data center cooling perspective, it is essential to understand how the ambient temperature affects the power use of the computers. Based on the inlet temperature of the computer, the overall system power will change; assuming a constant workload, the server fan power will increase as the inlet temperature increases. The data center cooling strategy must account for the operation of the computers to avoid an unintentional increase in energy use by raising the inlet temperature too high.

2.11 LEVERAGING IT AND FACILITIES

Based on current market conditions, there is a confluence of events that can enable energy optimization of the IT enterprise. It just takes some good planning and a thorough understanding of all the elements that affect energy use. Meeting these multiple objectives—service enhancement, reliability, and reduction of operational costs—once thought to be mutually exclusive, must now be thought of as key success factors that must occur simultaneously. Some of the current trends in IT operations that can be leveraged to reduce/optimize energy spending are discussed in the following.

2.11.1 Technology Refresh

Computational speed and the ability to handle multiple complex simultaneous applications has enabled new business applications and expanded the use of computers and technology into almost every imaginable market sector. This "tail wagging the dog" phenomenon is driving new capital expenditures on technology and data centers to record high levels. This appears to be an unending cycle: faster computers enabling new software applications that, in turn, drive the need for newer, more memory- and speed-intensive software applications requiring new computers! This cycle will typically increase energy use; but if there already is an upgrade planned, there now exists an opportunity to leverage the upgrade by looking at energy use optimization in addition to just the usual technology upgrade to improve business results.

2.11.2 Reducing IT and Operational Costs

In order for companies to maintain a completive edge in pricing products and services, reducing ongoing operational costs related to IT infrastructure, architecture, applications, real estate, facility operational costs, and energy use often come under scrutiny. This is an area where taking a multifaceted approach starting at the overall IT strategy (infrastructure and architecture) and ending at the actual facility where the technology is housed, will reap benefits in terms of reduction of annual costs. Avoiding the myopic, singular approach is of paramount importance. The best time to incorporate thinking on energy use optimization is at the very beginning of a new IT planning effort. This is not typically the norm, so it is very important to widen out the view portal and discuss energy use. Statistics show that within 3 years of obtaining a server, the purchase will be eclipsed by the energy costs to run the server.

2.11.3 Data Center Facilities—Dynamic and Unpredictable

One of the primary design goals of a data center facility is future flexibility and scalability knowing that IT systems evolve on a life cycle of 12–18 months. This, however, can lead to short-term overprovisioning of power and cooling systems until the IT systems are fully built out. Even at a fully built-out stage, the computers, storage, and networking equipment will experience hourly, daily, weekly, and monthly variations depending on what the data center is used for. This "double learning curve" of both increasing power usage over time plus ongoing fluctuations of power use make the design and operation of these types of facilities difficult to optimize. Using simulation tools can help to show how these changes affect not only energy use, but also indoor environmental conditions, such as dry-bulb temperature, radiant temperature, and moisture content.

2.11.4 Server Technology and Application Efficiencies

In the consumer PC and laptop market, the idea of different power settings to conserve battery life or to put the PC into hibernate have been around for years. Yet this strategy is still getting a foothold in the enterprise server market due to the perceived potential reduction of reliability and processing speed. The focus for enterprise servers has traditionally been on the power-per-instruction metric at the chip level, and when used as a benchmark of energy efficiency shows tremendous improvements in energy efficiency over the past several years. However, this metric is misleading if taken out of context. The power consumption of the servers themselves has been steadily increasing.

The reality is that most facility teams do not have the time or bandwidth to understand the dramatic changes that have occurred in the past 5 years in the IT industry (power/cooling density, reliability, etc.), so they need to be more proactive in working with their IT organizations. Looking beyond the processor power consumption and into how the servers are used within an enterprise, there are certainly viable opportunities in which a true reduction in power can be achieved.

2.11.5 Data Collection and Analysis for Assessments

The cliché, "You can't manage what you don't measure," is especially important for data centers, in view of the complex system interdependences exceptionally high energy use intensity. For example, knowing the relationship between server wattage requirements and mechanical cooling costs can help in determining the value of purchasing slightly more expensive but much more efficient server power supplies and IT equipment. Also, operating costs of double-conversion UPSs can be compared to those of line-reactive units to determine if the (possibly unnecessary) additional conditioning is financially justifiable. While much of the data collection process to optimize energy efficiency is similar to what is done in commercial office buildings, schools, or hospitals, there are nuances, which if not understood, will result in a suboptimal outcome. The following points are helpful when considering an energy audit consisting of monitoring, measurement, analysis, and remediation:

1. Identifying operational or maintenance issues—In particular, to assist in diagnosing the root cause of hot spots, heat-related equipment failure, lack of overall capacity, and other common operational problems. Due to the critical nature of data center environments, such problems are often addressed in a very nonoptimal, break–fix manner due to the need for an immediate solution. Benchmarking can identify those quick fixes that should be revisited in the interests of lower operating cost or long-term reliability.
2. Helping to plan future improvements—The areas that show the poorest performance relative to other benchmark data center facilities usually offer the greatest, most economical opportunity for energy cost savings. Improvements can range from simply changing set points in order to realize an immediate payback, to replacing full systems in order to realize energy savings that will show payback over the course of several years, but also increasing system availability and lifespan.
3. Developing design standards for future facilities—Conducting benchmarking studies performed at dozens of data center facilities in recent years has shed light on the best practice design and operational approaches that result in fundamentally lower cost, more efficient facilities. Design standards, among others, can help identify company-specific best practices that should be duplicated to reduce the cost of future facilities, and identify less efficient design approaches that should be avoided.
4. Establishing a baseline performance as a diagnostic tool—Comparing trends data over time to baseline performance can help predict and avoid equipment failure, improving long-term reliability. Efficiency can also benefit by identifying and therefore allowing the correction of typical performance decay that occurs as systems age and calibrations are lost.

The ASHRAE publication, *Procedures for Commercial Building Energy Audits*, provides material on how to perform an energy audit. The publication describes three levels of audit from broad to very specific, each with its own set of criteria. In addition to understanding and optimizing energy use in the facility, the audits also include review of operational procedures, documentation, and set points. As the audit progresses, it becomes essential that deficiencies in operational procedures that are causing excessive energy use be separated out from inefficiencies in power and cooling equipment. Without this, false assumptions might be made on poor equipment performance, leading to unnecessary equipment upgrades or replacement.

ASHRAE Guideline 14-2002, *Measurement of Energy and Demand Savings*, builds on this publication and provides more detail on the process of auditing the energy use of a building. Information is provided on the actual measurement devices, such as sensors and meters, how they are to be calibrated to ensure consistent results year after year, and the duration they are to be installed to capture the data accurately. Another ASHRAE publication, *Real-Time Energy Consumption Measurements in Data Centers*, provides data-center-specific information on the best way to monitor and measure data center equipment energy use. Finally, the document *Recommendations for Measuring and Reporting Overall Data Center Efficiency* lists the specific locations where monitoring and measurement is required (Table 2.4), including which locations in the data center facility. This is important for end users to consistently report energy use in non-data center areas such as UPS and switchgear rooms, mechanical rooms, loading docks, administrative areas, and corridors. (The energy consumption of lights, equipment, and cooling systems in the non-data center areas of the building are required to report PUE as defined by the Green Grid.) Securing energy use data accurately and consistently is essential to a successful audit and energy use optimization program (Table 2.5).

2.12 DETERMINING DATA CENTER ENERGY USE EFFECTIVENESS

When analyzing and interpreting energy use statistics in a data center, it is essential industry-accepted methods are used to develop the data collection forms, analysis techniques, and reporting mechanisms. This will ensure a high confidence level that the results are valid, and not perceived as a nonstandard process that might have built-in bias. These industry standards include ASHRAE 90.1; ARI Standards 340, 365, 550–590; and others. The information contained in the ASHRAE Standard 14 is paraphrased throughout this writing.

There are several methods available to collect, analyze, and present data to demonstrate both baseline energy consumption and projected savings resulting from the implementation of energy conservation measure (ECMs). A process called a calibrated simulation analysis incorporates a wide array of stages that range from planning through implementation:

1. Produce a calibrated simulation plan. Before a calibrated simulation analysis may begin, several questions must be answered. Some of these questions include: Which software package will be applied? Will models be calibrated to monthly or hourly measured data, or both? What are to be the tolerances for the statistical indices? The answers to these questions are documented in a simulation plan.

TABLE 2.4 Recommendations for measuring and reporting overall data center efficiency

System	Units	Data Source	Duration
Total recirculation fan (total CRAC) usage	kW	From electrical panels	Spot
Total make-up air handler usage	kW	From electrical panels	Spot
Total IT equipment power usage	kW	From electrical panels	Spot
Chilled water plant	kW	From electrical panels	1 week
Rack power usage, 1 typical	kW	From electrical panels	1 week
Number of racks	number	Observation	Spot
Rack power usage, average	kW	Calculated	N/A
Other power usage	kW	From electrical panels	Spot
Data center temperatures (located strategically)	°F	Temperature sensor	1 week
Humidity conditions	R.H.	Humidity sensor	1 week
Annual electricity use, 1 year	kWh/year	Utility bills	N/A
Annual fuel use, 1 year	Therm/year	Utility bills	N/A
Annual electricity use, 3 prior years	kWh/year	Utility bills	N/A
Annual fuel use, 3 prior years	Therm/year	Utility bills	N/A
Peak power	kW	Utility bills	N/A
Average power factor	%	Utility bills	N/A
Facility (total building) area	sf	Drawings	N/A
Data center area (electrically active floor space)	sf	Drawings	N/A
Fraction of data center in use (fullness factor)	%	Area and rack observations	Spot
Airflow	CFM	(Designed, Test and Balance report)	N/A
Fan power	kW	3Φ true power	Spot
VFD speed	Hz	VFD	Spot
Set point temperature	°F	Control system	Spot
Return air temperature	°F	10 k Thermistor	1 week
Supply air temperature	°F	10 k Thermistor	1 week
RH set point	RH	Control system	Spot
Supply RH	RH	RH sensor	1 week
Return RH	RH	RH sensor	1 week
Status	Misc.	Observation	Spot
Cooling load	Tons	Calculated	N/A
Chiller power	kW	3Φ true power	1 week
Primary chilled water pump power	kW	3Φ true power	Spot
Secondary chilled water pump power	kW	3Φ true power	1 week
Chilled water supply temperature	°F	10 k Thermistor	1 week
Chilled water return temperature	°F	10 k Thermistor	1 week
Chilled water flow	gpm	Ultrasonic flow	1 week
Cooling tower power	kW	3Φ true power	1 week
Condenser water pump power	kW	3Φ true power	Spot
Condenser water supply temperature	°F	10 k Thermistor	1 week
Chiller cooling load	Tons	Calculated	N/A
Backup generator(s) size(s)	kVA	Label observation	N/A
Backup generator standby loss	kW	Power measurement	1 week
Backup generator ambient temp	°F	Temp sensor	1 week
Backup generator heater set point	°F	Observation	Spot
Backup generator water jacket temperature	°F	Temp sensor	1 week
UPS load	kW	UPS interface panel	Spot
UPS rating	kVA	Label observation	Spot
UPS loss	kW	UPS interface panel or measurement	Spot
PDU load	kW	PDU interface panel	Spot
PDU rating	kVA	Label observation	Spot
PDU loss	kW	PDU interface panel or measurement	Spot
Target	Units	Data source	Duration
Outside air dry-bulb temperature	°F	Temp/RH sensor	1 week
Outside air wet-bulb temperature	°F	Temp/RH sensor	1 week

TABLE 2.5 Location and duration of monitoring and measurement for auditing energy use and developing recommendations for increasing efficiency

ID	Data	Unit
General data center data		
dG1	Data center area (electrically active)	sf
dG2	Data center location	—
dG3	Data center type	—
dG4	Year of construction (or major renovation)	—
Data center energy data		
dA1	Annual electrical energy use	kWh
dA2	Annual IT electrical energy use	kWh
dA3	Annual fuel energy use	MMBTU
dA4	Annual district steam energy use	MMBTU
dA5	Annual district chilled water energy use	MMBTU
Air management		
dB1	Supply air temperature	°F
dB2	Return air temperature	°F
dB3	Low-end IT equipment inlet air relative humidity set point	%
dB4	High-end IT equipment inlet air relative humidity set point	%
dB5	Rack inlet mean temperature	°F
dB6	Rack outlet mean temperature	°F
Cooling		
dC1	Average cooling system power consumption	kW
dC2	Average cooling load	Tons
dC3	Installed chiller capacity (w/o backup)	Tons
dC4	Peak chiller load	Tons
dC5	Air economizer hours (full cooling)	Hours
dC6	Air economizer hours (partial cooling)	Hours
dC7	Water economizer hours (full cooling)	Hours
dC8	Water economizer hours (partial cooling)	Hours
dC9	Total fan power (supply and return)	W
dC10	Total fan airflow rate (supply and return)	CFM
Electrical power chain		
dE1	UPS average load	kW
dE2	UPS load capacity	kW
dE3	UPS input power	kW
dE4	UPS output power	kW
dE5	Average lighting power	kW

Courtesy of Lawrence Berkeley National Laboratory.

2. Collect data. Data may be collected from the building during the baseline period, the retrofit period, or both. Data collected during this step include dimensions and properties of building surfaces, monthly and hourly whole-building utility data, nameplate data from HVAC and other building system components, operating schedules, spot measurements of selected HVAC and other building system components, and weather data.
3. Input data into simulation software and run model. Over the course of this step, the data collected in the previous step are processed to produce a simulation-input file. Modelers are advised to take care with zoning, schedules, HVACs stems, model debugging (searching for and eliminating any malfunctioning or erroneous code), and weather data.
4. Compare simulation model output to measured data. The approach for this comparison varies depending on the resolution of the measured data. At a minimum, the energy flows projected by the simulation model are compared to monthly utility bills and spot measurements. At best, the two data sets are compared on an hourly basis. Both graphical and statistical means may be used to make this comparison.
5. Refine model until an acceptable calibration is achieved. Typically, the initial comparison does not yield a match within the desired tolerance. In such a case, the modeler studies the anomalies between the two data sets and makes logical changes to the model to better match the measured data. The user should calibrate to both pre- and post-retrofit data wherever possible and should only calibrate to post-retrofit data alone when both data sets are absolutely unavailable. While the graphical methods are useful to assist in this process, the ultimate determination of acceptable calibration will be the statistical method.
6. Produce baseline and post-retrofit models. The baseline model represents the building as it would have existed in the absence of the energy conservation measures. The retrofit model represents the building after the energy conservation measures are installed. How these models are developed from the calibrated model depends on whether a simulation model was calibrated to data collected before the conservation measures were installed, after the conservation measures were installed, or both times. Furthermore, the only differences between the baseline and post-retrofit models must be limited to the measures only. All other factors, including weather and occupancy, must be uniform between the two models unless a specific difference has been observed that must be accounted for.
7. Estimate Savings. Savings are determined by calculating the difference in energy flows and intensities of the baseline and post-retrofit models using the appropriate weather file.
8. Report observations and savings. Savings estimates and observations are documented in a reviewable format. Additionally, sufficient model development and calibration documentation shall be provided to allow for accurate recreation of the baseline and post-retrofit models by informed parties, including input and weather files.

9. Tolerances for statistical calibration indices. Graphical calibration parameters as well as two main statistical calibration indices (mean bias error and coefficient of variation (root mean square error)) are required. Document the acceptable limits for these indices on a monthly and annual basis.
10. Statistical Comparison Techniques—Although graphical methods are useful for determining where simulated data differ from metered data, and some quantification can be applied, more definitive quantitative methods are required to determine compliance. Again two statistical indices are used for this purpose: hourly mean bias error and coefficient of variation of the root mean squared error [1, 2].

Using this method will result in a defendable process with results that have been developed in accordance with industry standards and best practices.

2.13 PRIVATE INDUSTRY AND GOVERNMENT ENERGY EFFICIENCY PROGRAMS

Building codes, industry standards, and regulations are used pervasively in the design and construction industry. Until recently, there was limited availability for documents explicitly written to improve energy efficiency in data center facilities. Many that did exist were meant to be used for a limited audience and others tended to be primarily anecdotal. All of that has changed over the past few years with the release of several peer-reviewed design guidelines from well-established organizations. I expect that in the near future, there will be even more detailed standards and guidelines to draw from as the industry continues to evolve.

There are several primary organizations responsible for the development and maintenance of these documents: The American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE), U.S. Green Building Council (USGBC), U.S. Environmental Protection Agency (US EPA), U.S. Department of Energy (US DOE), and the Green Grid, among others. The following is an overview of some of the standards and guidelines from these organizations, which have been developed specifically to provide advice on improving the energy efficiency in data center facilities.

2.14 USGBC—LEED ADAPTATIONS FOR DATA CENTERS

The new LEED data centers credit adaptation program was developed in direct response to challenges that arose when applying the current LEED standards to data center projects. These challenges are related to several factors including the extremely high power density found in data centers. In response, the USGBC has developed credit adaptations that address the primary challenges in certifying data center facilities. The credit adaptations were released with the LEED version 4 rating system and apply to both Building Design and Construction and Building Operations and Maintenance rating systems. Since the two rating systems apply to buildings in different stages of their life cycle, the credits are adapted in different ways. However, the adaptations were developed with the same goal in mind: establish LEED credits that are applicable to data centers specifically and that will provide tools for developers, owners, operators, designers, and builders to enable a reduction in energy use, minimize environmental impact, and provide a positive indoor environment for the inhabitants of the data center.

2.15 HARMONIZING GLOBAL METRICS FOR DATA CENTER ENERGY EFFICIENCY

In their development of data center metrics such as PUE/DCiE, CUE, and WUE, the Green Grid has sought to achieve a global acceptance to enable worldwide standardization of monitoring, measuring, and reporting data center energy use. This so-called global harmonization has manifested itself in the United States, the European Union, and Japan reaching in an agreement on guiding principles for data center energy efficiency metrics. The specific organizations that participated in this effort were U.S. Department of Energy's Save Energy Now and Federal Energy Management Programs; U.S. Environmental Protection Agency's ENERGY STAR Program; European Commission Joint Research Center Data Centers Code of Conduct; Japan's Ministry of Economy, Trade, and Industry; Japan's Green IT Promotion Council; and the Green Grid.

2.16 INDUSTRY CONSORTIUM—RECOMMENDATIONS FOR MEASURING AND REPORTING OVERALL DATA CENTER EFFICIENCY

In 2010, a taskforce consisting of representatives from leading data center organizations (7 × 24 Exchange, ASHRAE, the Green Grid, Silicon Valley Leadership Group, U.S. Department of Energy Save Energy Now Program, U.S. Environmental Protection Agency's ENERGY STAR Program, U.S. Green Building Council, and Uptime Institute) convened to discuss how to standardize the process of measuring and reporting PUE. One of the goals of the taskforce was to develop guidelines for data center owners with limited measurement capability and to allow them to participate in programs where power/energy measure is required, while also outlining a process that allows operators to add additional measurement points to increase the accuracy of their measurement program. The guidelines that were developed were meant to generate a consistent and repeatable measurement strategy that allows data center operators to monitor and improve the energy efficiency of their facility. A consistent measurement approach will also facilitate communication of PUE among data center owners and operators. (It should be noted that caution must be exercised when an organization wishes to use PUE to compare different data centers, as it is necessary to first conduct appropriate data analyses to ensure that other factors such as levels of reliability and climate are not impacting the PUE.)

2.16.1 U.S. EPA—Energy Star for Data Centers

In June 2010, the U.S. EPA released the data center model for their Portfolio Manager, an online tool for building owners to track and improve energy and water use in their buildings. This leveraged other building models that have been developed since the program started with the release of the office building model in 1999. The details of how data center facilities are ranked in the Portfolio Manager are discussed in a technical brief available on the EPA's website.

Much of the information required in attempting to obtain an Energy Star rating for a data center is straightforward. A licensed professional (Architect or Engineer) is required to validate the information that is contained in the Data Checklist. The Licensed Professional should reference the 2010 Licensed Professional's Guide to the ENERGY STAR Label for Commercial Buildings for guidance in verifying a commercial building to qualify for the ENERGY STAR.

2.16.2 ASHRAE—Green Tips for Data Centers

The ASHRAE Datacom Series is a compendium of books that provides a foundation for developing an energy-efficient design of the data center. These books are under continuous maintenance by ASHRAE to incorporate the newest design concepts that are being introduced by the engineering community. Arguably, the seminal book in the series dealing energy and sustainability topics (when these were still in unchartered territory), *Green Tips for Data Centers* was conceived to be an engineering resource that overtly provides energy and water consumption reduction strategies. It is akin to the ASHRAE *Green Guide* in overall format and organization. However, it presents chapter-by-chapter technical guidance on energy- and water-use mitigation approaches in the data center. The book is aimed at facility operators and owners, as well as engineers and other professional consultants.

2.16.3 Other International Programs and Standards

2.16.3.1 Singapore Standard for Green Data Centers—Energy and Environmental Management Systems Developed by Singapore's Green Data Centre Standards Working Group, this standard is a certifiable management system standard that provides data centers with a recognized framework as well as a logical and consistent methodology to achieve energy efficiency and continuous improvement. The standard also recommends best practices for data centers and lays out several metrics that can be used to measure performance and energy efficiency.

2.16.4 FIT4Green

An EU consortium made up of private and public organizations from Finland, Germany, Italy, the Netherlands, Spain, and the United Kingdom, FIT4Green "aims at contributing to ICT energy reducing efforts by creating an energy-aware layer of plug-ins for data centre automation frameworks, to improve energy efficiency of existing IT solution deployment strategies so as to minimize overall power consumption, by moving computation and services around a federation of IT data centers sites."

2.16.5 Guidelines for Environmental Sustainability Standard for the ICT Sector

The impetus for this project is that the ICT sector is increasingly asking a number of customers, investors, governments and other stakeholders to report on sustainability performance, but there is lack of an agreed-upon standardized measurement that would simplify and streamline this reporting specifically for the ICT sector. The standard provides a set of agreed-upon sustainability requirements for ICT companies that allow for a more objective reporting of how sustainability is practiced in the ICT sector in these key areas: sustainable buildings, sustainable ICT, sustainable products, sustainable services, end-of-life management, general specifications, and assessment framework for environmental impacts of the ICT sector.

There are several other standards that range from firmly established to still-emerging not mentioned here. The landscape for the standards and guidelines for data centers is growing, and it is important that both the IT and facilities personnel become familiar with them and apply them where relevant [3].

One of the essential objectives of this effort was to develop an integrated, cross-disciplinary toolkit containing sustainability requirements to guide data center owners through efforts to improve their eco-efficiency, and facilitating equitable and transparent sustainability reporting. The ITU-T with over 50 organizations and ICT companies

TABLE 2.6 Example of analysis and recommendations for increasing data center efficiency and improving operational performance

Title	Description
Supply air temperatures to computer equipment is too cold	Further guidance can be found in "Design Considerations for Datacom Equipment centers" by ASHRAE and other updated recommendations. Guidelines recommended range is 64.5–80°F. However, the closer the temperatures to 80°F, the more energy efficient the data center becomes.
Relocate high-density equipment to within area of influence of CRACs	High-density racks should be as close as possible to CRAC/H units unless other means of supplemental cooling or chilled water cabinets are used.
Distribute high-density racks	High-density IT hardware racks are distributed to avoid undue localized loading on cooling resources.
Provide high-density heat containment system for the high-density load area	For high-density loads, there are a number of design concepts whose basic intent is to contain and separate the cold air from the heated return air on the data floor: hot-aisle containment; cold-aisle containment; contained rack supply, room return; room supply, contained rack return; and contained rack supply, contained rack return
Install strip curtains to segregate airflows	While this will reduce recirculation, access to cabinets needs to be carefully considered.
Correct situation to eliminate air leakage through the blanking panels	Although blanking panels are installed, it was observed that they are not in snug-fit "properly fit" position, and some air appears to be passing through openings up and below the blanking panels.
Increase CRAH air discharge temperature and chilled water supply set points by 2°C (~4°F)	Increase the set point by 0.6°C (1°F) reduces chiller power consumption 0.75–1.25% of fixed speed chiller kW/TON and 1.5–3% for VSD chiller. Increasing the set point and widening the range of economizer operation will result in greater energy savings.
Widen %RH range of CRAC/H Units	Humidity range is too tight. Humidifiers will come on more often. ASHRAE recommended range for servers' intake is 30–80%RH. Widening the %RH control range (within ASHRAE guidelines) will enable less humidification ON time and hence less energy utilization. In addition, this will help to eliminate any controls fighting.

developed a toolkit that aimed at buildings, sustainable ICT, sustainable products and services, end of life management for ICT equipment, general specifications, and an assessment framework for environmental impacts of the ICT sector.

2.17 STRATEGIES FOR OPERATIONS OPTIMIZATION

Many of the data center energy efficiency standards and guidelines available today tend to focus on energy conservation measures that involve improvements to the power and cooling systems. Or if the facility is new, strategies that can be used in the design process to ensure an efficient data center. One topic that needs more exposure is how to improve energy use through better operations.

Developing a new data center includes expert design engineers, specialized builders, and meticulous commissioning processes. If the operation of the facility is out-of-sync with the design and construction process, it is probable that deficiencies will arise in the operation of the power and cooling systems. Having a robust operations optimization process in place will identify and neutralize these discrepancies and move the data center toward enhanced energy efficiency (Table 2.6).

REFERENCES

[1] Kreider JF, Haberl JS. Predicting hourly building energy use. ASHRAE Trans 1994;100(2):1104–1118.

[2] Haberl JS, Thamilseran S. The great energy predictor shootout II—measuring retrofit savings—overview and discussion of results. ASHRAE Trans 1996;102(2):419–435.

[3] International Telecommunication Union. *Guidelines for Environmental Sustainability Standard for the ICT Sector*. Geneva: International Telecommunication Union; 2012.

FURTHER READING

AHRI Standard 1060 (I-P)-2013. Performance Rating of Air-to-Air Heat Exchangers for Energy Recovery Ventilation Equipment.

ANSI/AHRI 365 (I-P)-2009. Commercial and Industrial Unitary Air-Conditioning Condensing Units.

ANSI/AHRI 540-2004. Performance Rating of Positive Displacement Refrigerant Compressors and Compressor Units.

ANSI/AHRI 1360 (I-P)-2013. Performance Rating of Computer and Data Processing Room Air Conditioners.

ASHRAE Standard 90.1-2013. (I-P Edition)—Energy Standard for Buildings Except Low-Rise Residential Buildings.

ASHRAE. *Thermal Guidelines for Data Processing Environments*. 3rd ed.

ASHRAE. Liquid Cooling Guidelines for Datacom Equipment Centers.

ASHRAE. Real-Time Energy Consumption Measurements in Data Centers.

ASHRAE. *Procedures for Commercial Building Energy Audits*. 2nd ed.

ASHRAE. Guideline 14-2002—Measurement of Energy and Demand Savings.

Building Research Establishment's Environmental Assessment Method (BREEAM) Data Centres 2010.

Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric, the Green Grid.

ERE: A Metric for Measuring the Benefit of Reuse Energy from a Data Center, the Green Grid.

Green Grid Data Center Power Efficiency Metrics: PUE and DCIE, the Green Grid.

Green Grid Metrics: Describing Datacenter Power Efficiency, the Green Grid.

Guidelines and Programs Affecting Data Center and IT Energy Efficiency, the Green Grid.

Guidelines for Energy-Efficient Datacenters, the Green Grid.

Harmonizing Global Metrics for Data Center Energy Efficiency Global Taskforce Reaches Agreement on Measurement Protocols for GEC, ERF, and CUE—Continues Discussion of Additional Energy Efficiency Metrics, the Green Grid.

Koomey JG, Ph.D. Estimating Total Power Consumption by Servers in the U.S. and the World.

Koomey JG, Ph.D. Growth in Data Center Electricity Use 2005 to 2010.

Lawrence Berkeley Lab High-Performance Buildings for High-Tech Industries, Data Centers.

Proxy Proposals for Measuring Data Center Productivity, the Green Grid.

PUE™: A Comprehensive Examination of the Metric, the Green Grid.

Qualitative Analysis of Power Distribution Configurations for Data Centers, the Green Grid.

Recommendations for Measuring and Reporting Overall Data Center Efficiency Version 2—Measuring PUE for Data Centers, the Green Grid.

Report to Congress on Server and Data Center Energy Efficiency Public Law 109–431 U.S. Environmental Protection Agency—ENERGY STAR Program.

Singapore Standard SS 564: 2010 Green Data Centres.

United States Public Law 109–431, December 20, 2006.

Usage and Public Reporting Guidelines for the Green Grid's Infrastructure Metrics (PUE/DCIE), the Green Grid.

US Green Building Council—LEED Rating System.

Water Usage Effectiveness (WUE™): A Green Grid Data Center Sustainability Metric, the Green Grid.

3

HOSTING OR COLOCATION DATA CENTERS

Chris Crosby and Chris Curtis
Compass Datacenters, Dallas, TX, USA

3.1 INTRODUCTION

"Every day Google answers more than one billion questions from people around the globe in 181 countries and 146 languages."[1] At 700,000 searches per minute, over 1,800 terabytes of new information is created. In a single year, over 88 quadrillion e-mail messages are generated. The vast majority of this information is not just transmitted but stored for repeated access, which means that organizations must continually expand the number of servers and storage devices to process this increasing volume of information. All of those servers and storage devices need a data center to call home, and every organization needs to have a data center strategy that will meet their computing needs both now and in the future. Not all data centers are the same, though, and taking the wrong approach can be disastrous both technically and financially. Organizations must therefore choose wisely, and this chapter provides valuable information to help organizations make an informed choice and avoid the most common mistakes.

Historically, the vast majority of corporate computing was performed within data center space that was built, owned, and operated by the organization itself. In some cases, it was merely a back room in the headquarters that was full of servers and patch panels. In other cases, it was a stand-alone, purpose-built data center facility that the organization's IT team commissioned. Whether it was a humble back room devoted to a few servers or a large facility built with a significant budget, what they had in common was that the organization was taking on full responsibility for every aspect of data center planning, development, and operations.

In recent years, this strategy has proven to be cumbersome, inefficient, and costly as data processing needs have rapidly outstripped the ability of a large number of businesses to keep up with them. The size, cost, and complexity of today's data centers have prompted organizations that previously handled all their data center operations "in-house" to come to the conclusion that data centers are not their core competency. Data centers were proving to be a distraction for the organization's internal IT teams, and the capital and costs involved in these projects were becoming an increasingly larger burden on the organization's IT budget. That created a market opportunity for data center providers who could relieve organizations of this technical and financial burden, and a variety of new vendors emerged to offer data center solutions that meet those needs.

Although these new businesses use a variety of business models, they may be categorized under two generalized headings:

1. Hosting
2. Colocation (Wholesale Data Centers)

3.2 HOSTING

In their simplest form, hosting companies lease the actual servers (or space on the servers) as well as storage capacity to companies. The equipment and the data center it resides in

[1] http://www.google.com/competition/howgooglesearchworks.html

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

are owned and operated by the hosting provider. Underneath this basic structure, customers are typically presented with a variety of options. These product options tend to fall within three categories:

1. Computing Capacity
2. Storage
3. Managed Services

3.2.1 Computing Capacity

Computing capacity offerings can vary widely in a hosted environment from space on a provider-owned server all the way up to one or more racks within the facility. For medium- to enterprise-sized companies, the most commonly used hosting offering is typically referred to as colocation. These offerings provide customers with a range of alternatives from leasing space in a single provider-supported rack all the way up to leasing multiple racks in the facility. In all of these offerings, the customer's own server and storage equipment is housed in the leased rack space. Typically, in multirack environments, providers also offer the customer the ability to locate all their equipment in a locked cage to protect against unauthorized access to the physical space.

Customer leases in colocated environments cover the physical space and the maintenance for the data center itself. Although some providers may charge the customer for the bandwidth they use, this is not common as most companies operating in this type of environment make their own connectivity arrangements with a fiber provider that is supported in the facility. Providers typically offer facility access to multiple fiber providers to offer their customers with a choice in selecting their connectivity company. The most important lease element is for the actual power delivered to the customer. The rates charged to the customer may vary from "pass through," in which the power charge from the utility is billed directly to the customer with no markup, to a rate that includes a markup added by the data center provider.

3.2.2 Storage

Although most firms elect to use their own storage hardware, many providers do offer storage capacity to smaller customers. Typically, these offerings are based on a per gigabyte basis with the charge applied monthly.

3.2.3 Managed Services

"Managed services" is the umbrella term used to describe the on-site support functions that the site's provider performs on behalf of their customers. Referred to as "remote" or "warm" hands, these capabilities are often packaged in escalating degrees of functions performed. At the most basic level, managed services offerings can be expected to include actions such as restarting servers and performing software upgrades. Higher-level services can include activities like hardware monitoring, performing moves, adds and changes, administering Internet security, and the availability of customer monitoring and tracking portals. These services are typically billed to the customer on a monthly basis.

3.3 COLOCATION (WHOLESALE)

The term "colocation" as used to describe the providers who lease only data center space to their customers has been replaced by the term "wholesale" data centers. Wholesale data center providers lease physical space within their facilities to one or more customers. Wholesale customers tend to be larger, enterprise-level organizations with data center requirements of 1 MW power capacity. In the wholesale model, the provider delivers the space and power to the customer and also operates the facility. The customer maintains operational control over all of their equipment that is used within their contracted space.

Traditionally, wholesale facilities have been located in major geographic markets. This structure enables providers to purchase and build out large-capacity facilities ranging from as little as 20,000 ft^2 to those featuring a million square feet of capacity or more. Customers then lease the physical space and their required power from the provider. Within these models, multiple customers operate in a single facility in their own private data centers while sharing the common areas of the building such as security, the loading dock, and office space.

3.4 TYPES OF DATA CENTERS

Within the past 5 years, wholesale providers have found that it is more cost-efficient and energy-efficient to build out these facilities in an incremental fashion. As a result, many providers have developed what they refer to as "modular" data centers. This terminology has been widely adopted but no true definition for what constitutes a modular data center has been universally embraced. At the present time, there are five categories of data centers that are generally considered to be "modular" within the marketplace.

3.4.1 Traditional Design

Traditional modular data centers (Fig. 3.1) are building-based solutions that use *shared* internal and external backplanes or plant (e.g., chilled water plant and parallel generator plant). Traditional data centers are either built all at once, or, as more recent builds have been done, are expanded through adding new data halls within the building. The challenge with shared backplanes is the introduction of risk due to an entire system shutdown because of cascading failures across the backplane. For "phased builds" in which additional data halls are added over time, the key drawback to this new approach is the

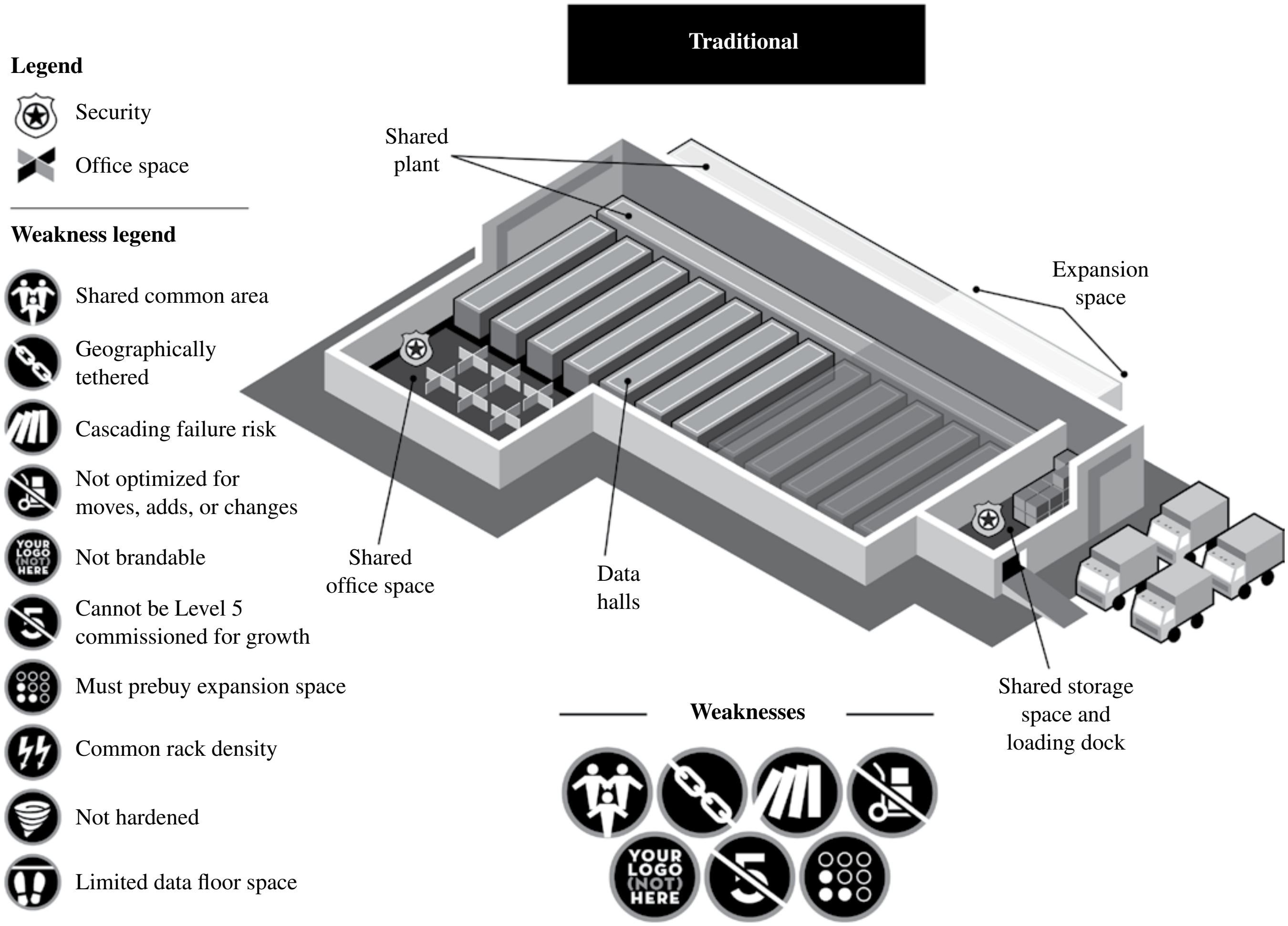


FIGURE 3.1 Traditional wholesale data centers are good solutions for IT loads above 5 MW. Courtesy of Compass Data Centers.

use of a shared backplane. In this scenario, future "phases" cannot be commissioned to Level 5 Integrated System Level[2] since other parts of the data center are already live.

In Level 5 Commissioning, all of the systems of the data center are tested under full load to ensure that they work both individually and in combination so that the data center is ready for use on day one.

Strengths:

- Well suited for single users
- Good for large IT loads, 5 MW+ day-one load

Weaknesses:

- Cascading failure potential on shared backplanes
- Cannot be Level 5 commissioned (in phased implementations)
- Geographically tethered (this can be a bad bet if the projected large IT load never materializes)
- Shared common areas with multiple companies or divisions (the environment is not dedicated to a single customer)
- Very large facilities that are not optimized for Moves/Adds/Changes

3.4.2 Monolithic Modular (Data Halls)

As the name would imply, Monolithic Modular data centers (Fig. 3.2) are large building-based solutions. Like traditional facilities, they are usually found in large buildings and provide 5 MW+ of IT power day one with the average site featuring 5–20 MW of capacity. Monolithic Modular facilities use segmentable backplanes to support their data halls so they do not expose customers to single points of failure and each data hall can be independently Level 5 commissioned prior to customer occupancy. Often, the only shared component of the mechanical and electrical plant is the medium-voltage utility gear. Because these solutions are

[2] Building Commissioning Association (http://www.bcxa.org/)

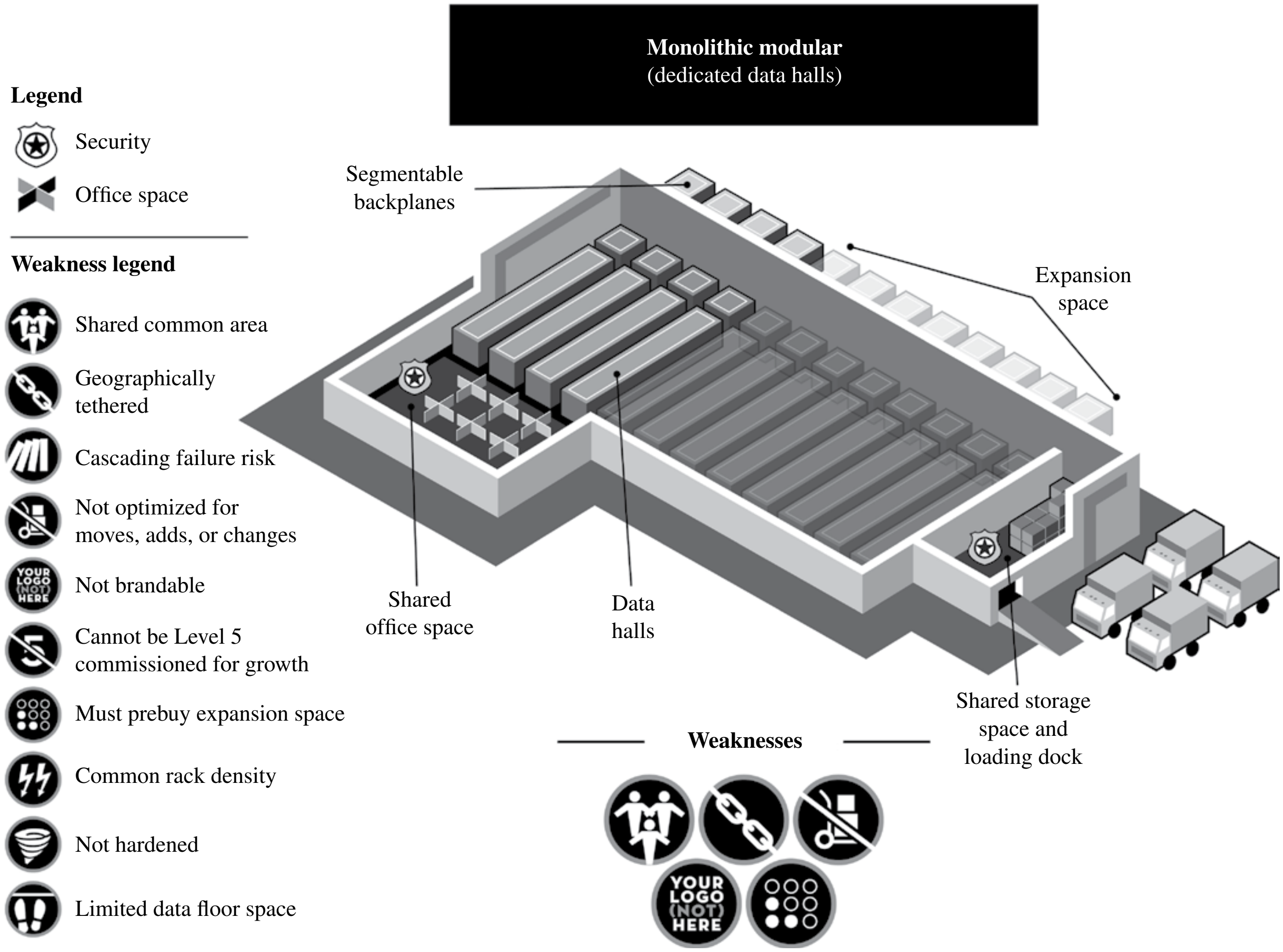


FIGURE 3.2 Monolithic Modular data centers with data halls feature segmental backplanes that avoid the possibility of cascading failure found with traditional designs. Courtesy of Compass Data Centers.

housed within large buildings, the customer may sacrifice a large degree of facility control and capacity planning flexibility if the site houses multiple customers. Additionally, security and common areas (offices, storage, staging, and the loading dock) are shared with the other occupants within the building. The capacity planning limit is a particularly important consideration as customers must prelease (and pay for) shell space within the facility to ensure that it is available when they choose to expand.

Strengths:

- Good for users with known, fixed IT capacity, for example, 4 MW day one, growing to 7 MW by year 4, with fixed takedowns of 1 MW per year,
- Optimal for users with limited Moves/Adds/Changes
- Well suited for users that don't mind sharing common areas
- Good for users that don't mind outsourcing security

Weaknesses:

- Must pay for unused expansion space
- Geographically tethered, large buildings often require large upfront investment
- Outsourced security
- Shared common areas with multiple companies or divisions (the environment is not dedicated to a single customer)
- Very large facilities that are not optimized for Moves/Adds/Changes

3.4.3 Containerized

Commonly referred to as "containers" (Fig. 3.3), prefabricated data halls are standardized units contained in ISO shipping containers that can be delivered to a site to fill an immediate need. Although advertised as quick to deliver, customers are often required to provide the elements of the shared outside plant including generators, switch gear, and,

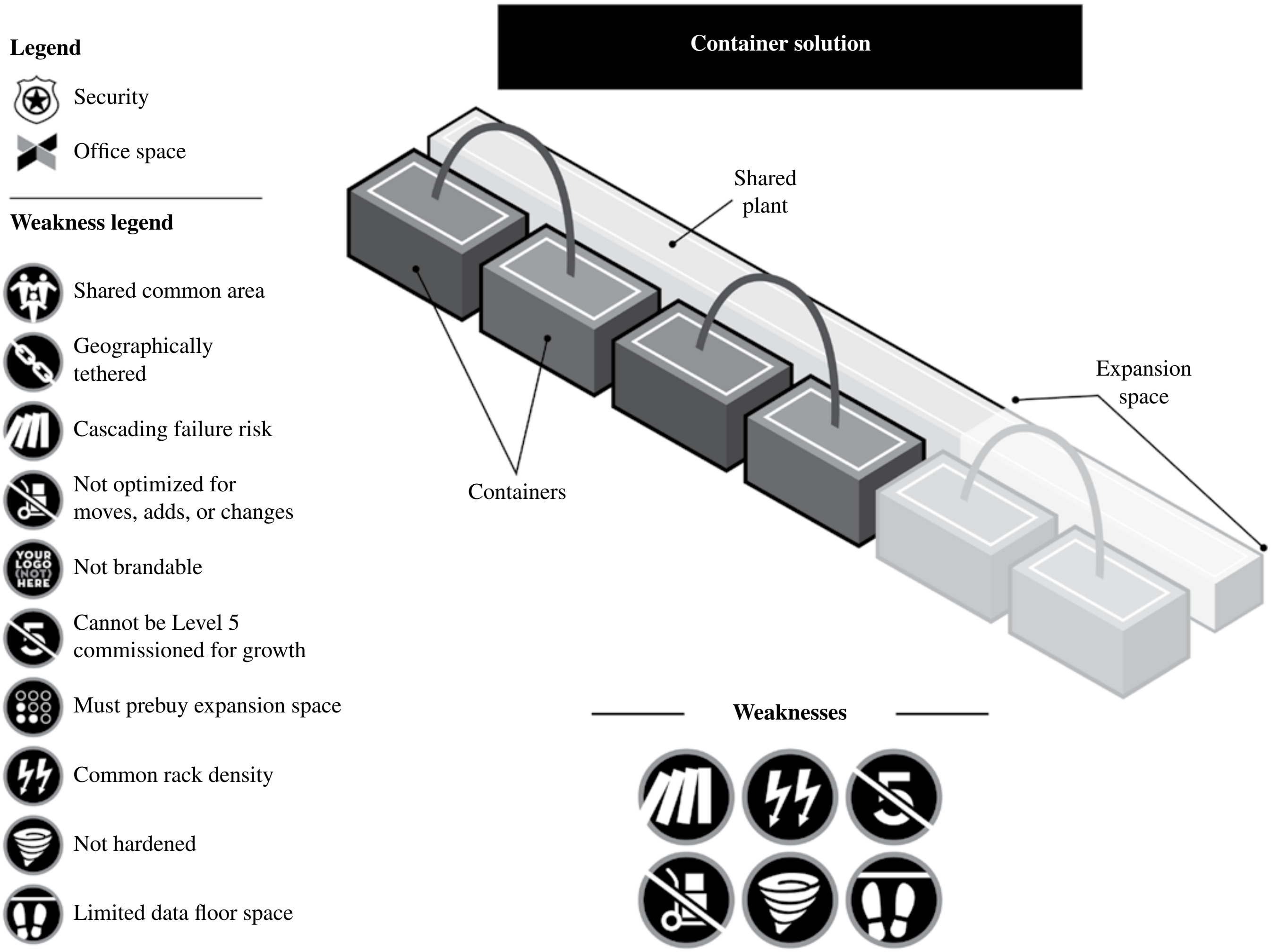


FIGURE 3.3 Container solutions are best suited for temporary applications. Courtesy of Compass Data Centers.

sometimes, chilled water. These backplane elements, if not in place, can take upward of 8 months to implement, often negating the benefit of speed of implementation. As long-term solutions, prefabricated containers may be hindered by their nonhardened designs that make them susceptible to environmental factors like wind, rust, and water penetration and their space constraints that limit the amount of IT gear that can be installed inside them. Additionally, they do not include support space like a loading dock, a storage/staging area, or security stations, thereby making the customer responsible for their provision.

Strengths:
- Optimized for temporary data center requirements
- Good for applications that work in a few hundred of KW load groups
- Support batch processing or supercomputing applications
- Suitable for remote, harsh locations (such as military locales)
- Designed for limited Move/Add/Change requirements
- Homogeneous rack requirement applications

Weaknesses:
- Lack of security
- Nonhardened design
- Limited space
- Cascading failure potential
- Cannot be Level 5 commissioned when expanded
- Cannot support heterogeneous rack requirements
- No support space

3.4.4 Monolithic Modular (Prefabricated)

These building-based solutions are similar to their data hall counterparts with the exception that they are populated with the provider's prefabricated data halls. The prefabricated data hall (Fig. 3.4) necessitates having tight control over the applications of the user. Each application set

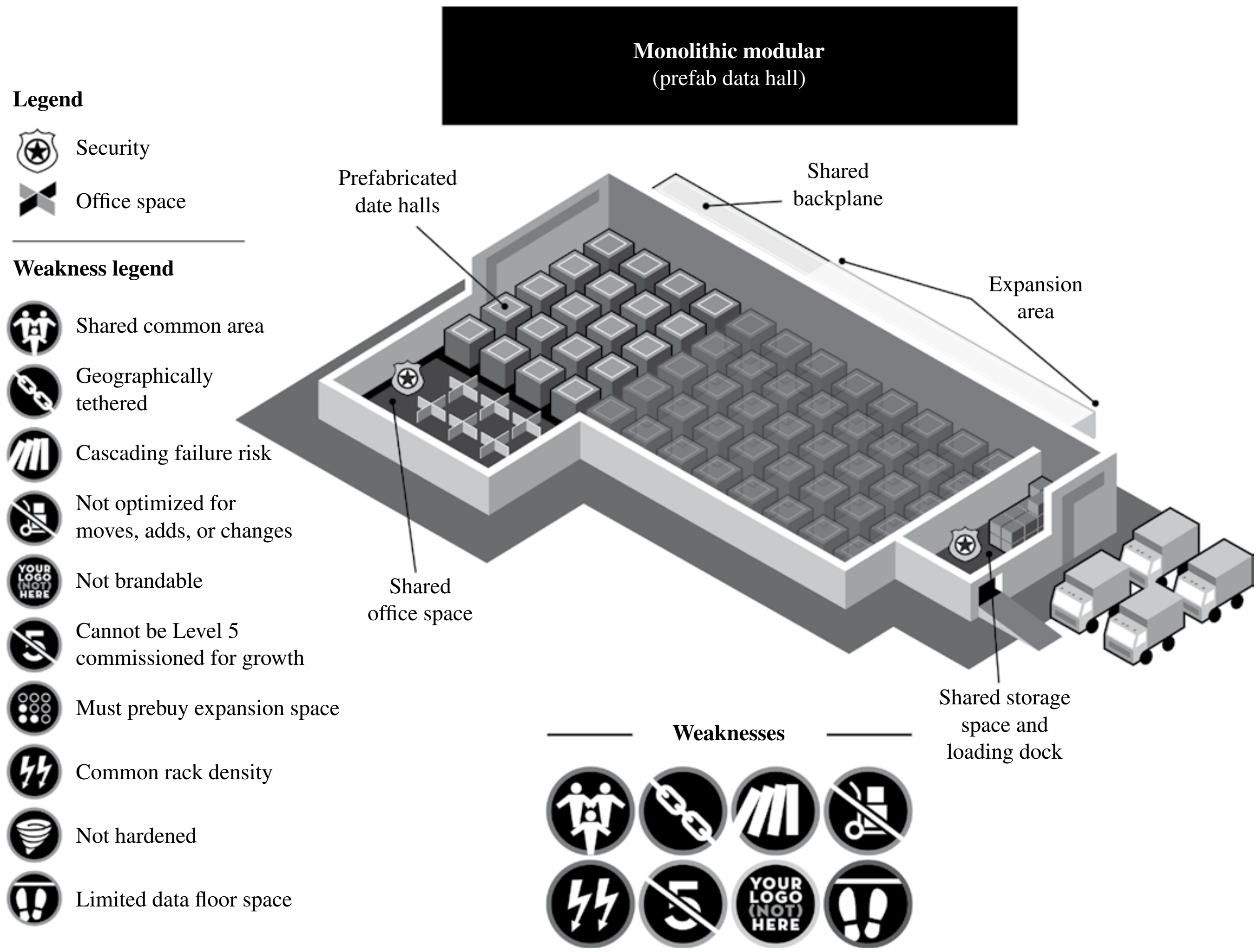


FIGURE 3.4 Monolithic Modular data centers with prefabricated data halls use a shared backplane architecture that raises the risk of cascading failure in the event of an attached unit. Courtesy of Compass Data Centers.

should drive the limited rack space to its designed load limit to avoid stranding IT capacity. For example, low-load-level groups go in one type of prefabricated data hall and high-density load groups go into another. These sites can use shared or segmented backplane architectures to eliminate single points of failure and to enable each unit to be Level 5 commissioned. Like other monolithic solutions, these repositories for containerized data halls require customers to prelease and pay for space in the building to ensure that it is available when needed to support their expanded requirements.

Strengths:

- Optimal for sets of applications in homogeneous load groups
- Designed to support applications that work in kW load groups of a few hundred kW in total IT load
- Good for batch and supercomputing applications
- Optimal for users with limited Moves/Adds/Changes
- Good for users that don't mind sharing common areas

Weaknesses:

- Outsourced security
- Expansion space must be preleased
- Shared common areas with multiple companies or divisions (the environment is not dedicated to a single customer)
- Since it still requires a large building upfront, may be geographically tethered
- Very large facilities that are not optimized for Moves/Adds/Changes

3.4.5 Stand-Alone Data Centers

Stand-alone data centers use modular architectures in which the main components of a data center have been incorporated into a hardened shell that is easily expandable in standard-sized increments. Stand-alone facilities are designed to be complete solutions that meet the certifications standards for reliability and building efficiency. Stand-alone data centers have been developed to provide geographically independent alternatives for customers who want a data center dedicated to their own use, physically located where it is needed.

By housing the data center area in a hardened shell that can withstand extreme environmental conditions, stand-alone solutions differ from prefabricated or container-based data centers that require the customer or provider to erect a building if they are to be used as a permanent solution. By using standard power and raised floor configurations, stand-alone data centers simplify customers' capacity planning capability by enabling them to add capacity as it is needed rather than having to prelease space within a facility as in the case of monolithic modular solutions, for example.

Because they provide customers with their own dedicated facility, stand-alone data centers use their modular architectures to provide customers with all the site's operational components (office space, loading dock, storage and staging areas, break room, and security area) without the need to share them as in other modular solutions (Fig. 3.5).

Strengths:

- Optimized for security conscious users
- Good for users who do not like to share any mission-critical components
- Optimal for geographically diverse locations
- Good for applications with 1–4 MW of load and growing over time
- Design for primary and disaster recovery data centers
- Suitable for provider data centers
- Meet heterogeneous rack and load group requirements

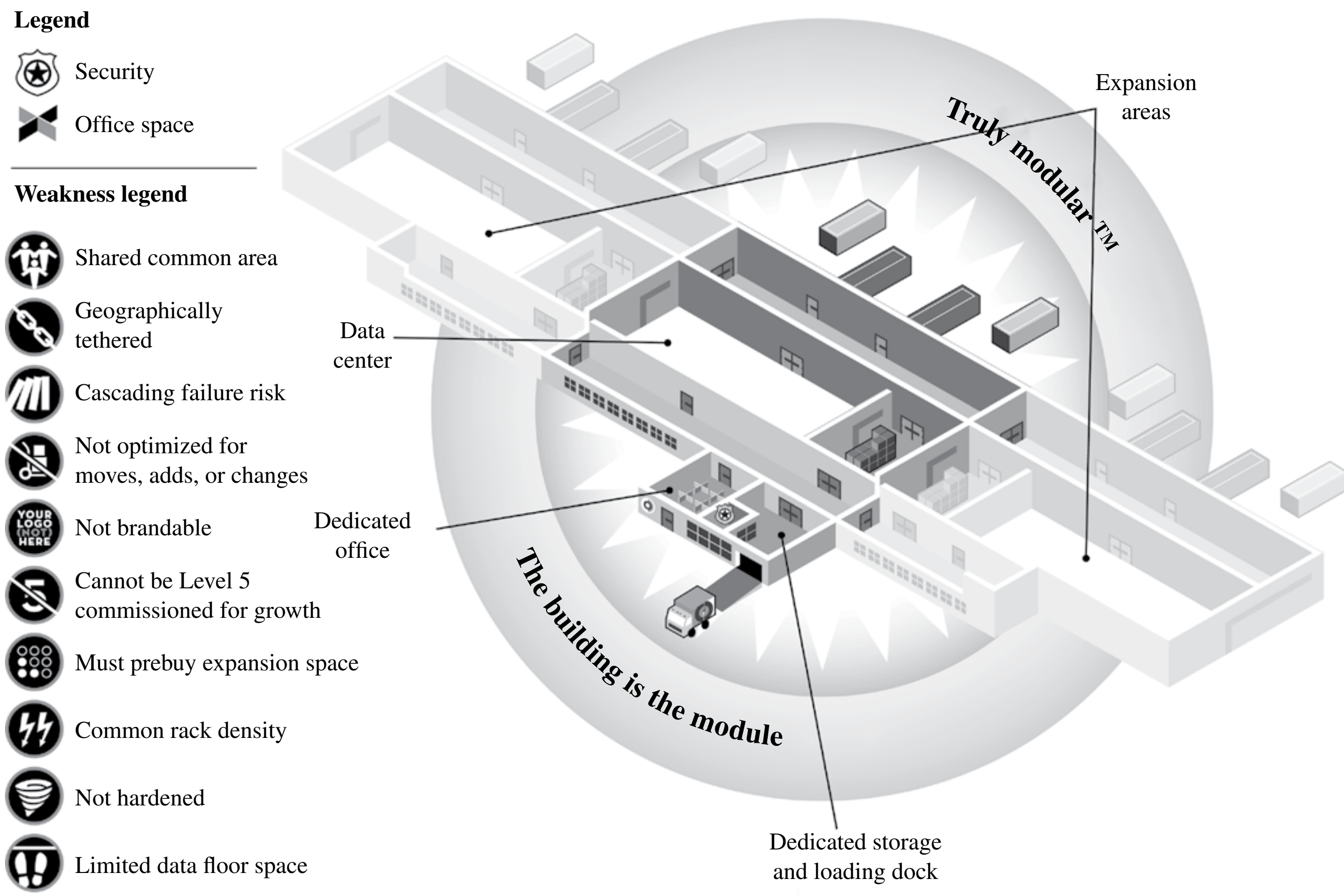


FIGURE 3.5 Stand-alone data centers combine all of the strengths of the other data center types while eliminating their weaknesses. Courtesy of Compass Data Centers.

Weaknesses:

- Initial IT load over 4 MW
- Non-mission-critical data center applications

3.5 SCALING DATA CENTERS

Scaling, or adding new data centers, is possible using either a hosting or wholesale approach. A third method, Build to Suit, where the customer pays to have their data centers custom-built where they want them may also be used, but this approach is quite costly. The ability to add new data centers across a country or internationally is largely a function of geographic coverage of the provider and the location(s) that the customer desires for their new data centers.

For hosting customers, the ability to use the same provider in all locations limits the potential options available to them. There are a few hosting-oriented providers (e.g., Equinix and Savvis) that have locations in all of the major international regions (North America, Europe, and Asia Pacific). Therefore, the need to add hosting-provided services across international borders may require a customer to use different providers based on the region desired.

The ability to scale in a hosted environment may also require a further degree of flexibility on the part of the customer regarding the actual physical location of the site. No provider has facilities in every major country. Typically, hosted locations are found in the major metropolitan areas in the largest countries in each region. Customers seeking U.S. locations will typically find the major hosting providers located in cities such as New York, San Francisco, and Dallas, for example, while London, Paris, Frankfurt, Singapore, and Sydney tend to be common sites for European and Asia Pacific international locations.

Like their hosting counterparts, wholesale data center providers also tend to be located in major metropolitan locations. In fact, this distinction tends to be more pronounced as the majority of these firms' business models require them to operate facilities of 100,000 ft^2 or more to achieve the economies of scale necessary to offer capacity to their customers at a competitive price point. Thus, the typical wholesale customer that is looking to add data center capacity across domestic regions, or internationally, may find that their options tend to be focused in the same locations as for hosting providers.

3.6 SELECTING AND EVALUATING DC HOSTING AND WHOLESALE PROVIDERS

In evaluating potential hosting or wholesale providers from the perspective of their ability to scale, the most important element for customers to consider is the consistency of their operations. Operational consistency is the best assurance that customers can have (aside from actual Uptime Institute Tier III or IV certification[3]) that their providers' data centers will deliver the degree of reliability or uptime that their critical applications require. In assessing this capability, customers should examine each potential provider based on the following capabilities:

- Equipment Providers: The use of common vendors for critical components such as UPS or generators enables a provider to standardize operations based on the vendors' maintenance standards to ensure that maintenance procedures are standard across all of the provider's facilities.
- Documented Processes and Procedures: A potential provider should be able to show prospective customers its written processes and procedures for all maintenance and support activities. These procedures should be used for the operation of each of the data centers in their portfolio.
- Training of Personnel: All of the operational personnel who will be responsible for supporting the provider's data centers should be vendor certified on the equipment they are to maintain. This training ensures that they understand the proper operation of the equipment, its maintenance needs, and troubleshooting requirements.

The ability for a provider to demonstrate the consistency of their procedures along with their ability to address these three important criteria is essential to assure their customers that all of their sites will operate with the highest degree of reliability possible.

3.7 BUILD VERSUS BUY

Build versus buy (or lease in this case) is an age-old business question. It can be driven by a variety of factors such as the philosophy of the organization itself or a company's financial considerations. It can also be affected by issues like the cost and availability of capital or the time frames necessary for the delivery of the facility. The decision can also differ based on whether or not the customer is considering a wholesale data center or a hosting solution.

3.7.1 Build

Regardless of the type of customer, designing, building, and operating a data center are unlike any other type of building. They require a specialized set of skills and expertise. Due to

[3]The Uptime Institute's Tier system establishes the requirements that must be used to provide specified levels of uptime. The most common of the system's four tiers is Tier III (99.995% uptime) that requires redundant configurations on major system components. Although many providers will claim that their facilities meet these requirements, only a facility that has been certified as meeting these conditions by the Institute are actually certified as meeting these standards.

the unique requirements of a data center, the final decision to lease space from a provider or to build their own data center requires every business to perform a deep level of analysis of their own internal capabilities and requirements and those of the providers they may be considering.

Building a data center requires an organization to use professionals and contractors from outside of their organization to complete the project. These individuals should have demonstrable experience with data centers. This also means that they should be aware of the latest technological developments in data design and construction and the evaluation process for these individuals and firms should focus extensively on these attributes.

3.7.2 Leasing

Buying a data center offers many customers a more expedient solution than building their own data center, but the evaluation process for potential providers should be no less rigorous. While experience with data centers probably isn't an issue in these situations, prospective customers should closely examine the provider's product offerings, their existing facilities, their operational records, and, perhaps most importantly, their financial strength as signing a lease typically means at least a 5-year commitment with the chosen provider.

3.7.3 Location

Among the most important build-versus-buy factors is the first—where to locate it. Not just any location is suitable for a data center. Among the factors that come into play in evaluating a potential data center site are the cost and availability of power (and potentially water). The site must also offer easy access to one or more fiber network carriers. Since data centers support a company's mission-critical applications, the proposed site should be far from potentially hazardous surroundings. Among the risk factors that must be eliminated are the potential for floods, seismic activity, as well as "man-made" obstacles like airplane flight paths or chemical facilities.

Due to the critical nature of the applications that a data center supports, companies must ensure that the design of their facility (if they wish to build), or that of potential providers if leasing is a consideration, is up to the challenge of meeting their reliability requirements. As we have previously discussed, the tier system of the Uptime Institute can serve as a valuable guide in developing a data center design, or evaluating a providers', that meets an organization's uptime requirements.

3.7.4 Redundancy

The concept of "uptime" was pioneered by the Uptime Institute and codified in its Tier Classification system. In this system, there are four levels (I, II, III, and IV). Within this system, the terms "N, $N+1$, and $2N$" typically refer to the number of power and cooling components that comprise the entire data center infrastructure systems. "N" is the minimum rating of any component (such as a UPS or cooling unit) required to support the site's critical load. An "N" system is nonredundant and the failure of any component will cause an outage. "N" systems are categorized as Tier I. $N+1$ and $2N$ represent increasing levels of component redundancies and power paths that map to Tiers II–IV. It is important to note, however, that the redundancy of components does not ensure compliance with the Uptime Institute's Tier level.[4]

3.7.5 Operations

Besides redundancy, the ability to do planned maintenance or emergency repairs on systems may involve the necessity to take them off-line. This requires that the data center support the concept of "concurrent maintainability." Concurrent maintainability permits the systems to be bypassed without impacting the availability of the existing computing equipment. This is one of the key criteria necessary for a data center to receive Tier III or IV certification from the Uptime Institute.

3.7.6 Build versus Buy Using Financial Considerations

The choice to build or lease should include a thorough analysis of the data center's compliance with these Tier requirements to ensure that it is capable of providing the reliable operation necessary to support mission-critical applications.

Another major consideration for businesses in making a build-versus-lease decision is the customer's financial requirements and plans. Oftentimes, these considerations are driven by the businesses' financial organizations. Building a data center is a capital-intensive venture. Companies considering this option must answer a number of questions including:

- Do they have the capital available?
- What is the internal cost of money within the organization?
- How long do they intend to operate the facility?
- What depreciation schedules do they intend to use?

Oftentimes, the internal process of obtaining capital can be long and arduous. The duration of this allocation and approval process must be weighed against the estimated time that the data center is required. Very often, there is also no guarantee that the funds requested will be approved, thereby stopping the project before it starts.

The cost of money (analogous to interest) is also an important element in the decision-making process to build

[4] Executive Guide Series, *Build versus Buy*, Data Center Knowledge, p. 4.

a data center. The accumulated costs of capital for a data center project must be viewed in comparison to other potential allocations of the same level of funding. In other words, based on our internal interest rate, are we better-off investing the same amount of capital in another project or instrument that will deliver a higher return on the company's investment?

The return on investment question must address a number of factors, not the least of which is the length of time the customer intends to operate the facility and how they will write down this investment over time. If the projected life span for the data center is relatively short, less than 10 years, for example, but the company knows it will continue to have to carry the asset on its books beyond that, building a facility may not be the most advantageous choice.

Due to the complexity of building a data center and obtaining the required capital, many businesses have come to view the ability to lease their required capacity from either a wholesale provider or hosting firm as an easier way to obtain the space they need. By leasing their data center space, companies avoid the need to use their own capital and are able to use their operational (OpEx) budgets to fund their data center requirements. By using this OpEx approach, the customer is able to budget for the expenses spelled out within their lease in the annual operations budget.

The other major consideration that customers must take into account in making their build-versus-lease decision is the timetable for the delivery of the data center. Building a data center can typically take 18–24 months (and often longer) to complete, while most wholesale providers or hosting companies can have their space ready for occupancy in 6 months or less.

3.7.7 The Challenges of Build or Buy

The decision to lease or own a data center has long-term consequences that customers should consider. In a leased environment, a number of costs that would normally be associated with owning a data center are included in the monthly lease rate. For example, in a leased environment, the customer does not incur the expense of the facility's operational or security personnel. The maintenance, both interior and exterior, of the site is also included in the lease rate. Perhaps most importantly, the customer is not responsible for the costs associated with the need to replace expensive items like generators or UPS systems. In short, in a leased environment, the customer is relieved of the responsibility for the operation and maintenance of the facility itself. They are only responsible for the support of the applications that they are running within their leased space.

While the cost and operational benefits of leasing a data center space are attractive, many customers still choose to own their own facilities for a variety of reasons that may best be categorized under the term "flexibility."

For all of the benefits found within a leased offering, some companies find that the very attributes that make these cost-effective solutions are too restrictive for their needs. In many instances, businesses, based on their experiences or corporate policies, find that their requirements cannot be addressed by prospective wholesale or hosting companies. In order to successfully implement their business models, wholesale or hosting providers cannot vary their offerings to use customer-specified vendors, customize their data center designs, or change their operational procedures. This vendor-imposed "inflexibility" therefore can be an insurmountable obstacle to businesses with very specific requirements.

3.8 FUTURE TRENDS

The need for data centers shows no signs of abating in the next 5–10 years. The amount of data generated on a daily basis and the user's desire to have instantaneous access to it will continue to drive requirements for more computing hardware for the data centers to store it in. With the proliferation of new technologies like cloud computing and Big Data, combined with a recognized lack of space, it is obvious that demand will continue to outpace supply.

This supply and demand imbalance has fostered the continuing entry of new firms into both the wholesale and hosting provider marketplace to offer customers a variety of options to address their data center requirements. Through the use of standardized designs and advanced building technologies, the industry can expect to see continued downward cost pressure on the providers themselves if they are to continue to offer competitive solutions for end users. Another result of the combined effects of innovations in design and technology will be an increasing desire on the part of end customers to have their data centers located where they need them. This will reflect a movement away from large data centers being built only in major metropolitan areas to meet the needs of provider's business models to a more customer-centric approach in which new data centers are designed, built, and delivered to customer-specified locations with factory-like precision. As a result, we shall see not only a proliferation of new data centers over the next decade but their location in historically nontraditional locations as well.

This proliferation of options, coupled with continually more aggressive cost reduction, will also precipitate a continued decline in the number of organizations electing to build their own data centers. Building a new facility will simply become too complex and expensive an option for businesses to pursue.

3.9 CONCLUSION

The data center industry is young and in the process of an extended growth phase. This period of continued innovation and competition will provide end customers with significant benefits in terms of cost, flexibility, and control. What will not change during this period, however, is the need for potential customers to continue to use the fundamental concepts outlined in this chapter during their evaluation processes and in making their final decisions. Stability in terms of a provider's ability to deliver reliable long-term solutions will continue to be the primary criteria for vendor evaluation and selection.

FURTHER READING

Crosby C. The Ergonomic Data Center: Save Us from Ourselves in Data Center Knowledge. Available at http://www.datacenterknowledge.com/archives/2014/03/05/ergonomic-data-center-save-us/. Accessed on March 5, 2014.

Crosby C. The DCIM Closed Door Policy in Mission Critical Magazine. Available at http://www.missioncriticalmagazine.com/taxonomies/2719-unconventional-wisdom. Accessed on August 13, 2014.

Crosby C. Questions to Ask in Your RFP in Mission Critical Magazine. Available at http://www.missioncriticalmagazine.com/articles/86060-questions-to-ask-in-your-rfp. Accessed on December 3, 2014.

Crosby C, Godrich K. Data Center Commissioning and the Myth of the Phased Build in Data Center Journal. Available at http://cp.revolio.com/i/148754. Accessed on August 2013.

SOURCES FOR DATA CENTER INDUSTRY NEWS AND TRENDS

Data Center Knowledge. Available at www.datacenterknowledge.com. Accessed on June 11, 2014.

Mission Critical Magazine. Available at www.missioncriticalmagazine.com. Accessed on June 11, 2014.

WHIR (Web Host Industry Review). Available at www.thewhir.com. Accessed on June 11, 2014.

4

MODULAR DATA CENTERS: DESIGN, DEPLOYMENT, AND OTHER CONSIDERATIONS

WADE VINSON[1], MATT SLABY[1], AND IAN LEVINE[2]
[1]Hewlett-Packard Company, Houston, TX, USA
[2]Hewlett-Packard Company, Albany, NY, USA

Within the data center (DC) construction market, analysts have recently described a category called Modular data centers (MDCs). In 2012, this type of build represented less than 5% of the total market, but it has been projected to almost double every year for the next several years. This chapter will describe Modular Data Center definition, design, and deployments, along with other consideration of similarities and differences with more traditional data centers.

4.1 MODULAR DATA CENTER DEFINITION

A critical understanding of the definition of a Modular Data Center (MDC) is that it can have no compromises with respect to a standard data center. Everything else in this handbook must apply with respect to usability, reliability, life expectancy, and the primary purpose to house, secure, power, cool, and provide connectivity for any type of IT device that would need to be deployed. The defining characteristic of an MDC is that when compared to a monolithic "brick and mortar" data center, some or all of the MDC components are not hand built on-site out of a collection of parts but are delivered as factory-built pre-engineered, preassembled, and pretested modules. Many industry experts liken the comparison of benefits to an automated assembly line versus hand-built automobiles. This allows a potential saving of time and cost, and allows optimal scalability to deliver just the right amount of data center needed to support the IT scalability needs. Neither method inherently produces a final end product of higher quality or more feature rich; it is simply the difference in the cost and time incurred during the construction phase of a data center. Once turned over to operations, it's just "a data center."

4.2 MDC BENEFITS AND APPLICATIONS

Using preassembled modules to build out a data center can save a company's assets of people, time, power, and money. These benefits most often occur when the MDC is aligned to the size, timetable, and reliability level of the IT it is built to support. This can be true whether deployed as single non-redundant modules of 50 or 200 kW, 2*N* power and cooled modules of 300 kW and 1.5 MW, and even aggregating these modules into 20 MW+ enterprise class data centers.

The lesson of manufacturing is that specialization allows paying the least amount for the right *people* skill set to get the product at the needed time. Demonstrated at the MDC, there is less need to have multiple of the highest level architects and engineers on site and in design and construction reviews, supported by dozens of master electrical and mechanical technicians to guarantee quality and performance, when all of it has already been done multiple times in a factory-quality environment. This allows the end client or their construction representative to need fewer people and overall lower total personnel costs.

MDC components are pre-engineered and preassembled at their source so the client does not manage the *schedule* for all the subsidiary elements in those modules. Scheduling concerns based on the availability of raw steel, copper, engine blocks, compressors, electrical components, and the like, including skilled labor for each, are removed from the client's responsibility. Not having to preplan around these interdependent schedules can allow time savings of more than a year, including the ability to even push out the start of the construction process to more closely align with future business needs.

Because module manufacturing versus on-site construction uses less personnel assets, and because all of the material

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

FIGURE 4.1 Example of an MDC. Each of the *white* and *black boxes* behind the office space is a manufactured module, and some are totally preassembled. Any one white IT module with its integrated cooling is powered by the four black modules, providing continuous redundant power. That row of five boxes is one 500–1000 kW complete MDC. Courtesy of Hewlett-Packard Company.

is managed in volume versus on a case-by-case time critical basis, the *total cost* to deliver a 2 MW power infrastructure or 1000 ton cooling system can be lower with MDC methods. Manufacturing also allows repeatable tested quality of larger assemblies in a validated process for lower costs and less time compared to quality assurance of the thousands of individual parts that make up an on-site constructed business critical data center.

The data center construction schedule dictates that the payment schedule and construction costs accrue from survey and ground breaking, through various serial phases of completion, all the way up to commissioning and turnover. MDC costs accrue based on delivery of modules and are often contingent on successful commissioning of the final collection of the modules so the client capital costs are much closer to the actual business productivity. And while the modular philosophy can scale to the largest 20 MW+DC projects, they can also build out complete DCs at the 200, 500, and 1000 kW scale (Fig. 4.1). When the smaller scales, shorter schedules, and lower costs are combined, many businesses find that a data center project can fall under lower corporate approval levels and allow data center projects to more closely align to a business region or function versus previously having to rely on only the fewer corporate mega data centers.

The *Power* savings due to energy efficiencies gained by properly sizing power and cooling equipment at the highest and most efficient utilization rates for power and cooling is another benefit of an MDC. While some MDC modules are used as construction elements, it is the complete 500 kW module running at 400 kW that benefits from greater than 90% power efficiency trains and sub 1.2 p PUEs[1] from compressor and fan cooling efficiencies. Conversely, 400 kW running in an open hall commissioned data center space built for 2.4 MW simply cannot be as energy efficient.

The primary purpose of data centers is to house the IT for the company, whether the IT is used to develop better products, serve the company's employees, or be the source of revenue for online business. The data center that best *aligns with the IT* need is going to be a superior solution. By being cost-effective and more energy efficient, this makes the MDC lower overhead on the IT. By being more predictable in schedule, with the potential of using inventory-held modules, the purchase and commissioning can occur closer to the IT need, with the up-front costs borne by the MDC vendor at their manufacturing site. This time alignment with IT can allow a client to push out decisions further and build the best DC that fits the unpredictable nature of the IT capacity and tier need. And ultimately, this allows the cost of the DC to conform to the IT requirements.

One of the largest benefits of the MDC comes from the application that *co-delivers IT* and *facility*. Traditionally, data centers had been described as "total square area of raised floor space" or defined in capacity by "power per area" metrics. This is shifting to a "power per rack" and

[1]http://www.thegreengrid.org/~/media/TechForumPresentations2011/Data_Center_Efficiency_Metrics_2011.pdf?lang=en

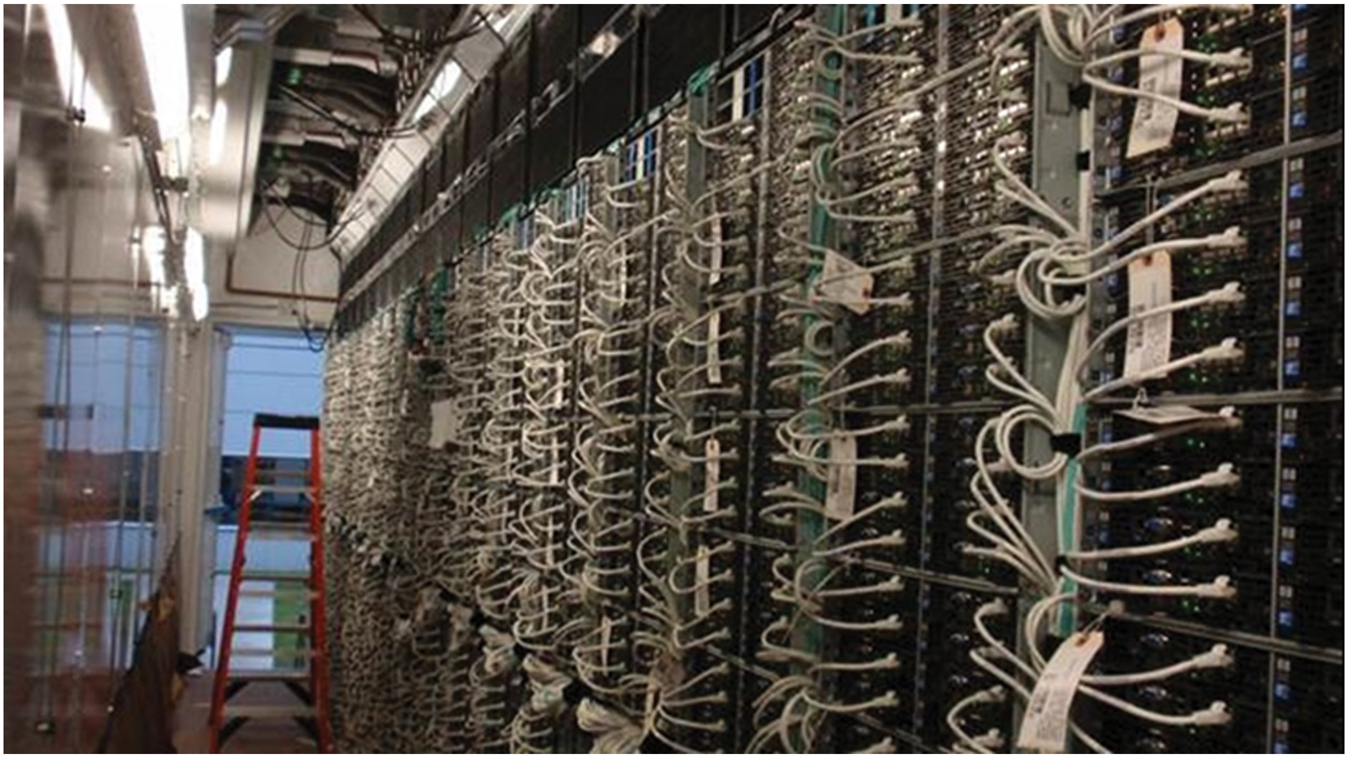


FIGURE 4.2 An MDC being configured and pretested with IT. This example shows twenty 18 kW racks all with their own UPS. This IT module can be fed by site chilled water, or is supported by another module specifically configured to deliver chilled water, or simply condenser temperature water for additional energy savings. Courtesy of Hewlett-Packard Company.

total rack availability to best define IT needs. Thus, an MDC might be optimized as a design that allows IT productivity of six 8 kW racks 6–12 weeks from permitting. (Fig. 4.2 is an example of an MDC factory preconfigured with IT.) In the past, a client typically tried squeezing that into an already full data center, or to build out a small computer room. While possible for small increases in IT capacity once an IT need grows to 10 racks or 200 kW, that approach is unsustainable. MDCs are available at those small total power levels but can be of lower cost at the 300 kW to 1 MW level due to complete utilization of a power circuit, common industry UPS sizes, or ease of procurement of single or redundant larger chillers. Another MDC use case is a collection of smaller IT module deployments repeatable until all of a 4000 A 480 V utility feed is utilized, or capacity is reached on a 3 MW back-up generator. While on-site-built data centers also are based on the size of these singular large components, MDC designs are divided to scale in size along any of the smaller elements that best align with the time and size of the IT. This scalability to natural boundaries of the utility supply is what allows the economics of many small modules to compete with the economy-of-scale associated with one large construction project.

While these examples of customization of an MDC to a particular IT deployment may limit the capacity of the MDC to be repurposed, it is often the uncertainty of the future IT that causes the highest costs and lead times for on-site-built data centers. Some MDCs can be designed so specific to the IT that they are treated like the IT from an expense and taxing purpose and could be disposed of with the IT at the end of the useful life to the client. Given the modular nature and size of MDCs, this could also easily allow the return of the module to the manufacturer to remove old IT and replace with the latest technology. This could occur several times over a 10- or 15-year lifecycle of a module.

While alternative applications could negate some of the benefits described in this section, once commissioned MDCs are identical to traditional DCs, thus eliminating the risk with investing in this alternative technology. It is for these reasons of size, cost, speed, and flexibility that the modular philosophy is being employed in certain aspects of even the most traditional on-site-built data center projects.

4.3 MODULARITY SCALABILITY PLANNING

Starting a construction project is a major undertaking for a business at any scale due to the number of people involved, the time from planning to operation, the external regulatory and nongovernmental influences, and the sheer size of the capital spend. Building one larger monolithic data center versus multiple smaller projects has been one way companies

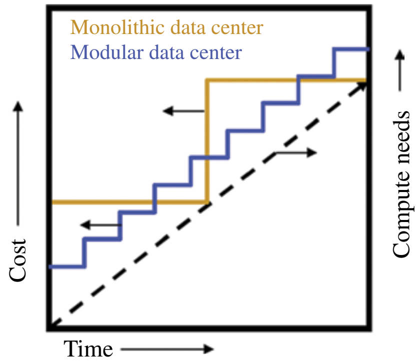


FIGURE 4.3 MDC capacity, traditional DC capacity, and IT capacity needs over time compared to costs. Courtesy of Hewlett-Packard Company.

have sought to reduce the cost per watt. The motivation for modularity has been to help scale the data center capacity and construction just in conjunction with the IT need. Conventional practices involve building in big chunks with large on-site-constructed data centers where the data center may take time to fill with IT. MDC provides a "pay-as-you-go" solution for the power, cooling, and building needed to support the compute capacity.

Figure 4.3 compares the cost of a large on-site construction project, often called a monolithic data center approach (solid lighter line), to a modular approach (solid darker line). The dashed line shows a typical linear need for IT over time. The IT demand line could have any slope or even multiple dips and peaks. An MDC approach can lower start-up costs since the requirements of the large capacity monolithic infrastructure does not burden the growing IT need of the dashed black line. Instead of building a large data center to insure adequate space for future growth of the IT operations, the MDC allows the costs to scale with the growing computing needs. The graph shows the ultimate cost of an MDC may be higher cost at the tail-end of the IT investment but depending on the timeframe for the build-out and the time value of money for that time, the ultimate cost may in fact be less. For example, a business that will completely fill its data center within 1 year would likely just consider a monolithic DC since the present value cost would be lower.

If the business does not expect to fill its data center for 10 years, the present value of the higher total cost at the end of 10 years is overshadowed by the savings in the first year. The typical space, power, and cooling forecast for brick and mortar is 5–10 years; it is a lot easier to have an accurate 12–36-month forecast. After more than 36 months, there is likely to be a disruptive technology change leaving stranded, insufficient, or inappropriate capacity in a brick and mortar design with 2/3 of the implementation's planned life left to go. The cost of the retrofitting needed to support the new IT can be higher than the first costs incurred to build the site. Alternatively, the broader adoption of MDCs is driving competition and optimization into the market place and lowering the cost of all the components. MDCs can have lower costs over the life of the project on any time scale.

Optimizing the total cost of ownership (TCO) of an MDC then becomes much easier compared to a monolithic build since the capacities for the IT are better aligned. A client starts by defining how much power and how many racks they will use over 12–36 months. This plan then assists with designing to the optimal size and duration of each step. More optimization occurs if other details are defined like the smallest and largest modules they want, at the right density and the tier level for the specific IT. The right lead times for any of these different modules should also be considered. This list gives a few examples of alternative designs and requirements:

- 10 MW needed over 36 months or 500 kW needed over 36 months;
- 25% at tier 1, 25% at fault tolerant, 50% at enterprise concurrently maintainable or other mix. This can lead to one type of power or cooling module supporting different IT modules;
- 16 week lead time for 1000 kW of 10 kW racks or 10 week lead time for 300 kW of 20 kW racks. This is often based on whether the IT is HPC, storage, virtual/clouds, or departmental;
- 40% predelivered with IT; 60% white space; 20% non-recyclable; 80% multi-IT lifecycle;
- 50 kW, 100 kW, 300 kW, or 1 MW IT need or aligned to 4000A 480 V circuit, 3 MW generator, or other easy to buy and install block like a 500 ton chiller;
- 30% renewable energy content; 50% at sub 1.4 PUE; or other environmental metric.

All of the earlier examples show that better planning is key to realizing the value of modularity in supporting the "pay-as-you-grow" philosophy.

4.4 MDC ANATOMY

While the future market likely has multiple premanufactured choices for building out full or partial MDC designs, this subchapter will define common features of many of the designs to assist in evaluating the elements that need to be considered for any one particular set of choices.

The most common set of deployment constraints concern the shipping logistics of cranes, trucks, and ships to get the 1 or 20 components that make up an MDC from its shell builder, to power/cooling/control integrator, to IT integrator, and finally to deploy at client's site. These constraints may apply beyond initial build-out since one benefit of MDCs is they may be moved or repurposed over the 3–20-year life depending on each type of module.

Every complete MDC must have very strictly defined elements of power and cooling because this directly relates to the service-level availability of the supported IT. Optimizing and right-sizing these based on the IT service-level agreement (SLA) can contribute more than 90% of the TCO of the whole DC project. Properly sizing the design is important to the utility modules and equipment internal to the IT module itself. These definitions must also include required initial commission, maintenance and support preventive, break fix, and life cycle servicing.

All modules must have defined network connectivity from any one rack or collection of racks out to the telco (telephone company) demarcation point (MPE or main point of entry). IT management and facility management DCIM (data center infrastructure management), BMS (building management system), and EPMS (electrical power management system) must also be architected and documented. This includes security, life safety and fire plans, conformance to the local and national codes, and other governmental or company insurability policies.

Even lights-out data centers still have people going into them and are built to keep weather out of them. The customer must understand how both are achieved, as well as codes associated with those attributes since they can vary across supplier's MDC offerings.

4.4.1 Structure of MDC (ISO Container)

The following are three different MDC configurations. Each of these has different approaches to meet the elements required to make up the proper anatomy of an MDC. While these three examples are not meant to be exhaustive of all the approaches, they have met with both market and technology success and so are representative of the choices across the industry today.

Figure 4.4 shows the physical and power layout of a small nonredundant DC at high density deployed indoors or outdoors that consists of one to four modules meant to be a standalone DC or as part of a larger DC project.

Figures 4.5 and 4.6 represent a four times larger, enterprise-redundant, medium-density MDC. It can be placed indoors or outdoors, deployed stand alone or as one module repeated and scaled ad infinitum. This MDC can be as small as two modules or more complex as needed.

Figure 4.7 is like the previous design, except that the IT space is a "building" that includes provisions for a network operations center and people space. It can be built as 1–4 modules at varying density and with differing cooling technologies.

The Figure 4.4 example is based on how most of the industry first thought about MDCs: as ISO 669 intermodal sea–land shipping containers. While there have been dramatic changes since these were first available as private stock keeping units (SKUs) in the 1990s, to the public press associated with Sun, BlackBox, Google containers, Rackable IceCubes, and HP PODS, these modules are still some of the simplest, fastest, lowest investment ways to stand up factory-configured IT in as little as 6 weeks from a purchase order. Well over half the suppliers offer variants of this nature.

The layout shows the elements that make up this style 50 kW to 1 MW IT "container" data center. This example is outdoors in a security/privacy fenced area with the required clearances for a small generator inlet and outlet airflow, and separation of the chiller to isolate it from the generator exhaust heat. The dotted lines indicate power that could be overhead or underground. The modules could be left on trailers or even anchored seismically. The footprint is reducible by stacking the modules, but that limits serviceability. Combining some of the modules onto single trailers allows more preassembly or simpler logistics. It is common for the site to not have a generator or even UPS if the IT SLA doesn't need it. The chilled water may also come from a central plant and the power switchboard could also come from the building, making this MDC as small as one single 20 ft container. The power input in this example arrives at 480 V, but this could be fed straight from the MV substation to both the IT and cooling. While these modules are available from multiple sources, the layout shows how important the coordination between modules is. The power size of the inputs and outputs affect all the pieces. Total energy efficiency of the MDC will vary depending on if the UPS is flywheel versus lead acid batteries with double conversion or if the chiller is low lift and/or has an economizer section. Efficiency also depends on if the IT module allows low airflow, or even if the IT can be run continuously at 90 F or greater. This example uses a readily available air-cooled water chiller and the IT module has rear door, under floor or overhead water cooling fan coils. The ability for maximum energy efficiency is limited because the design doesn't use outdoor air, but it protects the IT equipment from environmental contaminants. With proper climate, a water-side economizer can provide latent cooling to the water loop. Inside IT containers, individual fan coils can have separate power feeds and maintenance shunts to prevent complete downtime from a single point of failure. These types of MDCs can have PUEs from 1.05–1.5 and first costs of $8 to 20 per Watt. The specific example cited from dirt to live IT listed for $2 to 3 M deployed cost based on the chosen technology choices listed earlier.

Given the baseline example, it is easy to start thinking of these modules as building blocks that can be interconnected to support the specific IT needs for capacity, time, and budget. Figure 4.5 takes the idea further and still complies with the transportation elements of ISO668, but with more IT flexibility and capacity by having an IT and cooling module preassembled onsite. The electrical one-line is true enterprise class with no single point of failure for 7 × 24 × 365 continuous

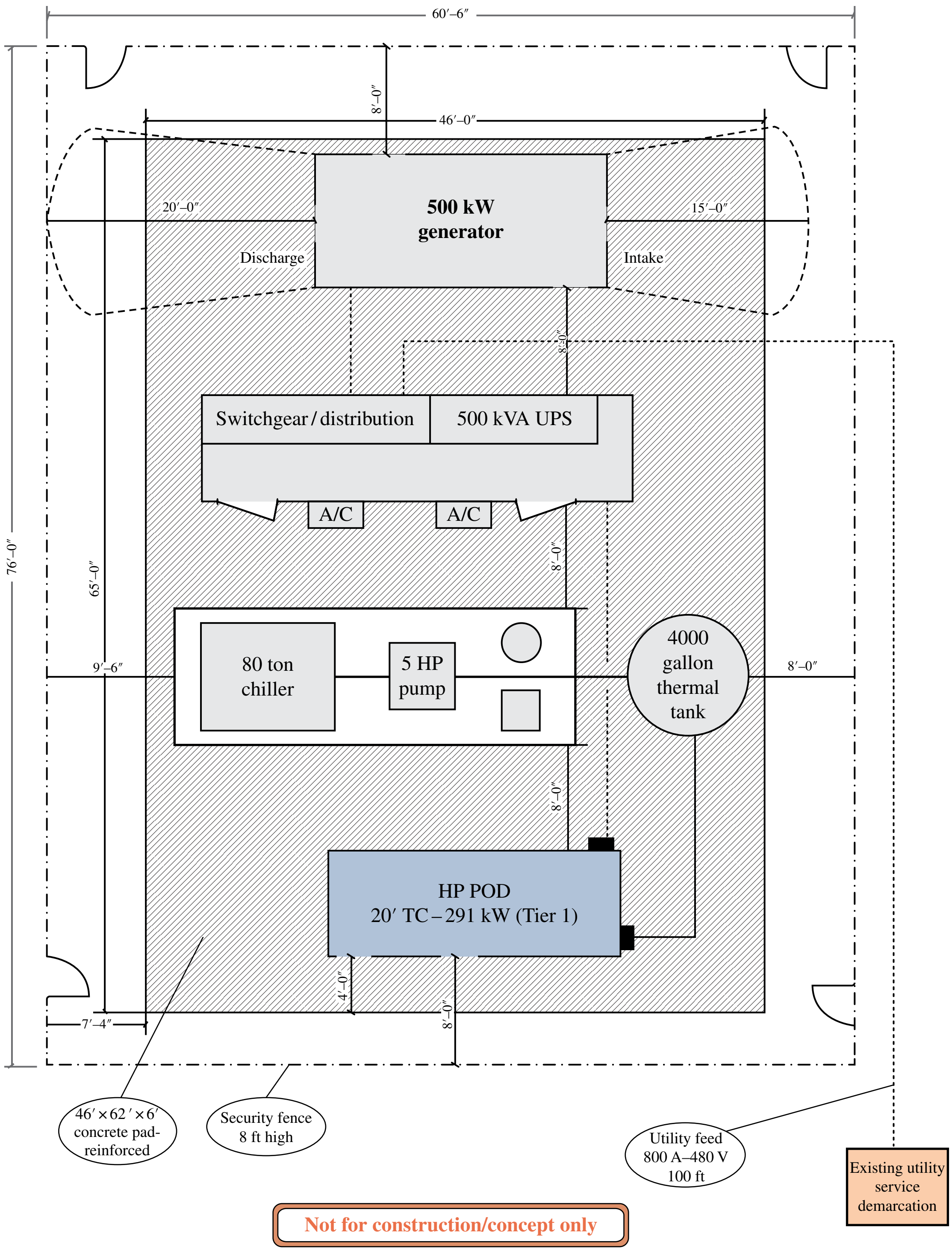


FIGURE 4.4 Layout of a complete small nonredundant high-density DC. Courtesy of Hewlett-Packard Company.

operation. It is important to note that while these seem block-like and many aspects can be pre-engineered, all of these require the right class of experts to get through the permitting and regulatory hurdles, commissioning, and assurances to make this arrive on schedule and on budget, delivering to enterprise SLA.

This example has two completely independent power paths. The power modules could be combined with multiple instances to have a shared single swing generator or UPSs reducing costs but having the same continuous uptime even if a failure occurred during a preventive maintenance event.

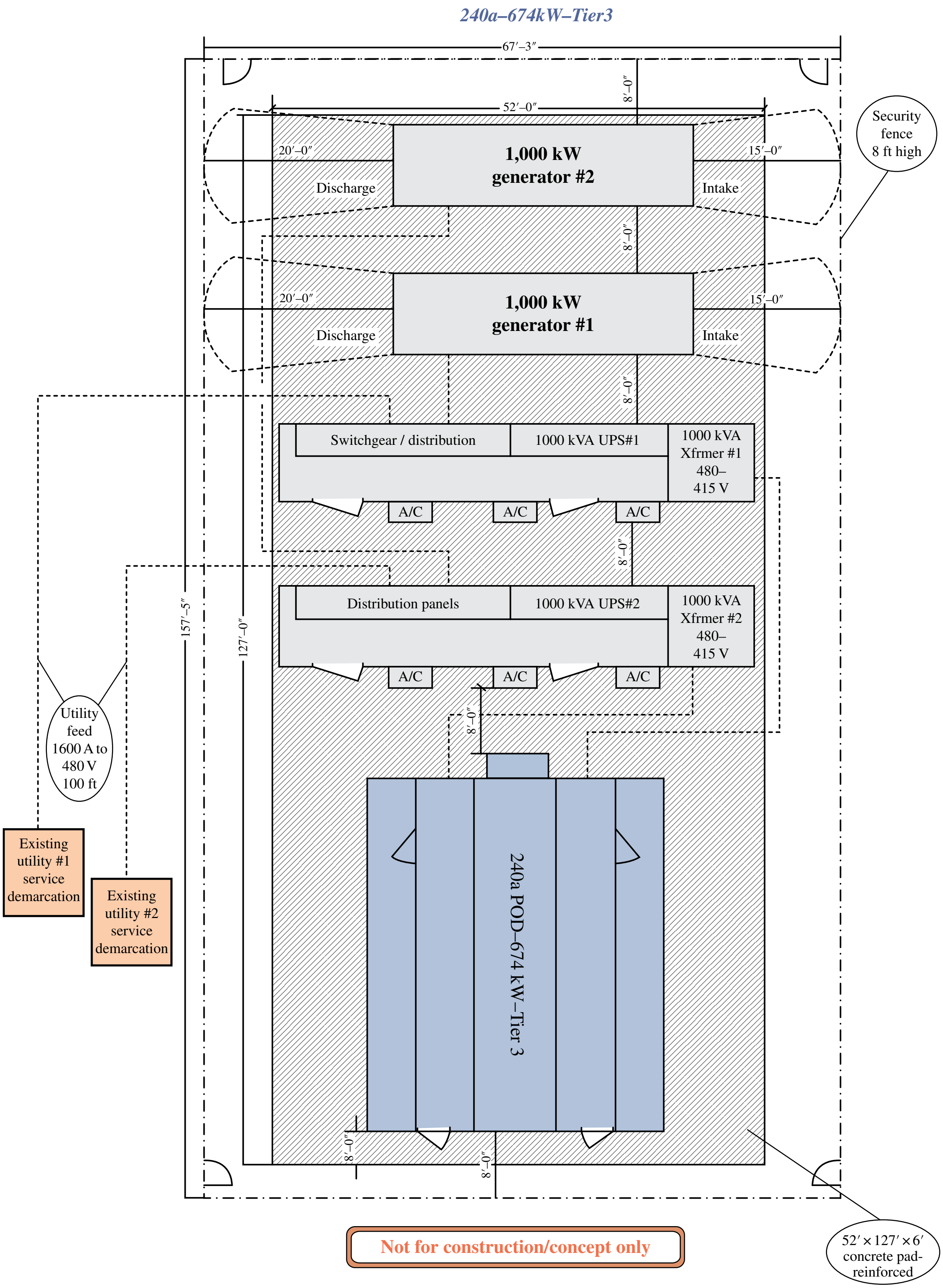


FIGURE 4.5 Layout for an enterprise-redundant, medium density MDC. Courtesy of Hewlett-Packard Company.

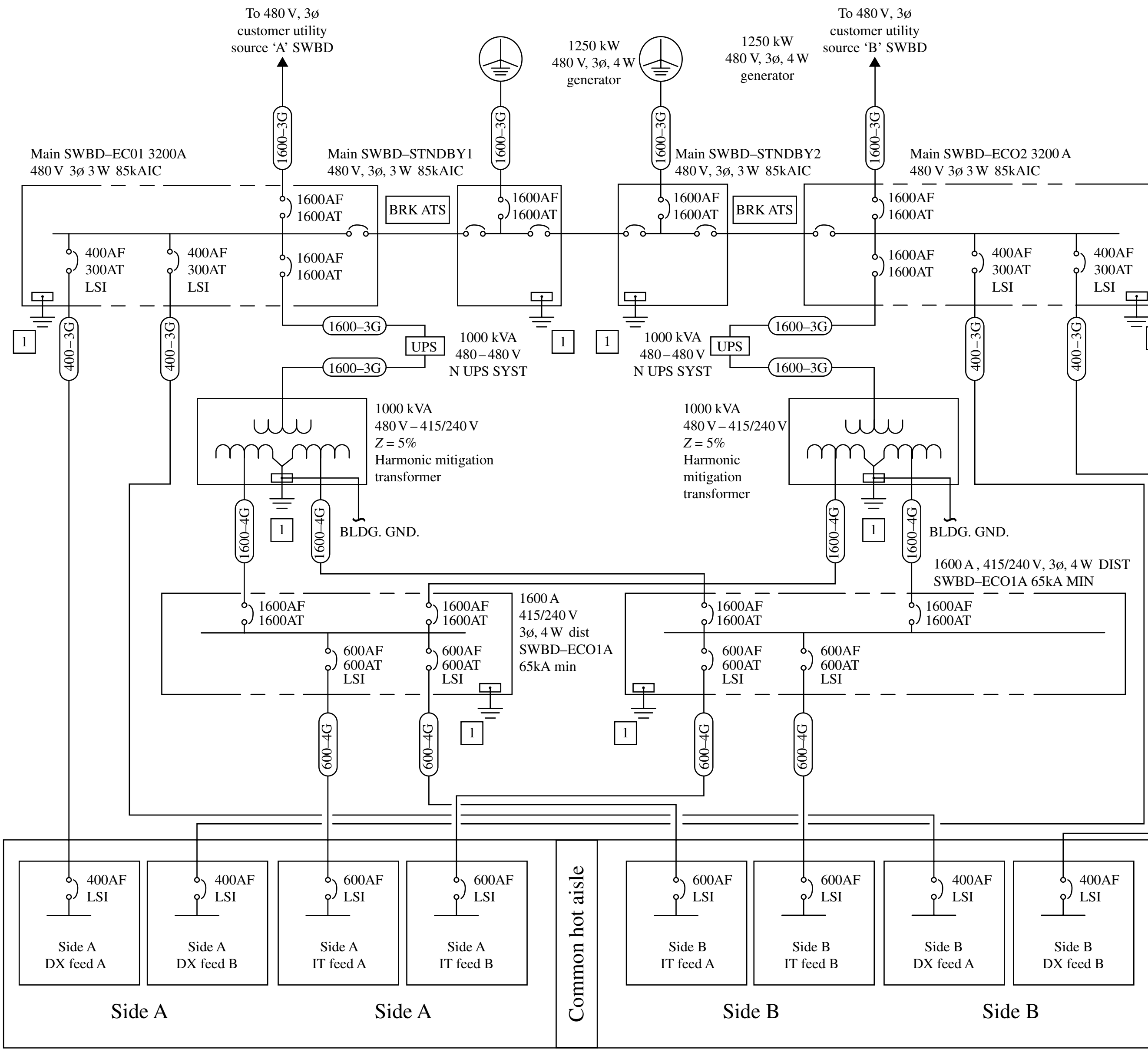


FIGURE 4.6 Power one-line for a high-density enterprise-redundant MDC with integrated cooling. Courtesy of Hewlett-Packard Company.

One significant difference in the power one-line in Figure 4.6 is that the 480V delta power for the cooler is part of the IT power module switch board even though the IT power is 415V/240V Wye. This example is for a US installation where the majority of IT power is run by 240V single phase power supplies, and the majority of compressors are 480V. Where 415V/240V power is the standard, these could be run from the same transformation. Conversely, it is possible that this MDC configuration could use adiabatic only coolers with no compressors with all power delivered as 415V location dependent. If the IT in the MDC is designed to run at 480V delta, then this could also be run from the single transformation. This illustration shows how the MDC architecture can vary to support different installations across the world, and matching the IT need and the site need, but also be flexible to align with the power and cooling specific to the IT.

This example resembles a more traditional data center with two cold aisles and a shared hot aisle, easily accessible and recognizable to service professionals. The module has the ability to be abutted to a building with a weather proof vestibule and security mantraps providing the same access and usability of a containment colocation. These MDC aisle access points could even be in the same facility that houses site-built colocation halls if this instance were adjoined to the building.

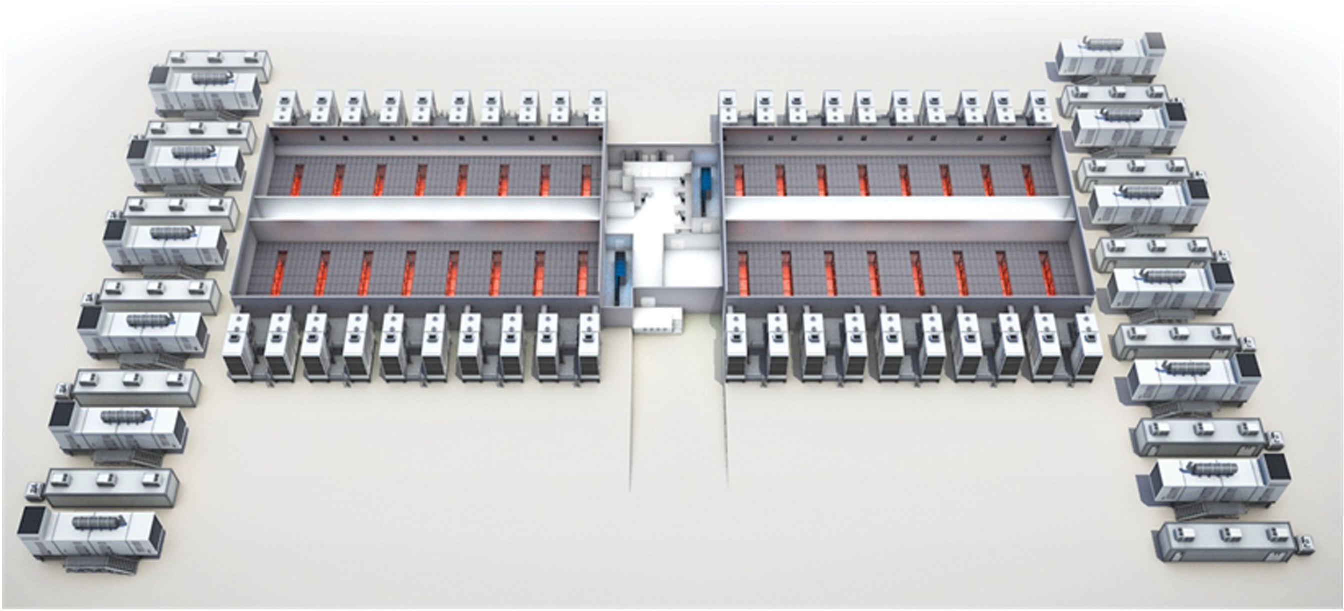


FIGURE 4.7 Layout for enterprise-redundant, air cooled building-like quadrant MDC. Courtesy of Hewlett-Packard Company.

The final example takes the modular concept closest to traditional site-built brick and mortar data center construction. Network operation centers, telco, demarcation rooms, offices, meeting rooms, lobbies, and restrooms can be delivered in a preassembled fashion or can be appended from tilt wall or cinder block appendages to the prebuilt modules. The example in Figure 4.7 shows a 3.2 MW site completely built out of four modules. With less-supporting coolers, generators, and switch boards in the same acreage, the total site could also be optimized for a lower destiny 1.6 MW. Each of the "quads" can be added over time as needed. Each quad can also have independence from a power and cooling perspective, and can each have their own density, cooling technology, and power topology tiering. Compared to the prior example with closed DX, air-side economization, adiabatic direct or indirect cooling on top of the IT aisles, this MDC has the units arrayed along the perimeter of the facility. Moving the air farther to get to the IT will use slightly more energy, but it allows more flexibility on how the IT rows are configured inside.

One benefit of this model is nonstandard, wider, or deeper IT racks and complex cabling requirements are more easily supported. In this example, the IT racks are rolled in later as needed, compared to the earlier examples where the IT could be preintegrated into the modules to reduce cost, improve quality, and reduce cycle time to live IT. Nothing prevents the type of IT modules used in the MDC examples in earlier Figures 4.4 and 4.5 from being deployed inside or as part of this facility either.

A typical example development of a complete larger MDC would be to start with the upper left quadrant example from Figure 4.7 to give the office space and site network room with 800 kW 2*N* capacity. The first racks delivered into that space arrive as two 200 kW skids from Figure 4.4 with their water cooling components removed. The remaining space and power in the quadrant is filled one rack at a time over the next 6–12 months.

In the second year, the second quadrant is developed at a lower tier level with no second utility feed but with the second IT bus run from a generator with a UPS. Later in the second year, the third quadrant is not built out but simply delivered with a 800 kW enterprise MDC as shown in Figure 4.5, that is half-filled with IT, and the remaining half is rolled in one rack at a time later. Finally in the third year, the fourth quadrant would also not be built out but instead arrive as a 1*N* adiabatically cooled 1.3 MW 44 30 kW rack HPC cluster like the one in Figure 4.5.

4.4.2 Architecture

Data centers exist to provide the needed support for the IT they house. These architectures provide cooling, power, and network for the IT, including controls for each and life safety for the people who work in the space. MDCs have all the same elements but are packaged in a way to provide preassembly and then final shipment. Figure 4.8 shows the external appearance of several MDC modules whose internal architecture contains these IT support systems. The upper left example is a 20-ft water-cooled container.

Internally, there may be hot aisle and cold aisle containment with the water-based coils above, below, in front of, or behind the racked IT coming from water supply and return mains. Similarly, power conductors supporting all

FIGURE 4.8 20-ft, 40-ft, and 40-ft double aisle IT, and 40-ft power MDC examples. Courtesy of Hewlett-Packard Company.

the racks come from a central switchboard and can also be located in a variety of locations. The upper right example is a 40-ft water-cooled container with the same type of power and cooling configurations. There are also network connections with cable ways running to each rack and again terminated at the module perimeter as a demarcation box or pass through. The third example is made up of two 40-ft IT aisles with a common hot aisle, but that architecture could also be reversed. The module conditions the outdoor air with overhead units, or the overhead units use the outdoor air to condition the recirculated internal air paths. In all three examples, the power can have metering and sense points and must have branch and main circuit breakers. Metering, sense, isolation, and shunt control is used on the water paths and air paths for the cooling control, including temperature, speed, and pressure for the fan coils in all three examples. The lower right example is a 40-ft module that transforms and delivers power from UPS, generators, and utility to the IT container and its cooling module.

All four examples have options for fire and smoke detection and notification, emergency power off, work, and safety lighting. An MDC can support fire suppression via internal gas-based discharge units or connections to site-based preaction dry pipe systems.

4.4.3 Power

Sections 4.1 and 4.2 have several detailed examples of how power in MDCs comes from the utility, is conditioned and backed up, and finally distributed all the way to the IT racks. These elements are also present in site-built data centers, and it is becoming increasingly more common that some of the power is delivered in a modular preassembled way on most construction projects. The widespread use of power modules in most data centers is allowing standard and custom offerings to be built in a variety of ways to accommodate site, SLA, and size specificity.

The features in these power modules allow choices across many parameters. Voltage can be delivered as medium, low, medium in and low out, ac in and dc out, and can be country specific. Many vendors even have SKUs that allow these choices to be made out of a bill of materials to allow them to have commonality of supply chain. Power generation and storage, like an uninterrupted power supply, is another criteria. Generation can be diesel (with options for storage), natural gas, biofuel, fuel cell, or optimized for alternative energy like wind or solar. Large providers like GE Natural Gas/Biofuel generators and BloomEnergy fuel cells both deliver their products in a modular fashion. This is also true with lead acid battery UPS suppliers

Schneider, Eaton and Emerson, and flywheel UPS suppliers ActivePower and Pillar.

These choices give customers many options for how and when they want to distribute power. Copper volume is reduced if medium voltage is run over longer distances and then the low voltage is right next to the IT. Power distribution can be paralleled to provide lower cost for redundancy, or it can be delivered with each IT module to allow tighter schedules and lower incremental costs. Cogeneration can be used as one of a $2N$ utility design. Coupling of UPS and generators can allow time on batteries to be as low as 15 s to reduce the size of the batteries needed. $2N$ IT bus distribution in the IT module with concurrent maintainability can allow for lower tier of the power sources feeding them. These choices also affect how the elements are serviced. Although regulatory codes dictate clearance around components, power modules can have some parts on external faces, or even be optimized if the known installation is indoor versus outdoor. Options can allow for $2(N+1)$ so faults can occur during maintenance events and keeping the IT is live; be designed for maintenance bypass to keep the IT live; or just be cost optimized to allow some or all of the IT to just be off during for service to reduce costs. In the market today power modules offer such flexibility in design and even multi-vendor sourcing strategy over a large range of power and tier capacities that their adoption will continue to grow.

Another element affecting the architecture of power modules is the power density trends occurring in IT. Many data centers will have a collection of each type of IT density, and MDC power module flexibility will allow the lower first costs and operating costs as they are optimized for each specific use case model. In storage, three classes are emerging, each having different power requirements. Enterprise business critical storage remains at under 10 kW per rack but needs to have $2N$ high availability. Big data storage used for analytics can be as high as 20 kW per rack and also needs to be $2N$. Data faults have to be rebuilt and the time it takes justifies the economics to have $2N$ power modules always available. The third class and growing trend is cold storage. This is best defined via the example of large collections of pictures or movies stored at a social networking sites that 90 days after distribution and initial view are hardly ever accessed again. This model can be as low as 2 kW/rack and only needs $1N$ power. This may represent 90% of the storage in a typical data center and having MDC power flexibility can offer dramatic savings opportunities.

On the computer application front, single CPU socket nodes and alternative processors like ARM are lowering the power for front-end application servers. In the past, this density was as high as 20 kW/rack, but could now be going down to 10 kW. However, IT designers continually figure out how to shrink these nodes, so having flexibility across that range is important. At the other end of the spectrum, high utilization of IT via virtualization can produce enterprise computing needs from today's 10 kW/rack to as high as 30 kW/rack. High-Performance Computing at governments, universities, and media studios, including graphic processors, have rack densities across entire rows of IT already running at 50–70 kW/rack averages. While most of these need $1N$ power, it has been common that more traditional enterprises like financial institutions are starting to develop code like this, and can start to push entire data centers up to 25 kW/rack averages with $2N$ power requirements.

IT flexibility and application flexibility in the SLA is also driving a change in how power provisioning occurs. In several large data centers, they are already setup to deliver $1N$ to some IT, $2N$ to others; $1N$ no generator but 5 min UPS to a third class, and $1N$ no UPS to a fourth class since those IT chassis have their own internal batteries. The trend will continue to progress toward IT and IT racks with their own batteries and even independent voltages. All of these examples show more power flexibility is needed and MDC designs are uniquely positioned to deliver exactly what is needed, where it is needed at the right time, price, and service level.

4.4.4 Cooling

MDC design allows for flexibility in designing the cooling subsystem to best match the needs of the IT based on density, service level, schedule, airflow requirements, and energy efficiency. The ideal IT to design cooling for would be one that needs no air or water movement to prevent it from overheating, such that the capital and operating expenses would be zero. Microsoft had demonstrated the concept by running HP Proliant DL580s outdoors in a tent in Tukwila, Washington, for a year with no ill effects. Enterprise data centers can't rely on that type of operation, but the modular concept can allow them to get the closest to that for the different types of IT they deploy. The goal of data center cooling is to keep the individual components internal to the IT and the IT racks as a whole within temperature tolerances as defined by the manufacturer specification or industry standard. The most common way of illustrating the range is the Psychrometric Chart. There are other texts that describe why and how this was developed, so here we will concentrate on using it to better define the choices MDC cooling modules are faced with. Any point on the chart fully defines the temperature and humidity (or enthalpy) of the air in the system. A region of the chart defines the allowable inlet temperatures and humidity for the IT components for proper operation. As servers take air in from the front and discharge at the rear, an MDC cooling module is set up to condition the IT entering air from either bringing out door air in or using the air coming from the server in a closed loop environment. Figure 4.9 shows how the air must be conditioned in order for it to be at the proper state to enter the IT rack.

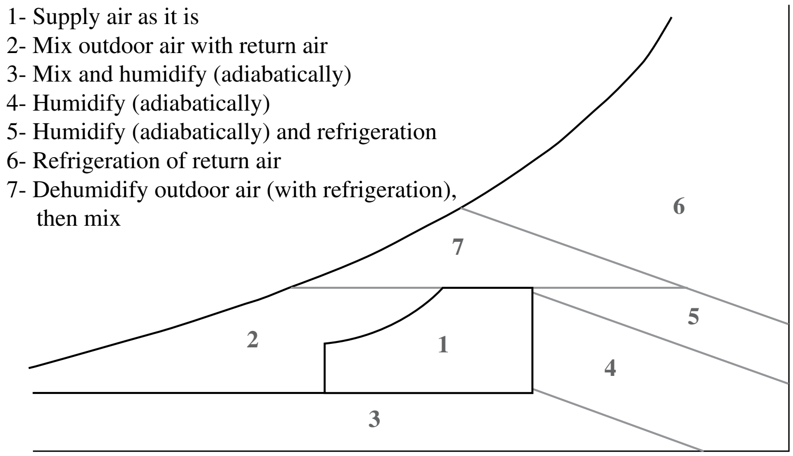


FIGURE 4.9 Psychrometric chart showing how air must be conditioned to make it acceptable to a define IT entering state. HP CFS Whitepaper "Global Data Centre Energy Strategy" 10-18-2009, Courtesy of Hewlett-Packard Company.

This example assumes the IT operating range is defined by the conditions of the air that is in the state defined by region or zone 1. Without describing a cooling module or technology yet, air from any of the numbered regions on the chart must be conditioned as described in order for it to be brought into a region 1 state. For example, if the outdoor air was very hot but not very humid as in region 4, just adding humidity to the air would bring it back into region 1. Whereas outdoor air that is not hot but too humid in zone 7 could be either mixed with return air from the server and passed over a dehumidifying cooling coil to condense the water out bringing air into zone 1.

Figure 4.10 is a specific psychrometric chart showing typical meteorological year data binned by hourly occurrence of the outdoor air state. Along the top are diagrams of how a cooling module might move the air from a state in one region in order to make it acceptable to the IT. The table at the lower left shows a summary of the hours by zone and the resultant pPUE of an MDC based on this accumulation of hours and application of the cooling technology. The small diagram to the upper left shows a top-mounted DX air conditioner allows filtered outdoor air to be drawn into the MDC and reach the IT unconditioned for all the hourly occurrences when that outdoor air is in zone 1. This is important because it eliminates running a compressor and saves a large amount of energy. The diagram in the middle shows how all the outdoor air in zone 2 h can be conditioned by mixing the air from the rear of the server directly with the outdoor air before it enters the IT. The inset diagram at the right illustrates how the module "close the window" when the outdoor air at the extremes via passing the air at the rear of the server over a refrigerated cooling coil before returning it to the front of the IT. Then the two diagrams in the middle describe how outdoor air can be conditioned if it's hot and dry (adiabatically) or if it's hot and humidity is not high (DX) before entering the server. This means that based on the conditions of the Los Angeles outdoor air, compressor-based DX conditioning use could be avoided by utilizing adiabatic humidifiers to cool the hot air for even more of the hours. This specific example is for an HP EcoPOD 240a placed in Los Angeles, California, USA. For 8421 h out of 8760 h in a typical year, only filtered or return air fan energy is needed when conditioning the air to what American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) defines as "A1 Allowable."

Because the IT was required to be at the ASHRAE allowable state, DX conditioners could not be eliminated completely due to the need to dry out the air via condensation on the cold coil. But Figure 4.11 shows that if the IT-operating envelope were expanded to allow slightly more humid air to enter the servers, which is actually acceptable for most of today's IT, then the DX units could be eliminated and the module would just use the adiabatic humidifying element to cool the air on the hottest days and hours.

From an energy usage perspective, the Los Angeles climate would be slightly better pPUE when the higher humidity standard is allowed. If this were Chicago, New York, or many other parts of the world, the pPUE difference between the three options could be much higher. The point here is that the climate is very important to the type of cooling used, and MDCs offer the ability to have common designs and then optimized cooling that is site specific.

Not shown in the pPUE data when looking at a city like Salt Lake City is the first cost difference that an adiabatic cooling unit can save versus DX compressors, evaporators, and condensing fans. Also not shown are the first cost savings associated with not having to install larger power feeds for the DX compressors. Last, there are even more first cost savings since the lower power adiabatic unit could be on the same UPS as the IT, whereas DX compressors would have to get their own UPSs, or the solution would need enough thermal mass to ride through until the generators started if there were no cooling UPS.

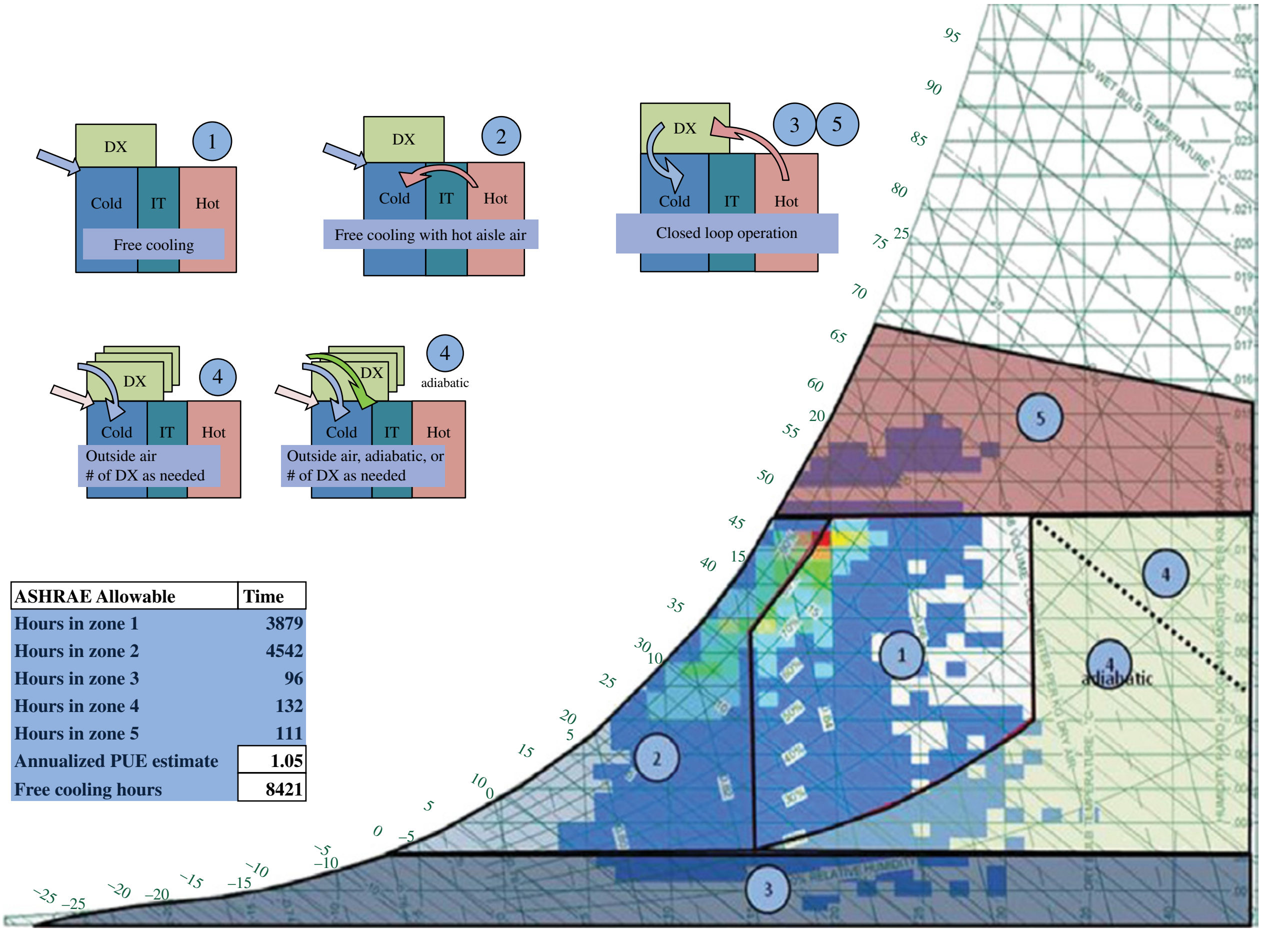


FIGURE 4.10 How a DX cooling module conditions IT air year round and how adding adiabatic cooling to the module allows less hours on DX. Courtesy of Hewlett-Packard Company.

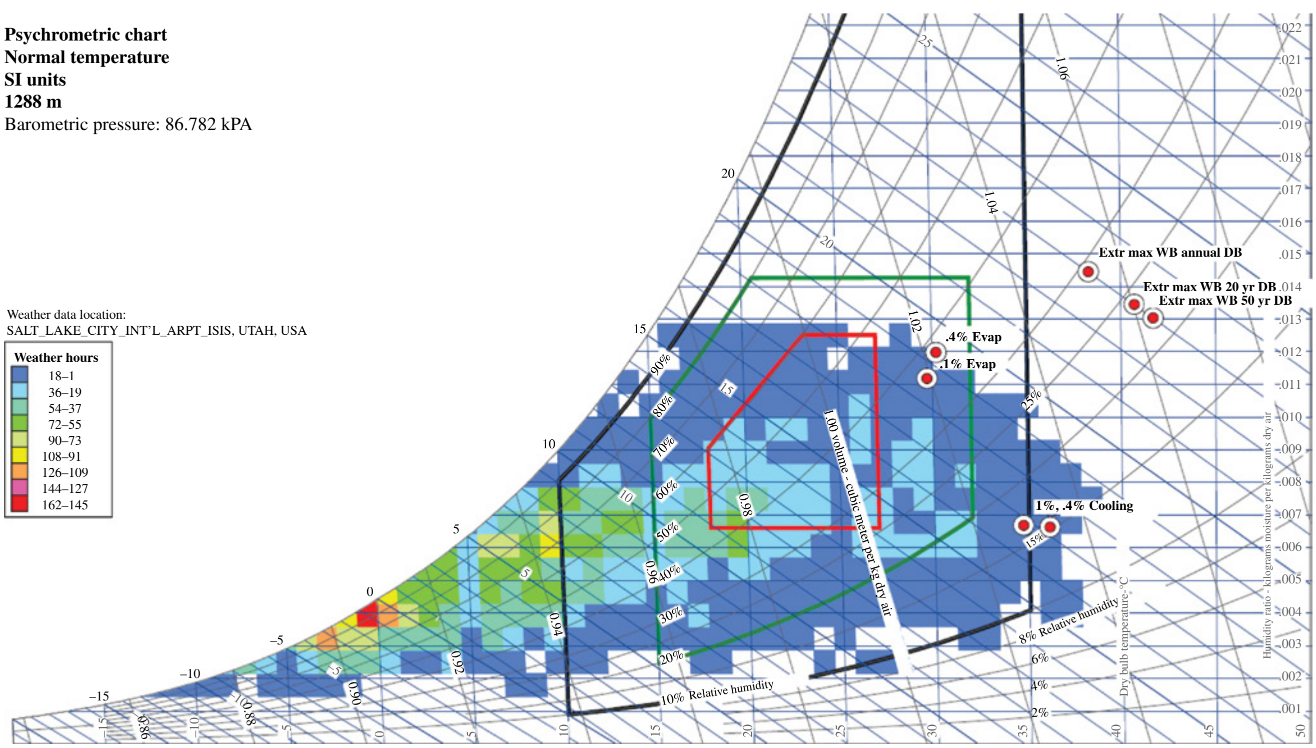


FIGURE 4.11 Cooling technology module choices must take into account extremes as well as typical. Courtesy of Hewlett-Packard Company.

Using site weather data is critical to determine the optimized cooling modules for the MDC based on first cost, energy usage, and service costs. Figure 4.11 shows that while the typical year data is a good indicator, attention must be paid to the extreme data points as well. In this example for Salt Lake City, if only the typical year data were used and the IT standard used the lower allowable humidity, then DX units could have been eliminated. But if the extreme data were considered and DXs were not installed, then the IT would be out of spec on numerous occasions. In this particular application adiabatic was used in conjunction with higher humidity IT specification and thus even the extreme data points could be kept in spec.

The type of IT cooling technology used in MDCs is very flexible. All of the cooling modules shown earlier described the use of outdoor air, typically called "air-side economization." While site-built data centers also use the same type of technology, this is more common with MDCs because the IT is physically closer to the outside air. In a large hall data center, or a data center built on site as part of a facility, not all of the IT can be located in this way. More fan energy would be needed to move the air to all of the IT. The earlier examples used direct expansion refrigerant in the cooling coils as the primary heat rejection means, once the hot air from the IT racks was pulled over it. Care must be taken in the refrigerant used in these coils as they are susceptible to changing regulations. Water is also commonly used in the cooling coils as a transport mechanism to cool the air. That water is then typically recooled using a refrigerant process. Based on the outdoor air state as described by the psychrometric chart, this cooling water can be cooled in certain location without the use of refrigerants. This process is called "water-side economization." One way to avoid the excess use of fan energy in site-built data center halls and MDCs is to locate the DX or water cooling coils close to the IT, then use the 1000× larger heat carrying capacity of these fluids to transport the heat versus trying to move the same heat via air and fans. Another way to reduce the cost of cooling is to evaporate water into the air via an adiabatic process. In a water-cooled coil, using this method for the ultimate heat rejection is called a "cooling tower." In the air cooled examples, these can be referred to as "swamp" or "adiabatic coolers." The key in all of these is cooling can be done free of the energy used by compressors, so all of these technologies are called "free cooling."

IT that was directly cooled via water was common in mainframe days, but most data centers eliminated water piping to the IT racks in the past decade. Water-cooled servers are again becoming more common in the high-performance computing space. The heat-carrying capacity of water versus air is not only allowing more efficient faster processors, but it is the only way of cooling the 50 kW/rack density and above that are becoming more prevalent. When the entire rack heat load is in water, then moving air can be eliminated completely. Compressors can also be eliminated because even the hottest climates can cool the water 20°C below the required case temperature[2] of today's processors. One other parameter driving water cooling is the greater ability to reuse the hot water, and thus lower the total energy cost of an IT campus. Adding water to buildings and into IT racks does not come without costs and risks. MDCs can have dedicated areas to provide for just this type of IT. The prebuilt aspect of the module and the IT allows for extensive testing to develop and the best configurations to optimize for reliability, using the lowest energy water for cooling, and allowing the hottest water return reuse application.

MDC cooling architectures have various approaches to deal with service levels required via different fault tolerances of the design when compared to first costs, power costs, and service costs. Fault tolerance of air cooling is supported by having $2N$ power to all the fans and $N+1$ or +2 quantity of fans in any fault domain. 1 N control signal to the fans and speed signal coming from the fans is acceptable as long as fail-safe is maintained if the signal is lost. In a water-cooled scenario, while $2N$ mains are achievable, more common is a single main fed from multiple sources with isolatable individual coils. The design can have $N+1$ coils and allow fault maintainability inside of any fault domain. This would also be true for water-cooled IT and the water used for adiabatically cooling the air. This same type of arrangement can also be done for the compressors. For continuous use during a fault, pumps and fans are often on UPS and generators just like the IT, whereas compressors are only on the generator and water storage or air thermal mass is used to keep IT in spec until the generators are online.

The efficiency of the cooling design has been a large driver for MDCs because it allows the easiest method to evaluate the TCO of the cooling needed for the IT. One example is modular cooling towers on the heat rejection side of a DX or water-cooled plant. These will always have higher first costs; but because they improve efficiency, it may allow less total cooling to be purchased. This has the ripple effect of lowing total power requirements for the cooling as well. Add to this the ability to build in redundancy via multiple modules that scale up only as needed to maintain peak efficiency, and it easy to see why modular coolers are even used on most large site-built data center construction projects.

Most MDCs are designed to be placed outdoors and are weatherized to NEMA3 or IP46. Being outdoors also means ensuring they are rated for the right wind and snow loads specific to the site. While insulation to prevent solar gain can lower the energy bills in a sunny hot environment, insulation is needed more for the winter. IT doesn't want to run below 50°F, so cold weather start-up procedures are required. Once the servers are running, the IT waste heat can keep the space conditioned, but then the

[2]http://www.intel.com/support/processors/sb/CS-034526.htm

primary concerns are too low or too high humidity. Common practice is to design the structure with thermal breaks so no cold metal outer member can conduct heat away from the interior air volume adjacent to the servers. Most MDCs are not 100% vapor tight, even those designed with gas-based fire suppression and to the highest weatherproof standards. This is because there are numerous access doors to the IT and air inlet/egress ports for free cooling. So, having the ability to add humidity in the dry winter is important if your IT standard requires it. Adding water for humidity in cold weather months introduces risk of condensation and frozen water lines, so some clients require IT that can be operated at 5% RH to forgo this expense. Too much humidity in the winter can also cause all of the coldest surfaces to form condensation, so having a desired place to condense water out of the air is also necessary. In the coldest climates, additional cost for more insulation and true vapor barrier may be necessary to eliminate condensation concerns.

Perhaps the largest driver for using MDC cooling modules is the ability to align the quantity of cooling to match the IT-specific cooling needs at the lowest possible costs. In a closed loop mode, this is the kW capacity heat rejection to the local environmental conditions based on per rack or per row capacity. In an air-cooled mode, this means airflow capacity per rack and per row. The architecture and controls of both allow airflow to be matched with variable speed fans to allow positive cold aisle pressurization. Even some low-density IT can have a high airflow requirement, so being able to spec, measure, and deliver airflow is critical. Temperature control is then achieved by metering in cooler outdoor air, bypassing more air over adiabatic media, over cooling water coils, or over DX coils. More of those coils or more air over those can be introduced into the system as the IT load goes up. But just like airflow, alignment with the IT is critical because too much adds first costs and lowers energy efficiency, and not having enough can shut down the entire IT fault domain.

Airflow separation with hot aisle and cold aisle containment is also more critical in the compact volume of MDCs. Rack blanking panels, inter-rack gaskets, rack-to-floor foam gaskets are all required to ensure the hot server exhaust air is properly conditioned before being returned to the server inlet. Some network equipment does not draw air front to back, so the exhaust air must also be properly ducted to not cause neighboring IT to overheat. The challenge is no more or less problematic than a site-built data center because airflow, thermal mass, and cooling capacity can also be overprovisioned for safety margins in the MDC too. More likely is that the use of the MDC makes it easier to understand the best cooling alignment and containment for the IT deployed, allowing the largest first cost and energy cost savings.

4.4.5 Control

The Controls architecture of MDCs have all of the elements that site-built data centers have. All of the power has main and branch circuit protection, down to the rack and plugs, including optional remote emergency power off. These can be monitorable with current and power transducers and electronically monitored as a data center EPMS function. All the cooling control points are visible to a BMS with monitoring and control points for temperature, humidity, pressure, airflow, water flow, valve position, and fan and pump speeds. Today's IT can have monitoring points, and these can be incorporated into a module's Environmental Control System. What is traditionally called DCIM is easier in an MDC because it can be all prebuilt and tested in the factory, and even be optimized with IT-specific interaction with the module power and cooling functionality because all are so closely coupled. This is demonstrated by the low total PUEs MDCs have achieved and the extremely high-density IT deployed.

Other control elements common to DCs and MDCs are VESDA smoke detectors and fire alarms and annunciators, both local and tied into large facilities. Fire suppression is easily designed and deployed because the interior volume is well known to make it easy to locate sprinkler heads or fill the volume with the right concentration for suppression.

The modular nature of the design benefits access control and security. Whether these are key-locked doors, entry detection devices, or biometric security, different modules can have different security levels based on the IT inside and even the type of employee doing work in one module versus another. The electrical technician can have access to the generator and switchboard module, the HVAC tech access to the cooling, and the network tech access to the IT, exclusive to one another.

The best MDCs take advantage of every cubic inch inside the shippable module to get the most IT, thus lowering the cost per deployment. There is rarely room left over to run extra wiring at the jobsite if it wasn't thought of up front. Every control wire, wire way, and conduit path must be fully defined and documented. Regulations or standards dictate that different voltages, security, fire, phone, controls, and monitoring wires all must have their own raceways, and that these raceways be maintainable and documented. This allows an MDC to behave the same as a site-built data center over the 15-year life span because somebody somewhere will need to access, service, or replace an element of control or other wiring, and the documentation and design must allow for that.

4.4.6 Redundancy and Reliability

The MDC must have the same availability options as those of any DC to ensure IT continuous functionality based on common standards. The Uptime Institute Tier ratings and TIA 942A Level ratings are both relevant to the IT support elements. Independent evaluation of the fault tolerance of the power, cooling, and control feeds from the street to the IT is required. Conceptually, the tradeoffs in the architecture choices can be viewed in these ways:

- Tier 1: Designs of this nature offer the opportunity to operate at higher capacities or to remove redundant infrastructure to reduce the total solution first expense. This can be true for some or all of the IT. Low-density cold storage and high-density render jobs benefit most by allowing this lower cost architecture.
- Tier 2-like: This is achieved by adding redundancy to the components that if failed would bring down the load—both internal to the IT module and in the power and cooling support modules. Consideration of these elements can bring operations up to four 9s reliability at a relative low cost and is increasing in popularity in the site-built and MDC space.
- Tier 3+: This design has $2N$ power to the IT and to the cooling, with everything fault tolerant and concurrently maintainable. While it could arguably be called tier 4, the compact nature of MDCs may not allow servicing of all path A elements with path B live. Most enterprise data center architects have determined a single MDC with this level of reliability meets all of their business needs for the applications housed in it.

A clear understanding of the tier-level requirements of an organization's IT infrastructure is required: that is, do not assume that everything is tier 3+ because potential savings are achieved in implementation and operation by matching the tier level with the real IT requirement. Aligning power and cooling tier to the IT capacity tier also eliminates sunk capital cost improving the short- and long-term TCO.

The modular preassembled nature of the architecture can also yield a redundancy strategy based on having two exact copies of the same thing at any one site or across multiple sites. For example, a large IT service provider could have an application stamp consisting of three compute racks, three storage racks, and a network rack. That gives them the ability to build two or more of these as low as tier 1 or tier 2, but get much higher availability by having software application redundancy across the two installations. The architecture could be further applied in an array fashion where perhaps one extra tier 1 container could give continuous availability to four others, further reducing a company's data center spend. This has less to do with a modular construction approach and more to do with the productization aspects of MDCs.

MDCs have also been positioned as part of a company's disaster recover strategy to enhance IT redundancy by having an identical stamp in a second location. The identical repeatable stamp nature of factory modules allows that. In addition, a multidata center company could afford to have a

FIGURE 4.12 Front-cabled racks being installed in an MDC at IT factory. Courtesy of Hewlett-Packard Company.

spare IT container in inventory that is deployable to a disaster site and connect to a rental chiller and rental generator in as little as a few days.

4.4.7 Network and Cables

IT servers connect to each other via cabling. That cabling is aggregated in each rack and then run from each rack to a core switch rack. From the core switch rack, it is run to the edge of the hall, where a demarcation panel then gives access to the network from the outside world. Site-built data centers design a defined location for each of these and may even prepopulate some of the trays before the IT is installed. MDCs may follow all of these practices; but as the cabling varies greatly with the IT, this wiring is usually defined specific to the IT installed.

Whether the MDC has 10, 20, or 40 racks, the racks have internal wiring that can be done at the IT supplier to enable total rack testing. That cabling can be at the front of the rack, at the back of the rack, or a combination of both. If these racks are moved into the MDC one at a time after cabling, care must be taken not to affect the cabling as it can be very sensitive to required bend radii. If this cabling is to be done on the blanked empty racks built into the MDC and the IT is installed later, then consideration for cable channels in between racks, and clearances to the front and back of the racks must be accounted for. Figure 4.12 shows a simple example of a front-cabled rack that has just been positioned in an MDC at the IT factory. This cabling must also allow for replacement of a strand and allow for the IT attached to it to be serviced at any time too. It is common for 500 cables aggregating to a pair of network switches, BMC aggregators, keyboard, video, and mouse (KVMs), patch panels, and power strips to be located internally to a rack. In the rack at the far right in Figure 4.12, the top of the rack has a patch panel installed and its network switch is at the 24th U. The second rack in the figure has the patch panels at the top of the rack with the network switches just below it. While network switches are called TORs (top of rack), putting them in the middle reduces cable bundle diameter. This matters because the larger bundle is more likely to be snagged in the aisle, and the fatter bundle means the aggregation has to be separated to service an individual cable.

Certain types of IT such as high-performance computing may aggregate across several racks into an InfiniBand or similar aggregator. This cabling requirement can result in as many as 1600 cables between eight racks, MDCs address this with large easily accessible cable trays positioned in the floor or overhead. Figure 4.13 is an example of a high-performance cluster in a water-cooled MDC that has more than 100 blade chassis collapsing onto a single switch. More common rack-to-rack aggregation occurs as the individual top of rack switches may have as few as 16 network cables running to the core switch rack. If there are 20 racks in the MDC, then the cable tray needs to support hundreds of cables up to 10 or 20 m in length. Figure 4.14 shows the back of an MDC aisle being prebuilt and prewired at the IT factory. The supports are removed once the aisle is joined onsite and the DX coolers are added to the top. The wires at the top of the racks go from rack to rack. This also shows how the MDC can be flexible enough to support rear-cabled blade chassis shown at the left, front cable scale out notes (not shown), and more traditional 1 U and 2 U servers like the racks at the right. With 2*N* power requirements the space gets very full similar to a site-built data center with this much dense IT and hot aisle containment.

The types of cables vary from single-strand fiber, CAT5 Ethernet, to 68 conductor storage cabling, so the wire ways internal to the racks, among the racks, and from the core rack to the outside demarcation must be considered. Since the primary purpose of an MDC is to run the IT, the signal integrity of this wiring must be maintained via proper spacing, bend radius, handling, and proper length connections from the core switch to the individual IT chassis spread

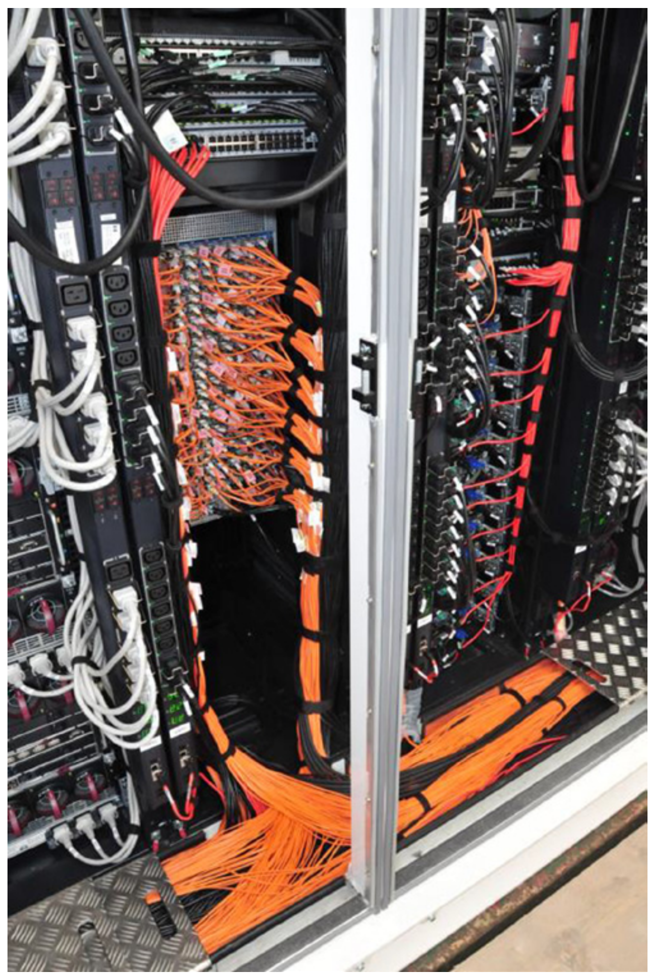


FIGURE 4.13 100 blade chassis collapsing onto a single switch. Courtesy of Hewlett-Packard Company.

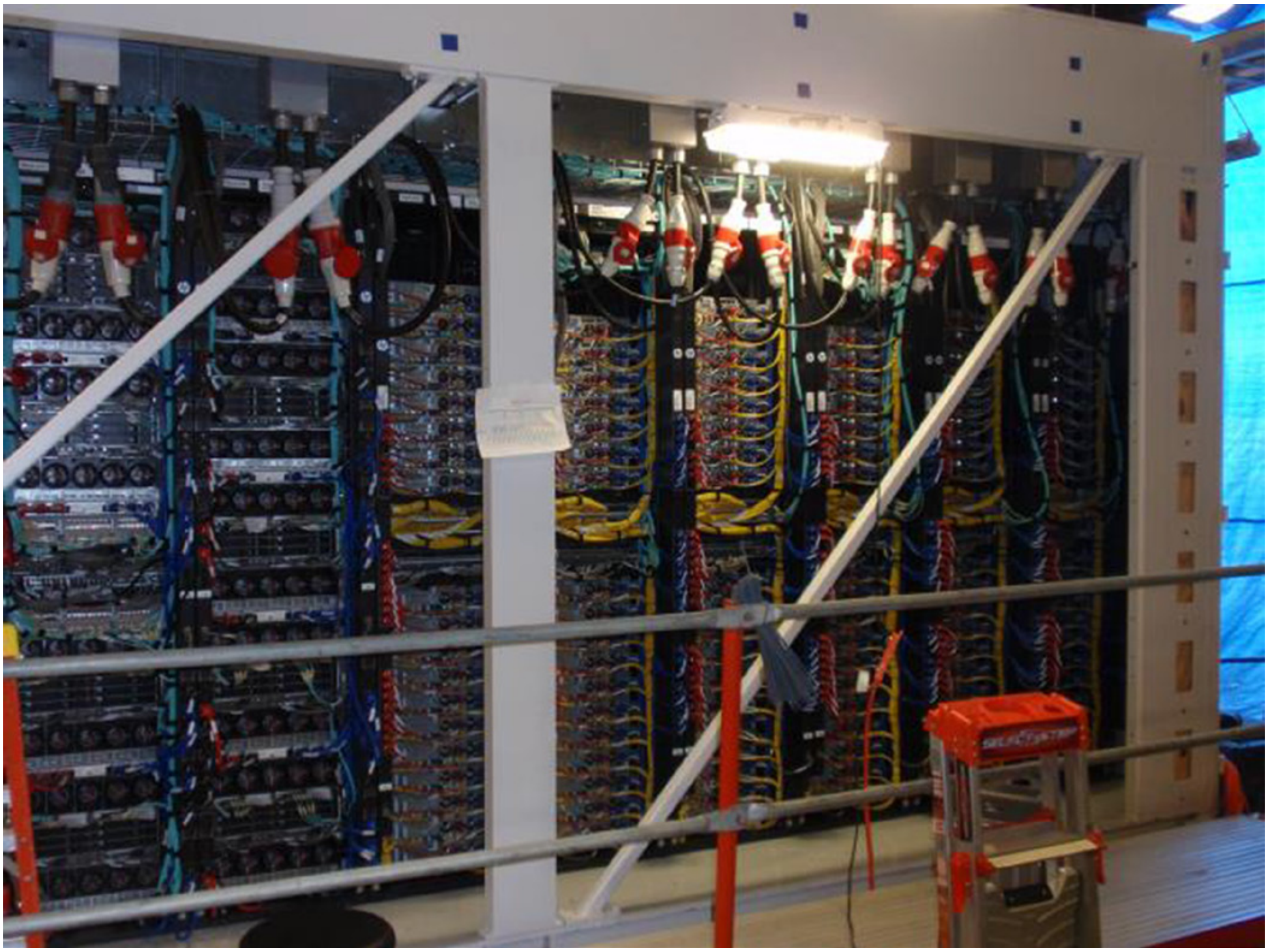


FIGURE 4.14 Inter-rack cabling across a diverse IT row in a double wide MDC. Courtesy of Hewlett-Packard Company. Courtesy of Hewlett-Packard Company.

across the MDC. It is important to understand the owner of routing design, schedule, and architect because this affects the construction company, the IT provider, and the network architecture, as well as the build and final test schedule. Figure 4.15 shows the same IT from Figure 4.14 while being configured for inter-rack cabling at the IT factory. With lots of skilled labor and a factory test process, the responsibility and schedule are well defined. If this work were completed onsite in the MDC aisle after it all arrived, the cost and schedule complexities would be greater. One way this complex wiring is more easily accomplished on-site is wider racks with integrated cable ways. Figure 4.16 shows an MDC with three wider centralized network racks to facilitate easier cabling, while the remaining 18 racks are normal.

Finally, all of this network must get to the outside world. In large site-built data centers, there are main distribution frames, intermediate distribution frames, and final demarcation rooms. The quadrant style MDC in Figure 4.7 has all of those features. Other MDCs just have simple pass-throughs or patch panels along the perimeter wall that are then connected to the local building intermediate distribution frame (IDF), or run straight through to the telco provider. Figure 4.17 is an example where the conduit came from the building-level networks to run straight from the site IDF, into the cable tray over the rack, then straight to the core switch rack.

This example is a double aisle adiabatically cooled MDC inside a warehouse. The network enters via the two pipes leading up from the central hot aisle door. Also in this figure at the far left is the power conduit for the left half and the controls demarcation box, with separate conduit for controls and life safety systems. The power conduits for the right IT aisle are the four metal pipes heading up at the far right. Insulated water supply and drain lines for the adiabatic coolers are installed as well.

FIGURE 4.15 Complex IT wiring could be much faster in an MDC factory environment. Courtesy of Hewlett-Packard Company.

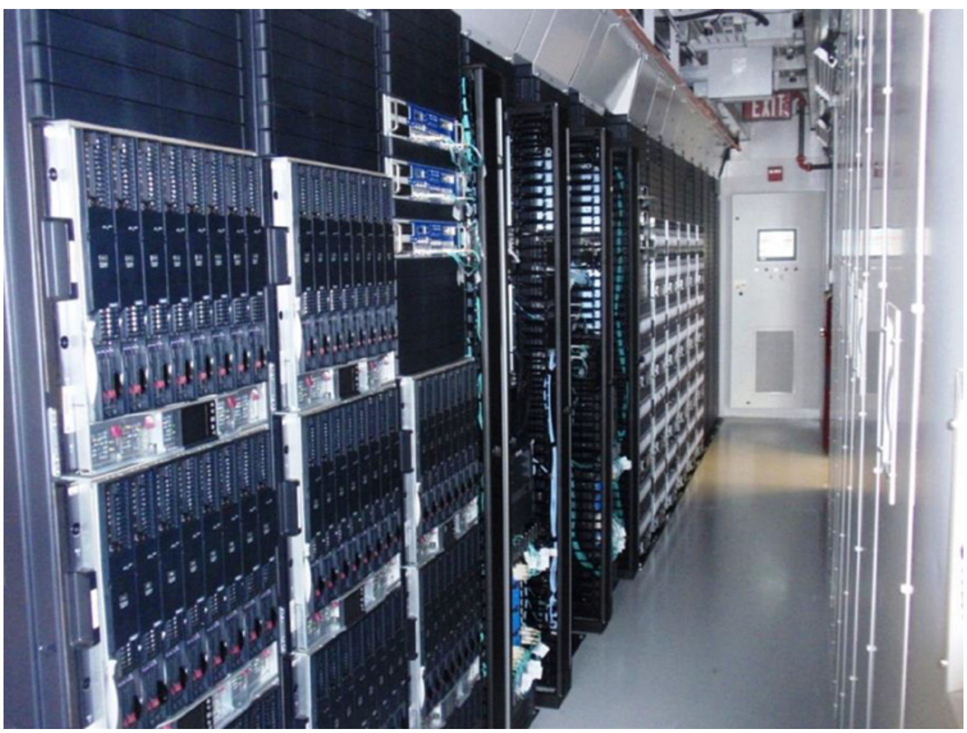


FIGURE 4.16 Having network cabling racks in a few locations can make complex cabling on-site easier at the trade-off of less racks. Courtesy of Hewlett-Packard Company.

4.4.8 Multiple Module Alignment

All of the earlier sections discussed the architecture elements inside an MDC or across an MDC provided by a single supplier. In the world of construction and multi-vendor sourcing models, MDCs also offer the opportunity for having different experts to provide different parts of the design. Clear documentation of interface points, schedules, and responsibilities is required, but this is not

FIGURE 4.17 Adiabatically cooled MDC in a warehouse with the network, control and power conduit connected to the outside world. The two large pipes in the hot aisle extension are a conduit pass through to run the network straight to the site IDF. Courtesy of Hewlett-Packard Company.

uncommon compared to a traditional site-built data center construction project. Given the newness of MDC design, the expectations of the owner, the general contractor (GC), and the module supplier may be different. The interplay of the modules can affect the total site energy efficiency, schedule and cost, so clear communication is needed between vendors.

Figure 4.18 shows each module can affect the energy efficiency and capacity of the total solution. The cited example shows a 40 ft, 500 kW 20 rack 2*N* powered power and cooling solution. That container is well documented to use 240 gpm of 55–75°F water with the colder water required for the higher IT load. The only non-IT power other than lights and controls are the fan coils that take the server waste heat and pull it over the colder cooling coil. The fan energy is documented to be variable and results in a pPUE of 1.01–1.02 from 100 to 500 kW.

The IT capacity and energy efficiency of the container are depending on the IT CFM, the chiller operational parameters, and the power module functionality.

If the IT doesn't come from the container vendor, the understanding of its airflow at various workloads and inlet temperatures is critical. One vendor may run IT with 55°F water stably at 520 kW, while another's 350 kW IT trying to use 75°F condenser tower water may ramp server fan speed so high that the IT power may not be achievable.

As a total solution, TCO per transaction is a better metric than PUE because servers with high fan power at higher ambient temperature can have better PUE but do less work per watt at the solution level.

Chiller decisions on capacity and efficiency should compare: part load value if it is at 20% duty cycle 90% of the year, operations across the year with high enthalpy conditions, and ability to run compressor-less for as many hours per year as possible. The run time of the UPS, synchronization time to the generator, maintainable bypasses and type of UPS technology are also variables. A choice of one module technology over another could also have larger local service costs that may negate any energy savings or first costs savings, so knowing the whole TCO is the most important.

These factors make this deployment vary from a 12 weeks, \$4M, 1.15 PUE site to a 26 weeks, \$7M, 1.50 PUE site. All of this is contingent on the right specification of the IT, IT module, chiller, and UPS that make up the solution.

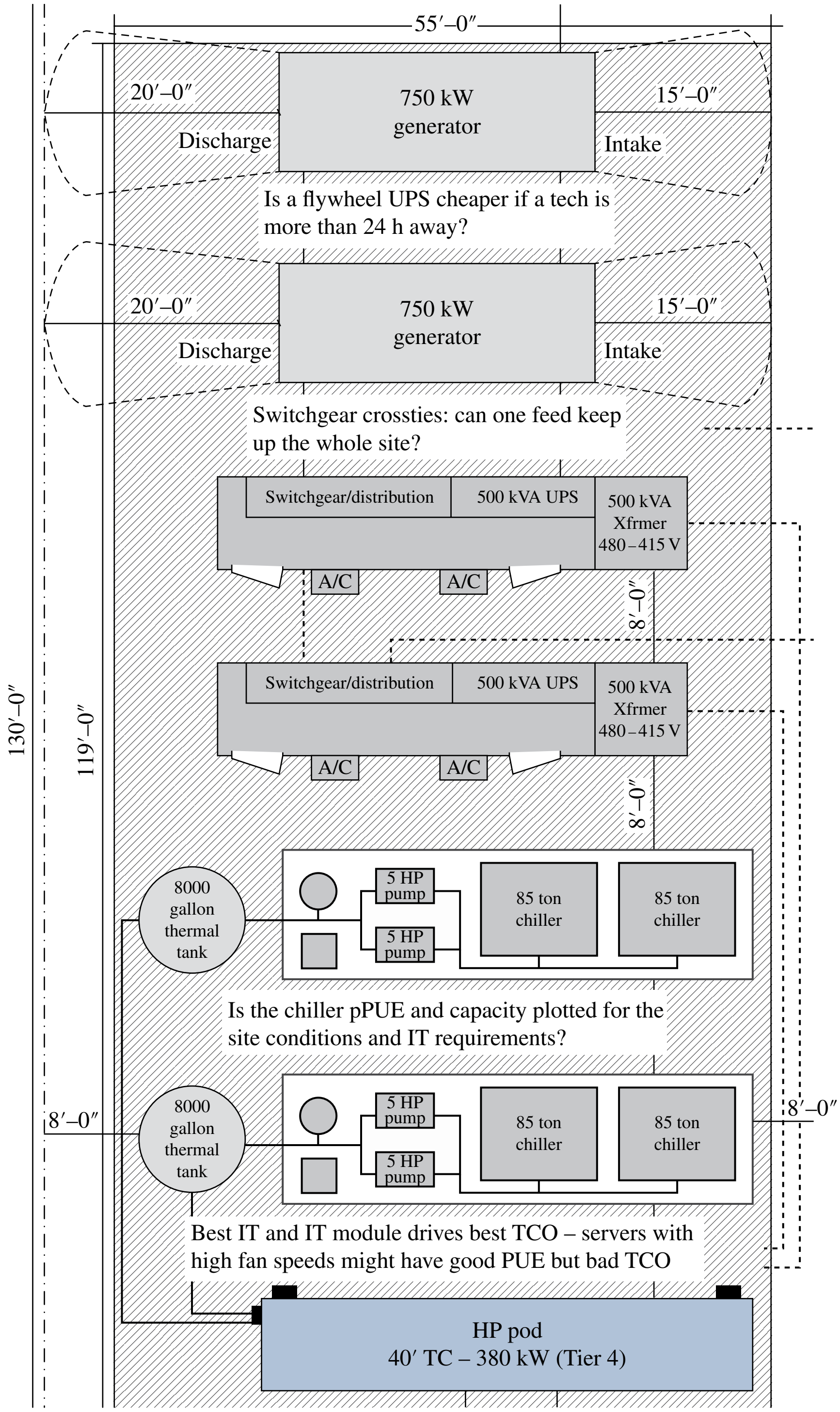


FIGURE 4.18 An MDC can have parts from all sources, so make sure you know who is responsible for what. All IT work/rack, container pPUE, and chiller efficiencies are not equal, and the TCO is based on how much IT performance you get out of the whole solution, not just one of the parts. Courtesy of Hewlett-Packard Company.

4.5 SITE PREPARATION, INSTALLATION, COMMISSIONING

Building an enterprise class data center from a collection of factory preassembled modules can be faster and cheaper than a traditional large data center construction program but only if executed as a construction project by the same class of professionals who would have built the site from scratch. Phrases like “plug and play” and “just add power” make it sound simple. Whether it is three 333 kW 2 *N* powered water-cooled modules or one 1 MW 2 *N* DX-cooled module, it still requires 2.5 MW of conduit and conductors. All of that interconnectivity needs to be planned for and scheduled up front. There is a lot of complexity from the medium-voltage utility, into and out of the generator, UPS and switch gear modules, and then finally into the IT and cooling modules. Getting it all designed, built, and verified is critical as featured in Figure 4.19. Once the concrete is poured, the design is locked.

In this figure, data center design professionals made sure there was plenty of room to locate all of the electrical conduit under the MDC site. These electrical duct banks whether underground or overhead follow all the same designs standards as if this were going to be used with a site-built data center of the same scope.

For an MDC design, the layout and site prep must also allot for the clearance required to get larger prebuilt modules into place. A traditional DC project needs a detailed timeline to complete all of the serial nature of the tasks. An MDC project is much the same from a planning perspective. It is just that many of the parts can be done in parallel at multiple sites and just come together at the final site. The challenge then is to optimize the schedule and make all of these elements come together at the same time just when needed. That can make for a lot of jockeying of trucks when it all comes in over just a few days time. The same example in Figure 4.19 had most of the prebuilt modules arriving 30 days after the pour. Figure 4.20 shows just how crowded the site can become and how clearances for the cranes, buildings, and modules all must be accounted for. Delivery trucks are long, tall, wide, and heavy—a 500 kW module of IT may weigh 120,000 pounds. This length and height will mandate the size of the crane needed to set them. MDC clearances are usually tight to take advantage of the available real estate, so this further impacts making sure there is enough room to place cranes, have truck turning radii, and sufficient room to allow electricians to pull all of their wire and cooling technicians to reach and repair their heavy equipment. All this is common on a site-built data center, so an MDC must all take the same care in planning.

A large site-built data center project can take 18 months due to the serial nature of the process and based on how long each of the individual elements takes because of all the

FIGURE 4.19 It takes about 30 days to dig a hole and prep 2.5 MW of electrical conduit to prepare to pour concrete for a 2 *N* 700 kW MDC expansion to an existing data center. Courtesy of Hewlett-Packard Company.

FIGURE 4.20 Tight clearances for long heavy MDC components all arriving at the same time requires complex engineering and planning. Courtesy of Hewlett-Packard Company.

on-site labor. An equivalent MDC can be online in 6 months and at lower costs, but only if the planning and design of the modules and the construction all come together as designed. For example, the first IT module shown in Figure 4.21, the warm water cooling plant (not shown), and the rooftop site all had to reach completion about the same time. If the chiller plant was site-built on the roof too, or the roof space was just going to be built like a regular hall, the time and expense would have been much more. Given the modular nature of the total design, additional IT modules, cooling towers, and power modules were added just when needed over time.

The "building" elements of a construction project aren't all that is needed, and the clock doesn't stop until the IT is on-line in the new DC. This is just as critical for the MDC because speed of deployment is dependent on the parallel efforts of site preparation and the IT module build. Even if the timing is perfect and the site is ready the day the IT module lands, its deployment can take 4–6 weeks minimum from landing onsite. The module has to be set and plumbed and power conduit must be pulled and connected. Typically, the IT module vendor has their own start up scripts before it is ready to be handed over to the GC and commissioning agents. One way to speed up that process is integrated factory test of cooling units and the IT. That can be simple if it is a water-cooled container and the supplier has chiller plants at their factory. It gets more complex to have a combined test of a multipart module like a DX if it is planned to come together until the final site. New power from a utility can take longer than 6 months, and a beautiful container full of IT and cooling can't even be turned on without power. All these plans could fall apart if the authority having jurisdiction (AHJ) requirements are not met. The AHJ often looks for guidance or precedence. Product Certifications like ETL, CE, and UL will help. Given all of the lead time on any of this, it is critical to plan this startup schedule and commissioning strategy prior to ordering the IT module.

Most data center designs have large operational margins to offer a buffer against multiple failure events and stay operational. Many MDC designs give the flexibility to reduce that margin to save first costs and operational expenses. Hot and cold aisle containment designs to increase density are one such design element. Putting 1000 kW into 200 racks spreads out the risks and fault domains but has a bigger footprint, needs more cooling, and simply costs more than operating all that power in just 40 racks. As applied to an MDC, high-density containment with DX outdoor air cooling will require quicker automated reaction to cooling failures because there is just less thermal inertia. One way to obtain this reaction is by keeping the DX fans and compressors on generator or on a faster UPS response. All of this needs to be tested as a system to ensure it operates as designed.

As most MDCs are not built inside multi-million dollar buildings, design for their environment must be considered.

FIGURE 4.21 Not uncommon for MDCs is to scale over time. The first water-cooled container IT is being commissioned, while the second one is being delivered.

Those that use outdoor air, that are densely packed, or those that sit close to or in a building must have clearly understood airflow paths. The deployment has to ensure that the IT inlet gets what it needs but must also ensure that the exhaust path is unimpeded. Placing the IT module in a ventilated shed is acceptable, but the airflow needed for condensers and generators can often dictate where they may be placed. Air-side economized site-built data centers have the same issues as an MDC requiring acceptable full cooling with poor air quality like smoke, volcanic ash, and dust infiltration. Similarly, adiabatic air cooling or water-side economizers need to account for site-water quality and treatment. Power quality also matters as the low PUE of an MDC has less capacitive loads, and provisions must be made to accept or deal with leading/lagging power factor, phase imbalance, or harmonics. Employees will require inclement weather access like an awning or a physical structure to aid in service and security. If an MDC isn't part of a larger campus, many other ancillary systems need to be designed such as lightning protection, grounding, drainage, fire, security, and managing system communications to the support staff. If it is part of a campus, then the module may need to adapt these systems to site standards like fire and security.

There are hundreds of thousands of service professionals worldwide who understand the care and feeding of data centers. Making sure your MDC gets the same TLC requires some up-front thinking. If your MDC is going to a remote site, there will be a need to have a spares kit with unique items kept on site like fans, coils, and transformers. Even if your MDC is next to your office, there may be long lead time, critical or specialized components that must be planned for. MDC space can be more compact and hotter than traditional DC space which may require special training to work in tight environments. Regulations like OSHA could limit technician time in the harsh conditions of a contained hot aisle or just the outdoor environment if your MDC uses free cooling in a place that isn't conditioned like a traditional office building.

Depicted in Figure 4.22 are two different types of containers at one site. The right one was based on an off-the-shelf model, whereas the left one had several unique modifications for this particular site and the IT inside. Like the changes for the left container, there are hundreds more

FIGURE 4.22 20-ft water-cooled custom module next to 40 ft standard module. Courtesy of Hewlett-Packard Company.

MDCs in the field, and many lessons learned are bubbling up in the industry.

- With all the doors on MDCs, the deployment effort is hard on door hardware, so consider the use of construction doors and hardware. Similarly, use shipping-size doors only where that functionally is required. All of these doors should to be weather-tight and fire suppression gas tight 15 years after deployment.
- Sometimes, cosmetic modifications are required to facilitate simple building interface like clearance for door swings, second floor access, and simply local architecture codes.
- Existing ancillary system (Fire, Security, BMS) compatibility drives customization—a more flexible interface may be needed to connect to existing designs.
- Top entry power feeds are not desirable due to difficult in weatherizing.

Many of the examples cited earlier and throughout this chapter show the flexibility in MDC design. Most of the citations are based on existing deployments. Care should be taken when compiling the wish list of the features for your MDC because some of these features may not be available off the shelf in every vendor's model. This matters because as the level of module customization increases, the time to delivery moves toward the time it takes to build out a full site-built data center. The two graphs in Figure 4.23 illustrate the comparison of the times from start to IT available. Custom site prep is like a regular data center project; it takes the most time and then the module is scheduled to be delivered at the right time to make the final schedule date. Adding custom features such as a specific type of door or security only extends the time slightly as it was some engineering work up front to make the custom change. Whereas, a custom module like the one on the left in Figure 4.22 will often require a more detailed site design as a deviation from a cookie-cutter design. The new custom module design will also drive additional time, so, for example, the right module project shown in Figure 4.22 took 28 weeks, but the left module project took 43 weeks. While this looks like deficiency of on MDCs, when both are compared to what the equivalent full site-built data center would take at 78 weeks, this was still a clear time saver for the client.

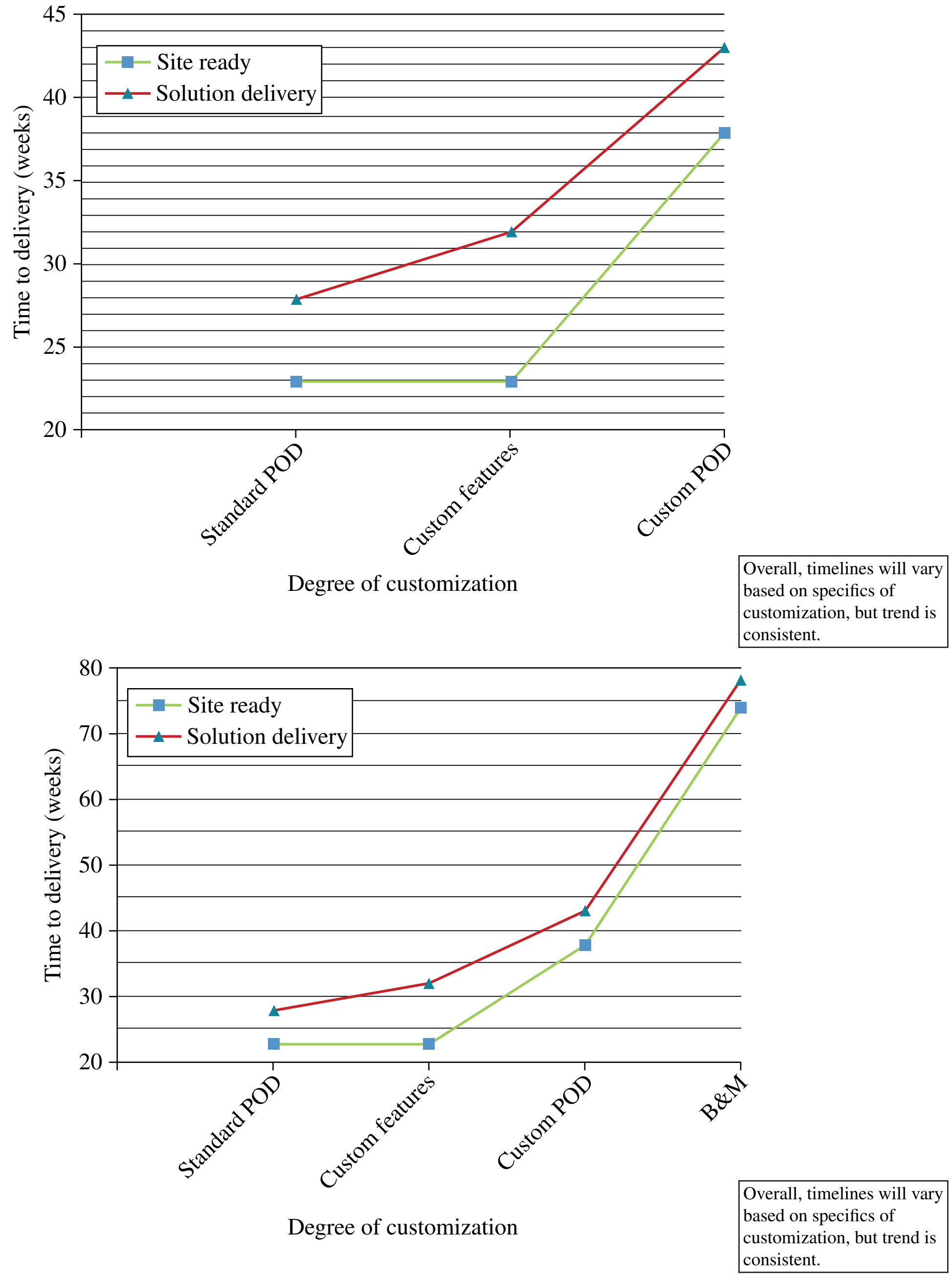


FIGURE 4.23 Comparison of time to live IT for off-the-shelf versus custom MDC versus regular DC. Courtesy of Hewlett-Packard Company.

Commissioning site-built data centers is such a normal practice, and it rarely leads to questions on how, when, or who does it. MDCs offer a unique opportunity to the commissioning activity that could further reduce costs and shorten lead time if properly accounted for. As part of product design, module manufactures offer some level 3 and level 4 commissioning purely as a function of product performance. The added focus will be on the full level 5 integrated design validation.

When it comes to commissioning work, the first question to consider is where to commission and at what load for IT, mechanical, and electrical. For the IT, it could be load tested at the factory during the module's startup, and again for continuity at the site. There will be a question of when the owner takes responsibility for the IT because testing may require dozens of power cycles; and if this is for large numbers of servers, some failures are inevitable, so ownership and how this is dealt with is critical. The electrical train can

FIGURE 4.24 Test the load with load banks that mimic server sealing and airflow. Courtesy of Hewlett-Packard Company.

be load tested at factory for continuity and as a functional test at site. This could be as simple as testing the branch circuits. Testing the cooling involves multiple options and requires evaluation of cost versus customer comfort level of validated performance.

Ideally, the IT vendor would load common software benchmark tools like Linpack right onto the real preconfigured IT and test at the factory and again at the site. But it is possible the client software prevents that, so that means site-level commissioning may be limited to the testing at an IT kW stress level below the design point. Adding load banks at the site is optional if spare racks are open, or more load could be induced with better stress software on the IT. The agent may ultimately need to remove IT racks to test at the MDC design load.

Data centers that have containment need to consider the load bank cost, lead time and type. Rack-mounted load banks are best because of the zero airflow bypass and behavior similar to servers, but then more of them are needed and power cords must be fit to them. It is ideal to run load banks on the IT power plug strips to test them, but it may be easier to install temporary power straight from the branch circuit. To get all of this right, having the IT module vendor use server simulator load banks in the factory like Figure 4.24 is preferred. This process could be replicated again on-site as required. Using IT with full-load software and rack with load banks, full power and cooling suites can be tested, including all conceivable fault domains.

It is not desirable to use larger floor-mounted load banks at 50 or 100 kW each. Integrating them into the MDC and installing them to act like IT can require great time and expense.

4.6 HOW TO SELECT AN MDC VENDOR

The key to select the right partner to build an MDC is to start with the definition of what you want your IT and data center to do over a 12–36 month period. If forecasting is not available, select a vendor that can help with forecasting. Look for proven experience with positive client references. Look for a stable company that will be around at least as long as your warranty period. Determine whether you have a comfort factor and experience with running a construction project. If not, look for a vendor that can provide a turnkey solution or best solution that meets your need. Conduct a request for information (RFI) based on that criteria to narrow down the field if you have a lot of uncertainty on what type of data center you need.

At the request for proposal (RFP) phase, solicit 3–5 proposals, conduct follow-up interviews and make multiple visits before awarding the contract. Here are examples of what could be documented in the contract. Schedule and fully define where and when the witness tests will be conducted. Document the type of support expected at each phase of the construction process. Ownership of deliverables and responsibilities are important. A single point of contact like a GC or architect may not be common with all MDC vendors, or may not be possible if sourcing is diverse. While all documentation can't be complete before award, reviewing sample documentation for prior installations will provide perspective.

4.7 EXTERNAL FACTORS

The funding of a new data center or data center capacity expansion can have significant budgetary and tax implications based on how the technology is viewed both by the tax code and by corporate finance policies. The nature of MDC solutions can blur the lines between IT equipment and facility infrastructure in many areas. As a result, this construction approach can lend additional flexibility to organizations to utilize the funds that are available to accomplish the organization's goals. Servers and storage devices will always be considered IT hardware, and permanent support structure, such as an underground fuel tank for extended generator operation, will usually be considered a capital improvement. In modular data center construction however, many of the modules in-between can be defined as either a part of a building solution or part of an IT solution. This ability may lead to the greatest flexibility in minimizing organizational tax exposure as well as using the budgets that are available.

Two other financial considerations in MDC construction are the incremental nature of construction supports incremental funding models. The ability of many modules to be repurposed for other customers lends the ability in many cases for modules to be leased, creating an expense model for components that may otherwise be considered capital infrastructure.

Other tax ramifications fall into similar categories that vary by corporate or even local regulatory governing bodies: expense versus capital; trailer based may equal temporary; container may mean 5 year; if it is a building, life safety requirements should be considered.

One way to avoid some of this hassle is to assist the local authority having jurisdiction (AHJ) with the ease of qualification. Some MDC vendors are NRTL "LISTED" as Information Technology product to UL 60950.

The industry has had a recent push to the new UL2755 standard: http://www.ul.com/global/eng/pages/offerings/industries/hightech/informationtechnology/ul2755/. While some of these are etched in stone, others are subject to local requirements and you should understand them all before you begin.

4.8 FUTURE TREND AND CONCLUSION

The pace of change in technology and business is another factor pushing fueled interested in MDC construction. In the first several decades of data centers, a data center was typically an investment considered primarily by large stable businesses. These businesses had consistent growth and operations models, and as IT infrastructure steadily reduced in size, power consumption resulted in more equipment installed in the same space. Today's IT world is full of small, medium and large business that are new, old, and re-invented. All are going through consistent changes between growth, contraction, and shifts into different technologies with different requirements. Only 10 years ago, a new cell phone meant that the size shrunk 30% and a full keyboard had been added; yet, today's cell phone replacement can be as powerful as last generation notebook computer. That same IT shift is happening in the gear data centers are designed to house.

An obvious result of the changing nature of data center operators from large stable companies to all sorts of companies, including collocation providers who provide datacenter space for others, is that building large infrastructure with limited flexibility in the expectation of keeping it nearly the same for 12–15 years does not fit most current business needs. MDC addresses these needs in several different ways. As discussed earlier, the most obvious is facilitating incremental growth that provides just a little bit more capacity than is currently needed with great flexibility to add more as needed. Another consideration in the granular construction approach is that when less capacity is needed, modules can be shut down and idled so that the energy required to keep them ready to turn up when needed is drastically reduced. This ability can save significant operating expense as well as carbon emissions when disruptive change occurs. The modular data center also allows for disruptive technology to be addressed by replacing modules rather than facilities. An example would be a company introducing high-density computing for the first time. While this may require retrofitting an entire 10,000 ft^2 data hall in a brick and mortar building to support high density in a small section of the room and yield unnecessary capacity in the rest, an MDC can retrofit or replace just enough modules to accommodate the new technology and have the necessary support deployed only as needed and without provisioning unusable capacity to dozens of other server racks.

All these trends in modularity are evolving with several competing positions. An understanding of your IT Tier requirements will give you Tier flexibility with this modularity. This is even more true when you understand the kW/rack across the spectrum of IT and applications being considered over 36 months.

Remember it is still a large capital construction project and deployment is complex like a conventional DC. Good planning is key to Delivery Time. Consultants, A&Es and GCs are important. Commissioning approach (IT and module) is best understood before order is placed. Implementation of an MDC has similar elements to a DC implementation, but many are different; you and your GC need to understand those differences.

There will be tighter integration of large blocks of IT and facility supporting clouds, Service providers and HPC. These trends are going to affect all aspects of data centers with fewer operators providing larger collections of more homogenous gear.

More governments worldwide want data and jobs local. Data local is great for MDC. Jobs local may face scrutiny as MDC components aren't built locally. Some states and countries are finding new ways to entice new data centers, while others are trying to optimize tax revenue from businesses inside their jurisdiction. Modular may not always mean mobile, but a modular data stamp repeatable in 6 months can be effective tool to arbitrage the shifting political winds.

FURTHER READING

- The Green Grid, http://www.thegreengrid.org/~/media/WhitePapers/WP42-DeployingAndUsingContainerizedModularDataCenterFacilities.pdf?lang=en
- Many companies in the industry are active participants in modular data center design and have their own websites. While this list is not comprehensive, all of these companies and organizations have source material on the subject.
- HP, IBM, Dell, SGI, Cisco, Oracle, Google, eBay, Emerson, ActivePower, Eaton, Gartner, Forrester, IDC, Tier1/The Uptime Institute, greenM3, ASHRAE, OpenComputeProject, UL, The Energy Efficient HPC Working Group, Datacenterdynamics and LinkedIn
- APC, http://www.apcmedia.com/salestools/WTOL-8NDS37_R0_EN.pdf
- http://www.apcmedia.com/salestools/WTOL-7NGRBS_R1_EN.pdf
- Microsoft,http://loosebolts.wordpress.com/2008/12/02/our-vision-for-generation-4-modulardata-centers-one-way-of-getting-it-just-right/
- http://loosebolts.wordpress.com/2009/10/29/a-practical-guide-to-the-early-days-of-datacenter-containers/
- LBNL/DOE, http://hightech.lbl.gov/documents/data_centers/modular-dc-procurementguide.pdf
- Rich Miller, http://www.datacenterknowledge.com/archives/2012/02/06/the-state-ofthe-modular-data-center/
- John Rath, http://www.datacenterknowledge.com/archives/2011/ 10/17/dck-guide-tomodular-data-centers/
- http://www.datacenterknowledge.com/archives/2011/05/24/video-what-doesmodular-mean/
- http://www.datacenterknowledge.com/archives/2006/10/18/suns-blackbox-gamechanger-or-niche-product/
- http://www.datacenterknowledge.com/archives/2008/04/01/microsoft-embraces-datacenter-containers/
- http://www.computerworld.com/s/article/357678/Make_Mine_Modular

5

DATA CENTER SITE SEARCH AND SELECTION

Ken Baudry
K.J. Baudry, Inc., Atlanta, GA, USA

5.1 INTRODUCTION

Almost all data center disasters can be traced back to poor decisions in the selection, design, construction, or maintenance of the facility. This chapter will help you find the right combination and eliminate poor site selection as a cause of failure. It begins with setting objectives and building a team, examining the process, and selection considerations. The chapter concludes with a look at industry trends and how they may affect site selection.

Site selection is the process of identification, evaluation, and, ultimately, selection of a single site. In this context, a "site" is a community, city, or other populated area with a base of infrastructure (streets and utilities) and core services such as police, fire, and safety, education, and parks and recreation. This definition is not meant to eliminate a location in the middle of a nowhere, the proverbial "corn field" location. However, experience indicates that unless an organization is really set on that idea, there are few such locations that have the key utilities (power and fiber) and core required to support a data center. Most organizations back away from this idea as they estimate the cost and logistics of operating such a facility.

Site search and selection can be as comprehensive or as simple as necessary to meet the goals of the organization. In practice, it consists of asking a lot of questions, gathering information, visiting potential communities, and perhaps entertainment by local and economic development officials.

The motivation behind conducting an extensive site search tends to be economic (i.e., find the site with the least total cost of ownership). While the site search and selection process attempts to establish an objective basis for the decision, it relies on estimates and assumptions about future conditions, and often includes criteria that have an economic value but are not easily fit into an economic model. These criteria include marketing and political aspects, social responsibility, as in giving back to the community, and quality-of-life issues. These issues tend to be subjective and are often weighted heavier in the decision matrix than the economics suggests. The end result is that the site selection process tends to be a process of site elimination based on economic criteria until the list is pared down to a few sites with similar characteristics. The final decision is often subjective.

There is no such thing as a single "best site" and the goal of a site search is to select the site that meets the requirements, does not violate any constraints, and is a reasonable fit against the selection criteria.

In general, the goal in site search and selection is to ensure that there are a sufficient number of development opportunities in the selected community. However, there are good reasons to search for specific opportunities (existing data centers that are suitable or that can be refurbished, buildings that might be converted, pad sites, or raw land) as part of the site search and selection process. There are often negotiating advantages prior to the site selection announcement and the identification of a specific facility or property might become the deciding factor between the short-listed sites.

5.2 SITE SEARCHES VERSUS FACILITY SEARCHES

Most of the discussion contained in this chapter is based on selection of a site for development of a data center from the ground up a.k.a. "a brown field site." However, there are

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

other viable data center acquisition strategies such as buying or leasing an existing single-tenant data center or leasing a co-location center. With co-location and existing facilities, much of the investigative work should have already been done for you. You will be able to ask the prospective landlords to provide answers to your questions. It is likely that some aspects, such as power company rates and tax incentives, have already been negotiated by the developer.

You will still need to understand what you want to achieve and what your requirements are, and it would be a mistake to assume that the developer had the uncanny ability to anticipate your requirements and has taken them all into consideration. Data centers are not a one-size-fits-all proposition, and existing centers may have been located where they are for reasons that aren't related to the data center business at all. It could be that the original tenant selected the location because it is where someone's grandchildren live.

You will still need to compare the information you are given against your requirements, and analyze the total occupancy cost for each prospective facility. In many cases, the lowest cost alternative may not be the one with the best lease rate. It might be the one with the best power company rate and the most opportunities for energy savings driven by the local climate.

5.3 GLOBALIZATION AND THE SPEED OF LIGHT

Globalization has been around since the beginning of time, when a group of adventurers left one area to seek better fortunes in another area. Today, it is largely driven by economic forces as organizations seek out competitive sources of raw materials and labor and new markets for their products. Technology, in particular air travel and voice and data communications, has made it possible for organizations, more than ever before, to consider sites outside of their country of origin.

Historically, organizations have located facilities overseas to secure raw materials or inexpensive labor and to avoid taxes and safety and regulatory policies of their country of origin. The question for us is: "Are there economic benefits to locating a data center outside of one's country of origin?"

The motivation for locating overseas, today, is very different from the raw materials and cheap labor considerations of the past. Data centers don't need a large pool of unskilled labor like manufacturing and call centers, and they are not tied to location by raw materials. Data centers can be located almost anywhere that can provide significant amounts of power and connectivity to high speed national and international communications networks.

Including sites in foreign countries adds a layer of complexity as the differences in tax structure, laws, ownership of real property, security of data, political unrest, etc. need to be considered. There is, however, one difference that is not easily eliminated by technology or money: communications signals cannot propagate any faster than the speed of light.

As an example of how this figures into site selection constraints, let us consider a route from Bombay, India, to New York City, New York, United States. The speed of light in free space is 299,792 kilometers per second (km/s) or approximately 186,282 miles/s. It is slightly slower in optical fiber, but we are going to ignore that for simplicity. The distance is about 7800 miles and assuming that a fiber route would be 50% longer than a direct path, it takes about 70 ms (1 millisecond is equal to 0.001 second) for the signal to propagate one way. This figure does not include any latency for network interfaces, transmission gear, etc. So unless Einstein was wrong, 70 ms is the best possible signal latency.

Today, the expected latency between the United States and India is between 250 and 500 ms. Network improvements will likely drop this figure over time to about 200 ms. How this limits site selection depends largely on the applications that you use.

One of the most common latency-sensitive applications is storage array replication. In data replication, an application will write data to a disk or storage array and the array will replicate the same to a remote array. When confirmation comes back that the data has been successfully written, the transaction is complete. If the latency is high, then the performance is seriously degraded. Other latency-sensitive applications include transaction-oriented applications like banking and retail sales where, due to a very high transaction rate, the latency must be very low and burst capabilities are required to meet high demand.

Without a doubt, some organizations will realize significant benefits from selecting an overseas location. But there are challenges, and it will not be a suitable option for everyone.

5.3.1 The Site Selection Team

Like any process with many moving parts, site search and selection requires diligence, clear expectations, schedules with defined deliverables and due dates, and good communication between all stakeholders. This doesn't happen by itself.

Project management is the art of managing a process from beginning to end. It concerns itself with reaching the end state by defining and organizing the work effort, communicating and leading stakeholders, and driving decision making consistent with requirements, within constraints, and in a timely manner.

Project management sets metrics for constant feedback on performance and adjusts accordingly to keep on track. It is a profession, and the role of the project manager cannot be overstated. If you do not have someone in-house who has successfully demonstrated their competence, you should look outside of the organization.

Some key reasons why projects fail are similar across almost all endeavors are as follows:

- Lack of User Involvement
- Unrealistic Expectations
- Incomplete Requirements and Unsupportable Criteria
- Lack of Planning
- Lack of Executive Support

Before your organization can make good progress with the site search, it will be important to identify potential team members and start building the team. The timing of "when" you bring your team on board is as important to your success as "who" you bring on board. Assumptions about cost and feasibility are often made early in the program, before Subject Matter Experts (SMEs) are traditionally on board, and often spell doom for a project when it is determined that they are not realistic and, more often than not, need to be adjusted upward. You will need your team on board early in the process to avoid this sort of pitfall and to create plausible criteria and constraints list.

As you build your site selection team, you should talk with key executives and ask for recommendations. By involving them early, reporting progress in a concise and effective manner, you will gain their support. If you are the key executive, then you should involve your board, investment committee, and your key subordinates as you will need their support as you push your search forward. As a minimum, this effort should lead to a better understanding of the project and identification of who will cooperatively back the project and who won't.

One of the key consultants will be a site search consultant. A good site search consultant will not only be an SME but an advisor as well. He will not only guide you through the site selection process but will know what to expect from economic development agencies. He will know where to get answers, know other professional resources in the industry, and understand how to evaluate proposals. He will understand when it is time to bring the search to a close and how to bring it to a close. He will render advice and help you make good decisions. In many cases, he will be able to save you time and money, by eliminating some sites, based on recent, relevant experiences prior to beginning a search.

Your team will need to include SMEs such as data center planners, data center consultants, architects, engineers, lawyers, tax accountants, and real estate brokers. Any specialty that can vary in cost between sites will need some level of representation on your team. The greater the expected cost variance between sites, the greater the need for the SME.

An architect might only be needed to create an initial estimate of the space required and nothing more if construction costs are believed to be the same regardless of site selection. An engineer might only be needed to create a power consumption profile, energy consumption and demand, for the proposed facility. While a utility expert might be retained to evaluate energy costs for each prospective site.

When building a team, there is often a distinction between consultants and vendors. Consultants typically charge a fee for their services. Vendors typically provide preliminary support for free or for a minimal fee in hopes of making the big sale later on in the project. This includes real estate brokerage firms that will assist with site selection, general contractors that will perform preliminary design and develop budgets, and others. While you should surround yourself with the best resources that you can afford, the key is to surround yourself with resources that have the experience, competence, and that you can trust to act in your best interest. All three are important!

5.3.2 The Nature of the Site Search and Selection Process

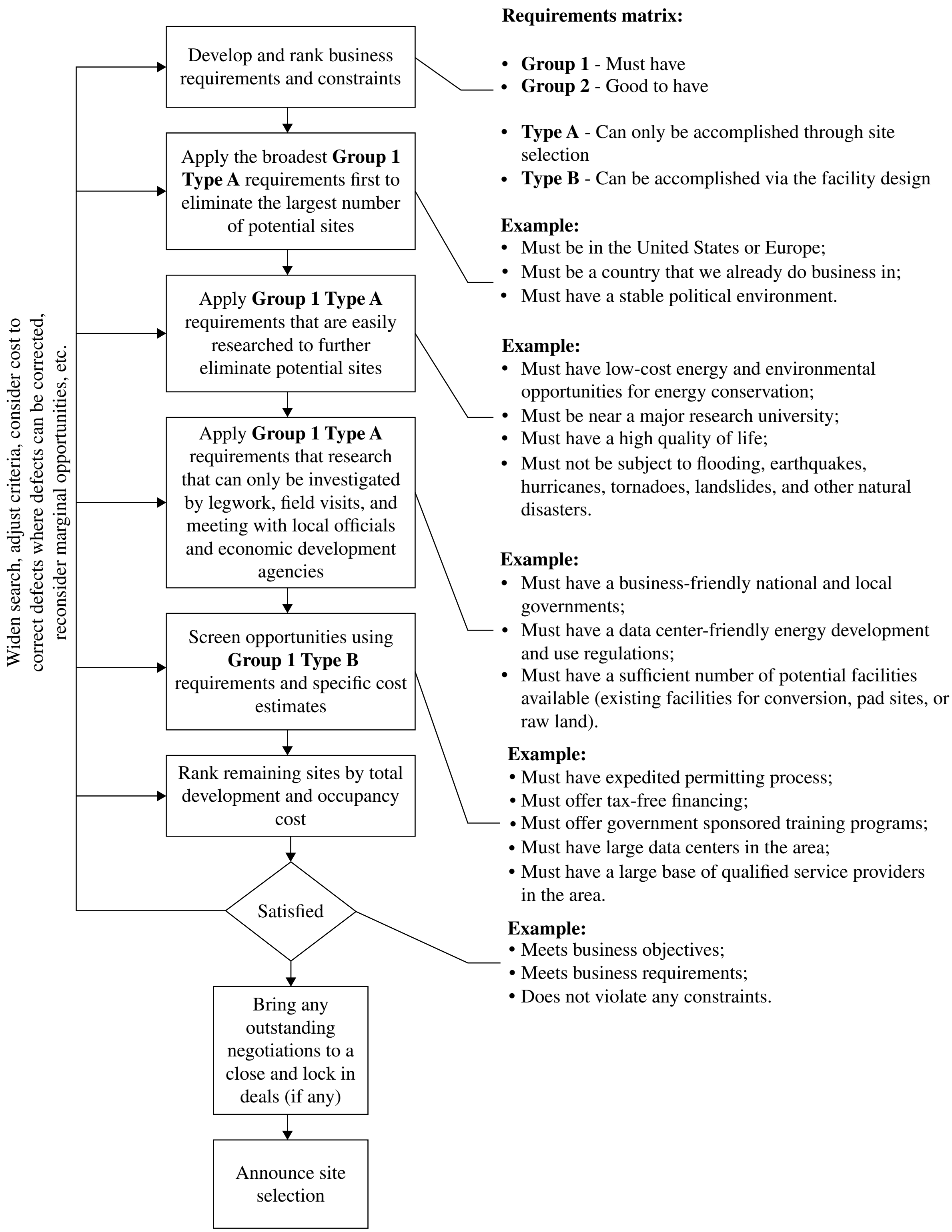
There are a few characteristics of site search that require special mentioning; it's generally performed in secrecy and it's not a search process but an elimination process, and there comes a time when continuing the process will no longer produce further value.

5.3.2.1 Secrecy More often than not, a site search is conducted in secret. There are many reasons including the following:

- Site searches don't always result in moves or in the development of new facilities. Announcing a site search and then not following through may be perceived as a failure.
- News of a site search may cause concerns among employees over potential facility closures.
- Most businesses don't find it competitive to telegraph future plans to their competition.
- Once the word is out, you will be inundated by vendors wanting to get a piece of the action.

Regardless of your reasons, it is likely that management will expect that the search be conducted in secret. Most site search consultants are going to be aware that this will be the case. As the size of your team grows and many aspects of the work will require team members to consult with resources outside of the team, such as equipment vendors, it is likely that by the end of a lengthy search, there the number of individuals aware of your search will be quite large.

If secrecy is important to your organization, then you will need to take precautions. The first step should be execution of simple confidentiality and nondisclosure agreements with every team member. Many of the consultants will be looking to perform additional services once the site is selected, and well-timed reminders about disclosure can be very effective.



Site search and selection process

FIGURE 5.1 Site search and selection process. Courtesy of K.J. Baudry, Inc.

Some companies give their site selection projects a code name. This works well as it gives everyone a name to reference and serves as a constant reminder that the project is secretive.

When it comes to secrecy, we are often our own worst enemy, and many individuals unknowingly let out the secret by wearing their company access card with picture id, wearing clothing or carrying pens or clipboards with a corporate logo, handing out business cards, providing phone numbers which can be easily traced back to a company, etc. Even discussing your travel arrangements and where home is can provide significant clues to the curious. While all of these things are obvious, it's the most obvious things that we forget to deal with. If you really want secrecy, you will leave nothing to chance, and let your site search consultant handle all of the communications.

5.3.2.2 Process of Elimination Site search and selection is a bit of a misnomer as it is more a process of elimination than selection. It starts with a broad universe of possibilities and moves through rounds of elimination (Fig. 5.1). The broadest criteria are applied in the first round to eliminate the most number of sites and narrow the search. The broadest criteria are usually geographic. For example, North America, Europe, cities of five million or more in population, within my service region, etc.

Each successive round uses more selective criteria until all, but one site is eliminated. This may sound a bit daunting; but in practice, the number of possibilities falls rapidly and narrows to a dozen or so very quickly.

By the time it gets down to two or three of sites, the remaining sites are all very similar in terms of economics and the search becomes as subjective as it is objective. At this stage, organizations may eliminate a prospective site based on perception, and previously unknown criteria may spring up as a way to differentiate between the remaining sites.

One thing is certain, once you make your announcement as to final selection, the game is over. There are no more concessions to be won and no more dealing to be done. It is important to keep two or three candidates actively involved until the very end.

5.3.2.3 Relative Costs and the Law of Diminishing Returns Estimating all of the cost may be important to the funding decision and may be developed as part of the overall project strategy; but as long as there isn't a difference between sites, it's not critical to the decision. For the purpose of site selection, it is only necessary to compare costs that differ between sites. For example, if license fees for operating systems and applications are the same regardless of where deployed, then this is not a cost that needs to be evaluated as part of the site selection process.

Site search and selection involves many forecasts of future events: how much power you will need, how fast you will grow, how many people you will employ, how the political and economic climates will change over time, etc. There is a cost to the site search and selection process and at some point the cost of obtaining more information, performing more analysis, and perfecting the estimates starts to exceed any benefits that may be derived from the effort.

In the end, worrying about costs that aren't significant or key to making the site selection decision and overanalyzing the costs tend to only delay the decision. There may be a clear choice; but more often than not, there will be two or three sites that are all well suited; and if you have done a good job, all very similar. At this point, additional analysis is not likely to improve your decision and it is time to bring the process to a close.

5.4 THE SITE SELECTION PROCESS

Site search and selection can be boiled down to several key activities as follows:

- Develop Business Requirements and Constraints
- Define Geographic Search Area
- Define Site Selection Criteria
- Screen Opportunities
- Evaluate "The Short List"
- Close the Deal

5.4.1 Develop Business Requirements and Constraints

The first activity of a site search is to define the business requirements. Typical motivations for conducting a site search may include supporting growth, reducing cost, expanding into new markets, etc. Requirements are the items that we believe are necessary to conduct business including space, power, cooling, communications, etc. Some might add "in a profitable manner," while others might argue that profitability is not part of a requirement.

The industry has created its own rules and vernacular that includes requirements, constrains, criteria, etc. Don't get hung up on the subtle difference between these terms. What you need to end up with is a set of statements that define what results are wanted and by whom segregated into "must have" (deal breakers) and "must have if available and affordable" (negotiable), a ranking by priority, and finally rough expectations of cost and practicality.

By definition, you would think that a "requirement" must be satisfied. But reality is that requirements come from a collection of stakeholders with different motivations, fears, and concerns. Some are more perceived than real. They often conflict with one another, and some are more important than others. As part of cataloging and ranking the requirements, you will need to resolve the conflicts.

You can catalog and track this kind of information in a spreadsheet, database, or an intranet web-based collaboration tool. There are some specialized "requirements management" type products available as well. How you do this is not important, but it is important that it is well done and documented.

In all probability, by the time an organization funds a site search, they have a pretty good idea of what they want to accomplish and have something akin to a requirements list. Key to being successful will be flushing out this list, eliminating requirements that are not supportable by the business, adding requirements that are missing, and building a consensus among the decision makers.

A good place to start is to look at why you are searching for a new site and establishing if expectations of the benefits that will be gained from the endeavor are reasonable. Even if the reasons for the search seem obvious, there are always alternatives and all too often, when the final cost estimate is put together, the program changes. An organization may elect to refurbish an existing data center rather than relocate, may decide that a new data center cannot be supported by the customer base (sales revenues), or may find that despite tall tales of cheap electrical power deals, cheap power is a fleeting opportunity at best.

This first high-level pass should be based on what we know today and on any additional information that can be gathered in a timely and inexpensive manner. If the key business requirement is to reduce operating expenses and if the belief is that cheaper power is available elsewhere, then a purchased power survey/report and industry average construction costs might be enough for a first-round estimate of the potential annual energy savings and a potential payback. This effort might confirm, change, or disprove the savings expectation and reason for the site search, saving the organization countless hours and expense.

Each requirement and expectation should be challenged in this manner, as well as assumptions about executive and business unit support. Don't assume that what is in the interest of one business unit is in the best interest of all business units. Spend some time and effort confirming the initial project program with the stakeholders (i.e., anyone who has a credible interest in the outcome). This may include senior management, business units that will use the facility, customers, consultants, etc. The goal is not to kill the site search before it starts, but a search that starts badly usually ends up badly.

It is surprising, and all too frequent, that halfway through the site search selection process, the question of what the organization really needs or wants gets raised as it was never suitably addressed in the beginning. This kind of questioning down the home stretch will lead to delays and missed opportunities. It can also cause confusion among your stakeholders and with the local and economic development leaders that are after your business.

The key here is "plausible." Beginning the search with too specific or unrealistic criteria risks eliminating potential sites and tends to get the search bogged down with detail that often isn't available until further along in the process. On the other hand, beginning the search with too few or incomplete criteria wastes a lot of everyone's time. When you don't sound like you know what you are doing, people are less responsive and don't offer their best.

No matter how hard you try to build a great team (Fig. 5.2), you will have some team members that are not of your choosing, some of whom will be uncooperative and will present site selection criteria that ensure the search will not be successful. It is important that inappropriate criteria do not make it to your site selection criteria list. Also, once you begin the search process, it is not uncommon to find that a certain requirement is not easily met or that it is prohibitively expensive to meet, and there may be other reasons to redress the requirements.

Be cautious and exercise due diligence when a change is imitated, but keep in mind that the goal is not to adhere to a statement of requirements at all cost but to make the right decision for the business. Business conditions do change, and the original requirements may have been misdirected in the first place or are perhaps no longer appropriate.

5.4.2 Round 1—Define Geographic Search Area

Geographic area is generally the first requirement, or a constraint, that the search team will use to narrow the universe of possibilities. Why? First, because the process is one of elimination and geographic area generally eliminates more potential sites than any other constraint. Second, the research and effort required is practically nil. Connecting these two points, there is simply more bang for the buck with this approach.

The geographical search area can be anything reasonable. Keep in mind that there are often legitimate requirements or constraints that may not be directly cost related and aren't easily estimated or accounted for.

- A power company or other regulated utility might define their search area as their service territory because locating outside their territory might be considered both politically incorrect and akin to suicide with the public service commission or other regulators.
- A bank with sensitive customer credit data might constrain their search to the locations in their home country based on regulatory requirements that certain data be maintained in-country.
- An international business such as Coca-Cola or Home Depot might look to any country where sales are substantial and contribute to their annual income, in an effort to give something back to the community.

Typical site selection team members

- **Program manager**
- **Stakeholders**
  - Customers
  - Employees
  - Stockholders
  - Business unit manager
  - C-level sponsors
- **Consultants and vendors**
  - Site selection consultant
  - Architect and engineer(s)
  - Power and energy procurement broker
  - Network and communications specialist
  - Analyst-cost accountant
  - Tax accountant
  - Attorney
  - Construction contractor
  - Data center move consultant
  - HR specialist

Typical site selection considerations

- **Geopolitical**
  - Stable government
  - Business friendly government
  - Favorable tax rates and policies
  - Favorable energy production and use regulations
- **Infrastructure**
  - Fiber networks
  - Power capacity and reliability
  - Water and sewage systems
  - Ease of travel (roads and airports)
  - Municipal services such as police and security, fire protection, medical, and health care
  - Availability of business and support services such as construction contractors and preventative maintenance vendors
- **Operating expenses**
  - Low energy cost
  - Environmental opportunities for energy savings
  - Low property and income taxes
  - Government funded training programs
- **Low risk of local and regional disaster**
  - Not subject to natural disasters: hurricane, tornado, monsoons, flooding, earthquakes, landslides, wild fires, etc.
  - Proximity to transportation arteries (rail, highway, and air)
  - Proximity to petrochemical plants
  - Proximity to nuclear plants
- **Quality of life**
  - Low cost of living and home prices
  - Short commute times
  - Employment and educational opportunities for spouses and children
  - A climate that provides residents a year-round playground with plenty to do—mountains, lakes, rivers and beaches
  - Entertainment including major league sports and theater
  - World-class cultural attractions
  - Exciting night life and convenience of travel

FIGURE 5.2 Typical search team members and criteria. Courtesy of K.J. Baudry, Inc.

- A regional organization that will relocate a number of employees to the new site may limit the search to cities in their region with a population between 1,000,000 and 5,000,000, in an effort to make the relocation more palatable to their existing employees.
- A business in a very competitive environment where being the lowest cost provider is key to winning business might not have any geographic limitation. It simply comes down to the least expensive site available.

Geographic constraints carry a lot of risk if the impact is not fully thought out. Many organizations set geographic constraints based on proximity to existing facilities, the assumption being that there are practical benefits (cost savings) to be gained. However, there is little basis for this assumption. Any operational support or convenience that might be available from the headquarters, office, warehouse, or other facilities in the immediate vicinity is often insignificant when compared to other savings such as energy cost and tax incentives that might be available outside the immediate vicinity. There also may be existing facilities in locations outside of the immediate vicinity that are available because of an unrelated merger or acquisition, bankruptcy, etc. that might be leased or purchased at substantial discount to the cost of new construction.

Once you have the search area, you will want to develop a list of potential communities. There aren't a magical number of prospective communities. Given the costs involved, more than 20 opportunities are probably excessive. Your site search consultant may suggest eliminating some communities based on relevant experience in dealing with them. With fewer than a dozen, you might not end up with any acceptable candidates as you cut the list down, applying more detailed criteria.

5.4.3 Round 2—Site Selection Criteria

We have opted to present this information as taking place in a serial manner for ease of presentation. Depending on your schedule and how committed (willing to spend money) the organization is, there is some benefit to developing the statement of requirements and selection criteria as a contiguous event in advance of starting the search. With this in mind, selection criteria that require little research might be incorporated as part of Round One. The idea being to eliminate as many possibilities before the in-depth (expensive) research is required.

The following sections identify concerns that are common to many data center users when searching for a site. The list is not comprehensive but covers the major areas. Some of the items may not be applicable, depending on the type of organization and industry which you operate in.

The idea is that for each of these items you assess how they affect your organization and if they are potentially impacting, identify how they can best be dealt with.

5.4.3.1 Political Environment Most national governments, and especially emerging economies, have reduced barriers to market entry, property ownership, and deregulated privatized industries and encourage capital investment. When looking overseas, the chances are that you will receive a warm welcome. However, there may be significant challenges in the following areas:

- Security
- Laws
- Regulatory
- Employment
- Property Ownership and Investment
- Censorship

5.4.3.2 Quality of Life for Employees Data centers can operate with a few employees. In many cases, quality-of-life issues might not be important, as employees will be hired locally. However, if an organization plans on relocating employees, quality-of-life issues will become important in retaining employees. What are people looking for? Some combination of the following:

- Low home price, taxes, and energy cost
- Short commute times
- Employment and educational opportunities for spouses and children
- A climate that provides residents a year-round playground with plenty to do—mountains, lakes, rivers, and beaches
- Entertainment including major league sports and theaters
- World-class cultural attractions
- Exciting night life
- Convenience of travel, etc.

There are numerous publications such as *Forbes*, *Business Week*, and *Time Magazine* that create "top ten" lists. Best cities for college graduates, best cities for technology, best cities for entrepreneurs, most high-tech employment, etc., make good sources for this kind of information.

5.4.3.3 Business Environment Site selection is largely about taxes and other costs that are a large part of operating expenses. Taxes come in all shapes and sizes, vary in how they are calculated, and are applied from state to state. Historically, some communities have looked at the economic benefits that data centers provide in terms of direct and indirect jobs created (payroll), investment, and taxes and have identified data centers as good sources of tax revenue. Data centers pay a lot in taxes and demand very little in terms of publicly provided services. They do not use significant amounts of sewage capacity, generate trash, fill up the local schools, and don't require new roads or extra police services. For a politician, it's a significant source of new "unencumbered" income.

The largest single category of taxes tends to be property taxes. In many communities, both the value of the real property, land, and buildings and the value of the fixtures, furnishings, and equipment are taxed. When you consider the cost of the facility is often well over $1000/sf and that the IT equipment installed can easily exceed this figure, even a low tax rate can result in a significant annual outlay.

Local communities typically have offered incentives in terms of tax abatement on property taxes and reduced sales taxes on equipment purchased for installation in the facility. Programs will vary, but almost all communities phase the incentives out over a period of years. Other state and local incentives may include an expedited permit process, land grants, improved road access, extension of utilities, tax rebates, and financing through industrial revenue bonds. There may also be Community Development Zones and associated development grants. Many states offer job credits and training programs through local community colleges.

You will need to calculate the economic benefit of the incentive package and include it in your overall analysis. A word of caution: a significant percentage of incentives, well over 50%, are never collected due to failure on the part of the organization to follow through after the facility is opened or due to changed economic conditions, overenthusiastic rates of growth, etc. For example, a delay in growth that pushes large purchases of equipment beyond the first couple of years could result in significant reduction in tax savings if the incentive was highest in the first year and fell off over successive years.

When considering overseas locations, there are differences in the way taxes are levied, and it will affect your cost structure. This is one of the reasons for having an accountant on your team who can evaluate the implications for your economic modeling.

Politicians, local officials, and economic development groups love to make a splash in the news headlines by announcing big deals, and data centers tend to be big deals in terms of dollars invested. But it is a two-way street, and it will be up to you to show them the value that you bring to the community. In short, the better you sell yourself, the more successful you will be at winning incentives.

5.4.3.4 Infrastructure and Services While airports, roads, water, sewage, and other utilities are all important, two utilities are showstoppers. A data center must have electrical power and telecommunications.

Telecommunications engineers often represent networks as a cloud with an access circuit leading in/out of the cloud at the "A-end" and in/out of the cloud at the "Z-end." It's a bit of a simplification but is a good representation. The in/out circuits are called end or tail circuits and are provided by a Local Exchange Carrier (LEC). They typically run to an exchange where traffic can be passed to long-haul carriers.

Depending on how you purchase bandwidth, you may make a single purchase and get a single bill, but it is very likely that your traffic is carried over circuits owned by several different carriers. Having more than one carrier available means that there is always a competitive alternate carrier who wants your business. At one time, network connections were priced on bandwidth miles; but today, supply and demand, number and strength of competitors in the local market, and available capacity all factor into the cost of bandwidth. The only way to compare bandwidth cost between site options is to solicit proposals.

While small users may be able to use copper circuits, T1s and T3s, many organizations will require fiber medium services, OC192 s or 10 Gb carrier grade Ethernet. LECs typically build fiber networks using self-healing networks, such as Synchronous Digital Hierarchy (SDH) or Synchronous Optical Network (SONET) rings. These are very reliable. Typically arrangements can be made to exchange traffic to long-haul networks at more than one exchange making the system very reliable and resilient.

Many organizations require that there be two LECs available. Reasons for this include reliability, pricing, and perhaps a lack of understanding about networks. In general, site search and selection tends to be initiated and led by the financial side of the organization, and most organizations do not involve their network engineers in the site search and selection process. However, network connectivity is one of the fundamental requirements of a data center. It is a key cost and performance issue and a network engineer should be considered a critical part of one's team.

Electrical power is the other big piece of infrastructure that is a must have. Smaller data centers with less than 5 MW (megawatts) of load can generally be accommodated in most large industrial and office parks where three-phase service exists. Larger data centers require significant amount of power and may require planning with the power company and the upgrading of distribution lines and substations and, in some cases, construction of dedicated substations. All of this must be addressed with the power company prior to the site selection.

Energy can make up as much as a third of the total occupancy cost, and more if the rates are high. Almost all power companies use the same basic formula for setting rates: recover the capital cost of serving the customer, recover the cost to produce the energy, and make a reasonable return for the stockholders. Understanding this is important to negotiating the best rates. Yes, most utilities can negotiate rates even if they are subject to public service commission regulation, and regulated utilities can be every bit as competitive as nonregulated utilities!

If you want to get the best rate, you need to know your load profile and you need to share it with the power company. Your design may be 200 W/ft^2 and that is great, but that's not your load profile. Your load profile has to do with how much energy you will actually consume and the rate at which you will consume it. Retaining a consultant who understands rate tariffs and can analyze your load is important to negotiating the best rates.

Negotiating rates is best done in a cooperative manner, sharing information as opposed to the more traditional negotiation stance of sharing little, demanding a lot, and constantly threatening to take your business elsewhere. Demanding oversized service only results in the utility spending more money to serve you and more money has to be recovered from you in upfront capital contribution or through higher rates; therefore, properly sizing the service results in the most competitive rates. Further, most utility companies, either because of regulation or because of board governance, will not knowingly provide you with below-cost rates and then make it up by charging another customer higher rates.

Climate is not part of the infrastructure but makes the checklist twice: first, as a factor that affects your energy cost and second as a quality-of-life issue. Your energy cost is dependent on the rate, but it is also dependent on how much energy you use. Recent thinking has led air-conditioning engineers to consider ways of reducing cooling costs such as air-side and water-side economization. The potential savings are greatly increased when the outside air is cool and dry for substantial parts of the year.

However, there is more that should be considered than just cost; the capability of the utility and responsiveness when and if an emergency should occur is also important. A short example is probably worth more than a dozen paragraphs. In September 2005, Hurricane Katrina devastated the Mississippi coast. Mississippi Power, a small operating unit of Southern Company with 1,250 employees, restored power to almost 90% of their customers within 11 days (10% were too decimated to receive power). They rebuilt 300 transmission towers, over 8,000 poles, and 1,000 miles of overhead lines against all odds by bringing in a workforce of over 10,000, providing temporary housing in large circus tents, food and water, over 8,000 tetanus shots, and burning 140,000 gallons of diesel fuel a day. A small municipal utility, affiliated with a regional utility, may not be able to come through when the going gets tough. The preparation and logistics necessary to maintain operations during a disaster gets exponentially more difficult as the duration becomes longer.

There are three concepts in business continuity that need to be considered in the site selection process: Walk To, Drive To, and Fly To. In major regional outages such as earthquakes and hurricanes and after 9/11, it became apparent very quickly that accessing to a primary or fall back site can be challenging if not impossible. Roads become blocked with traffic or debris, and even air travel may become curtailed. If your business continuity planning requires that your data center continues operating in an emergency, then it becomes important to have multiple means of transportation.

It is a bit unusual to think of maintenance vendors as part of the local infrastructure. However, the role of preventive maintenance is of as much importance as it is to have redundant systems, and perhaps even more important. While it is important that a vendor know everything there is to know about their equipment, they must also be familiar with and have the discipline to work in data center environments. If you are the only data center in the region, you may not find suitable maintenance vendors locally. If qualified maintenance vendors are not available, you will need to consider how this will affect your normal maintenance program as well as your ability to recover from a failure in a timely manner. Operator errors, including errors made by maintenance vendors, account for a significant percentage of all failures. Having other significant data center operations in the area is a good indication that maintenance vendors are available.

Finally, if there are other data centers in the area, ask the local economic development authority to introduce you. The odds are that they played some role in their site selection decision and will already have connections within the organization. Buying one of these contacts lunch is probably the best investment you can make in the search process.

5.4.3.5 Real Estate and Construction Opportunities Depending on your requirements, you may be looking for an existing facility to purchase or lease, for pad sites within a development or for raw land. Regardless of the need, you will want to make sure that there are a sufficient number of opportunities.

During the dot-com boom, a new breed of business, co-location, was created. Aiming at the outsourcing of information technology needs by corporate America, and armed with easy money, these companies built large, state-of-the-art facilities across the country, some at a cost of \$125 M or more and as large as 300,000 ft^2. Many of these companies failed, some because of a poor business plans and others because they were simply ahead of their time. These facilities were placed on the market at substantial discounts to construction cost. Some were practically given away.

Today, these types of opportunities are rare, but occasionally companies merge or change computing strategies to find that they now have excess data center facilities. Other facilities become available because of lease expirations, growth, changing technologies, etc. While these may be great opportunities for some companies, they are often based on outdated standards and were built prior to the current focus on energy costs. It may not be necessary that a facility meet "current standards," but it must meet your standards and needs. Great care should be taken to properly assess the opportunity and cost to upgrade the facility if necessary.

It is not uncommon to find industrial or mixed-use parks advertising themselves as a "data center opportunity." This may mean that the developer has already worked with the local power utility, brought multiple fiber carriers to the park, negotiated for tax abatement, and taken other steps to prepare the site for a new data center. However, it often doesn't signify anything more than someone could build a data center on this site if they choose to.

When selecting a community, it is important that there are multiple sites available. These sites should be competitive (owed by different investors) and suitable for constructing a data center. It is important that your team verify any claims made by the owner, economic development agency, or other organizations trying to attract your business.

If there is a single existing facility or only one site available, then the negotiations with the various parties involved need to be tied together for a simultaneous close or with dependencies written into any purchase agreements so that the failure of any one part invalidates any other agreements.

Construction costs tend to be relatively uniform, and there are construction indexes readily available that can be used to estimate differences between communities or geographic regions. Development costs, permitting process, and construction requirements may vary between communities and, in some cases, might represent a significant impediment to your project, especially in terms of schedule. If you have a target date in mind, then finding a community that is willing to expedite the permit process may be important.

5.4.3.6 Geography, Forces of Nature, and Climate Avoiding forces of nature is one area where there is often a discrepancy between what people say and do. This is due in part to a checklist mentality: the idea that by checking off a generic list of standard issues, we can avoid during our own legwork.

The general recommendation is to avoid risk due to forces of nature. Yet, California is home to a large number of data centers and continues to capture its share of growth despite an elevated risk of seismic activity. Almost half of the continental United States falls within a 200 mph or greater wind speed rating and is subject to tornado activity. Yet these areas continue to see investment in new data center facilities. Finally, East Coast areas with elevated risk of hurricane activity such as around New York and Washington, DC, continue to be desirable.

There are a couple of reasons why this happens. First, if there are compelling reasons to be located in a specific area, the risk may be substantially mitigated through good design practices and construction. Second, many individuals in key decision-making positions cling to the idea that they need to be able to see and touch their IT equipment. For these people, sometimes referred to as "server huggers," the idea of a remote lights out operation entails more risk than the local forces of nature.

Regardless of the reason, it is important to assess the level of risk, identify steps to mitigate the risk, estimate the cost of risk mitigation, and include that cost in your financial modeling. Depending on whom you ask, Natural Disasters account for anywhere from 1% to almost 50% of all major data center outages. Another report puts Power-Related Causes at 31%, Weather and Flooding (including broken water lines) at 36%, Fire at 9%, and Earthquake at 7%. A lot depends on the definition of a disaster, the size of data center under consideration, and what product or service is being promoted. The fact is that many data center outages are avoidable through good site selection, proper planning, design, and maintenance.

Design and construction techniques that effectively mitigate risk of damage from earthquakes, tornados, hurricanes, and flooding are well known and, while expensive, can be economically feasible when considered as part of the total cost.

While protecting your facility may be feasible, it is generally not possible to change the local utility systems and infrastructure. Many seemingly well-prepared organizations have found out the hard way that it is easy to end up being an island after a regionwide event such as a hurricane. An operating data center is worthless without outside communications and without its resupply chain in place (fuel, food, and water).

5.4.3.7 Manmade Risks Avoiding manmade risks ranks up there with avoiding natural disaster, and the variance between what one asks for and what one does is even greater with manmade risks. Requirements that we often find on checklists include the following:

- Two miles from an airport
- Two miles from a broadcast facility
- Four miles from a major state highway or interstate highway
- Five miles from a railroad
- Ten miles from a chemical plant
- One hundred miles from a Nuclear Facility

Many of the items on this list appear obvious and reasonable on first review. However, trains and tankers full of hazardous material traverse our railroad tracks and highways every day of the week and at all hours of the day and night. So being a minimum distance from potential accidents makes sense. But most accidents are not contained within the highway right of way, and when you consider that released toxic chemicals can be transported by wind for miles, you realize that one-half mile, two miles, or four miles do not substantially change the risk.

Now consider where fiber and power is readily available. Utility providers built their networks where they expect customers to be, in industrial, mixed-use, and industrial parks (i.e., where there is commerce), and where there is commerce, there is manmade hazards.

It is important that all risks be considered and they should be considered from both short- and long-term points of view. It is even more important that the nature of the risk, relationship between the risk and distance, and the potential impact be understood.

5.4.4 The Short List—Analyzing and Eliminating Opportunities

The site search and selection process is built around elimination and, to be most effective, we apply broad brush strokes in order to eliminate as many sites as possible. In doing this, we take a risk that we might eliminate some good opportunities. We eliminate some communities because they have significant shortcomings and others because they might have simply lacked the preparation and coordinated effort to provide a credible response to our request for information.

We are at a critical point in the process; we have eliminated most of the sites and are down to the short list. Each remaining community will need to be visited, the promotional claims and statements made during the initial rounds researched and supported, specific proposals requested, high-level costs estimates brought down to specific costs, and agreements negotiated. If there is a specific facility that figures into the site selection decision, you may want to have an architectural and engineering team evaluate the facility, create a test fit, identify the scope of work, and prepare a construction cost estimate (Fig. 5.3). This is a considerable effort for each site, and most organizations will want to limit the number of sites to three, the two front-runners and a replacement.

At this stage, it is inappropriate to eliminate a site without identifying a significant shortcoming. Earlier in the chapter, we stated that there was no such thing as the best site, just the best combination of site characteristics. In light of that idea, sites can be flawed but fixable or flawed and unfixable. In most cases, the reason for conducting an extensive site search is to minimize costs, and a site that has a fixable flaw, even if it is expensive to fix, might be the least expensive opportunity due to other line items. Consider a community without a competitive LEC. Could another carrier be attracted to the area because of your business? Could you negotiate a long-term agreement with the existing carrier to assure good service and pricing?

If you identify an unfixable flaw early enough in the process, it may be worthwhile eliminating the opportunity from the shortlist and moving one of the lower-ranked sites up to the list. It may be that a promise or claim made by the

Comparison of sites

	Selection criteria	Weighting	Criteria ratings				
			Option A	Option B	Option C	Option D	Option E
1	Ability to meet schedule	4	2	5	2	4	1
2	Availability of property	3	3	5	3	3	1
3	Cost of property	3	3	4	4	3	1
4	Business environment—economic incentives	4	4	1	4	3	1
5	Availability of low-cost financing	3	4	1	4	3	1
6	Availability of power	5	4	3	5	4	4
7	Reliability of power system	4	4	4	5	4	4
8	Low-cost energy	5	4	3	5	4	4
9	Environmental energy savings opportunity	3	2	5	3	4	5
10	Availability of fiber networks	5	3	4	2	4	2
11	Availability of skilled vendors	3	2	4	1	4	2
12	Availability of IT labor	3	2	4	1	4	2
13	Easy to "fly to"	4	1	3	4	4	2
14	Easy to "drive to"	4	1	3	4	4	1
15	Proximity to major technical university	3	1	4	3	4	1
16	Local job opportunities for family members	3	1	3	3	4	0
17	Quality of life	4	1	4	3	4	1
18		0	0	0	0	0	0
19		0	0	0	0	0	0
20		0	0	0	0	0	0
Total weighted score			**161**	**220**	**214**	**239**	**129**

Weighting—The default weighting is 2.5 and is the mean rating.

Rating—Ratings are high (5) to low (1).

Score—Score is calculated as "weighting × rating". For example, a weighting of 3 and a rating of 2, produces a score of 6 (3×2). The resulting score is an indication of how well a vendor meets the weighted selection criteria. The higher the score, the better the vendor meets the criteria.

FIGURE 5.3 Simple visual presentation of sites. Courtesy of K.J. Baudry, Inc.

development agency was not correct or turns out to be incomplete. This can and does occur because of deception, but for legitimate reasons as well. The community may have expected a bond referendum to pass that didn't; another company may step in front of you and contract for the available power; etc.

If a specific facility is part of the selection decision, some of the business requirements will need to be translated into technical or construction requirements. This can be a challenge. The bulk of our efforts will have been financial. This does not change with technical requirements but does take on a new complexion. It is not uncommon for the engineering members of your team to designate specific properties as unacceptable, especially when dealing with upgrading an existing data center or retrofitting an existing structure. In many cases, this will mean that the engineer perceives one or more aspects as too difficult or expensive.

It is important that questions of suitability, especially when they potentially remove a site from the list, be questioned and thoroughly reviewed. Often the cost, when taken into perspective, is not significant, or that the many cost savings features of the site easily offset the added cost. Not every member of your team will be fully aware of the overall cost and potential benefits that a community or facility brings to the program.

5.4.5 Closing the Deal

The process is complete when the site selection is announced. The site selection choice may be made long before it is announced and long before the process is complete, and it is important that at least two communities be kept in the running until all concession and agreements are in place and the deals are executed. The reason is very simple; you lose all negotiating power once all the other sites are eliminated. So despite the effort involved, it is important to continue to pursue your number one and two choices until the very end. Just like secrecy, the team members are the ones most likely to let out the decision through innocent conversations with power company employees, local consultants, permitting agencies, and other entities that are involved in your due diligence.

5.5 INDUSTRY TRENDS AFFECTING SITE SELECTION

5.5.1 Globalization and Political and Economic Reforms

Globalization has been a consistent force since the beginning of mankind. It has moved slowly at times and with great speed at other times. The past 30 years have seen great advancements. Trade barriers have fallen and governments have sought out investment from other countries. We may be on the verge of a temporary slowdown as governments move to protect their turf during economic downturns.

Globalization will continue as long as there are opportunities to sell more products and services and as long as there are differences in the cost of key resources such as utilities, differences in energy policies, and taxes.

5.5.2 Global Strategic Locations

While the world keeps getting smaller and smaller, it is still a pretty big place. Every search needs a starting point. Traditional economic criteria, such as per capita income, cost of living, and population might be applicable. Perhaps, locations where your competitor has facilities or maybe where major international IT companies are located such as Facebook, Google, HP and IBM.

Google, according to www.datacenterknowledge.com, has facilities in over a dozen US cities: Mountain View, Pleasanton, San Jose, Los Angeles, Palo Alto, Seattle, Portland, Dallas, Chicago, Atlanta, Reston, Ashburn, Virginia Beach, Houston, Miami, Lenoir (NC), Goose Creek (SC), Pryor (OK), and Council Bluffs (IO); and in many cities outside the United States: Toronto, Berlin, Frankfurt, Munich, Zurich, Groningen, Mons, Eemshaven, Paris, London Dublin, Milan, Moscow, Sao Paolo, Tokyo, Hong Kong, Taiwan, and Singapore. It is important to note that these locations are not all major data centers. Some may simply be peering points. For more information on Google Data Centers, see http://www.google.com/about/datacenters/locations/index.html#.

In addition to traditional criteria, we offer one more that might be worthwhile considering: volume or other measures of IP exchange traffic.

The largest exchanges in the United States are the New York Area, Northern Virginia, Chicago, San Francisco Bay Area, Los Angeles Area, Dallas, Atlanta, and Miami.

The largest exchanges outside the United States include Frankfurt, Amsterdam, London, Moscow, and Tokyo. A full list of exchanges can be found on Wikipedia by searching on "peering" or "list of IP exchanges."

5.5.3 Future Data Centers

Today, we design and build data centers from a facility-centric approach. They are conceived, designed, and operated by facility groups, not by the real customer, IT groups. Facilities design has had hard time keeping up with the rapidly changing IT environment. Our facility designs have changed little in the past 20 years. Sure, we have adapted to the need for more power and cooling, and there have been incremental improvements. Energy efficiency has become important. But by and large, we have simply improved old designs incorporating time-proven techniques used in other types of facilities such as industrial buildings, schools, hospitals, and offices.

Today, every data center is a unique, one-off design to match an organization's unique requirements and local conditions. Tomorrow's data center will need to be designed with a holistic approach, one that marries advancements in IT technology and management with facilities design and life cycle economics.

One approach that we expect in the future will be a throwaway data center with minimal redundancy in the mechanical and electrical systems. A large appliance, inexpensive enough to be deployed at sites around the world selected based on business-friendly governments and inexpensive power.

Equipped with self-healing management systems, meshed into a seamless operation, populated with IT systems prior to deployment, these systems would serve their useful life and the data center would be decommissioned.

With sufficient redundancy and diversity of sites, any one site could be off line due to planned or unplanned maintenance. Built to be taken off line and redeployed based on changing economic needs such as customer demand and short-term energy and fiber transmission services. At the end of life, it could be refurbished in the field, shipped back to a central depot for refurbishment, or sold to a low-end non-competing organization.

The combination of the reduction in initial facility development cost, mass production to a consistent standard and performance level, speed of deployment, and the flexibility to meet changing economic environments is a very attractive package. The ability to implement this approach largely already exists for the most part, considering the following:

- Light out operations already exist. Management systems have progressed to a point where most operations can be performed remotely. They allow applications to exist at multiple locations for production and backup purposes, for co-production, for load balancing, or to meet performance objectives.
- IT platforms are getting smaller and smaller. Servers have progressed from several rack units (RUs) to one RU and blades. You can fit as many as 70 blades servers in a single rack. Processing capacity is almost doubling every 3 years, and virtualization will greatly increase the utilization of the processing capacity.
- While the amount of data we store today has increased tremendously, the density of storage systems has increased tremendously and will continue to do so. In 1956, IBM introduced the random access method of accounting and control (RAMAC) with a density of 2000 bits/in.^{2} Today's latest technology records at 179 Gigabits/in.^{2} (Sony). That's from 2,000 to 179,000,000,000 bits/in.^{2} in 52 years.

A containerized data system could be configured to provide the same level of computing that is provided in today's average data center of perhaps 10–20 times its size.

There are advantages from the facilities perspective as well. Data centers are built for a specific purpose. We struggle to increase the availability of the physical plant and every improvement comes at an ever increasing expense (the law of diminishing returns). No matter how much redundancy we design into the facilities plant, we cannot make one 100% available. By comparison, inexpensive data center appliances with dynamic load balancing and redundant capacity deployed throughout the world would not need five nines of availability to be successful.

According to The Uptime Institute, the cost of a Tier III facility is generally believed to be $200/sf and $10,000/kW or higher. While the building does depreciate, it's the physical plant (the $10,000/kW) that becomes obsolete over time. In a manner of thinking, we are already building throwaway data centers, at a great expense and without a good exit strategy!

There has always been a mismatch between the investment term for the facility and IT systems that the facility houses. IT refresh cycles are typically 3 years (probably more likely 5 years for most organizations), yet facilities are expected to last 15 years or more. This mismatch between investment terms means that data centers have to be designed to accommodate an unknown future. This is an expensive approach.

While such a concept will not be suitable for companies that are the largest of centers, it could be very beneficial for smaller operators, the very clients that collocation centers are seeking to put in their centers today.

FURTHER READING

Best Practices Guide for Energy Efficient Data Center Design, EERE, DOE. Available at http://www1.eere.energy.gov/femp/pdfs/eedatacenterbestpractices.pdf. Accessed on March 2011.

National Oceanic and Atmospheric Administration. Available at http://www.nhc.noaa.gov/. Accessed on June 11, 2014.

National Weather Service. Available at http://www.weather.gov/. Accessed on June 11, 2014.

Rath J. Data Center Site Selection, Rath Consulting. Available at http://rath-family.com/rc/DC_Site_Selection.pdf. Accessed on June 11, 2014.

The Uptime Institute. Available at http://uptimeinstitute.com/. Accessed on July 4, 2014.

U.S. Geological Survey. Available at http://www.usgs.gov/. Accessed on June 11, 2014.

6

DATA CENTER FINANCIAL ANALYSIS, ROI AND TCO

Liam Newcombe
Romonet, London, UK

6.1 INTRODUCTION TO FINANCIAL ANALYSIS, RETURN ON INVESTMENT, AND TOTAL COST OF OWNERSHIP

Anywhere you work in the data center sector, from an enterprise business that operates its own data centers to support business activities, a colocation service provider whose business is to operate data centers, a cloud provider that delivers services from data centers, or for a company that delivers products or services to data center operators, any project you wish to carry out is likely to need a business justification. In the majority of cases, this business justification is going to need to be expressed in terms of the financial return the project will provide to the business if they supply the resources and funding. Your proposals will be tested and assessed as investments; and therefore, you need to be able to present them as such.

In many cases, this will require you to not simply assess the overall financial case for the project but also deal with split organizational responsibility or contractual issues, each of which can prevent otherwise worthwhile projects from going ahead. This chapter seeks to introduce not just the common methods of Return on Investment (ROI) and Total Cost of Ownership (TCO) assessment, but also how you may use these tools to prioritize your limited time, resources, and available budget toward the most valuable projects.

A common mistake made in many organizations is to approach an ROI or TCO analysis as being the justification for engineering decisions that have already been made; this frequently results in the selection of the first project option to exceed the hurdle set by the finance department. To deliver the most effective overall strategy, project analysis should consider both engineering and financial aspects to identify the most appropriate use of the financial and personnel resources available. Financial analysis is an additional set of tools and skills to supplement your engineering skill set and enable you to provide a better selection of individual projects or overall strategies for your employer or client.

It is important to remember as you perform or examine others' ROI analysis that any forecast into the future is inherently imprecise and requires us to make one or more estimations. An analysis that uses more data or more precise data is not necessarily any more accurate as it will still be subject to this forecast variability, precision should not be mistaken for accuracy. Your analysis should clearly state the inclusions, exclusions, and assumptions made in your TCO or ROI case and clearly identify what estimates of delivered value, future cost, or savings you have made, what level of variance should be expected in these factors, and how this variance may influence the overall outcome. Equally, you should look for these statements in any case prepared by somebody else, or the output is of little value to you.

This chapter provides an introduction to the common financial metrics used to assess investments in the data center and provides example calculations. Some of the common complications and problems of TCO and ROI analysis are also examined, including site and location sensitivity. Some of the reasons why a design or project optimized for data center A is not appropriate for data center B or C and why the vendor case studies probably don't apply to your data center are considered. These are then brought together in an example ROI analysis for a realistic data center reinvestment scenario where multiple options are assessed and the presented methods used to compare the project options.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

The chapter closes with a discussion from a financial perspective of likely future trends in data centers. The changing focus from engineering to financial performance accelerated by the threat of cloud and commoditization is discussed along with the emergence of energy service and guaranteed energy performance contracts. A sample of existing chargeback models for the data center is reviewed and their relative strengths and weaknesses compared. The impact on data centers of the current lack of effective chargeback models is examined in terms of the prevalent service monoculture problem. The prospect of using Activity-Based Costing (ABC) to break out of this trap provides effective unit costing and fosters the development of a functioning internal market for enterprise operators, and per customer margin management for service providers is examined. The development from our current, energy-centric metric, PUE, toward more useful overall financial performance metrics such as cost per delivered IT kWh is discussed and, last, some of the key points to consider when choosing which parts of your data center capacity should be built, leased, colocated, or deployed in the cloud are reviewed.

This chapter provides a basic introduction to the financial analysis methods and tools; for a more in-depth treatment of the subject, a good management finance text should be consulted such as Wiley's *Valuation: Measuring and Managing the Value of Companies* (ISBN 978-0470424704).

6.1.1 Market Changes and Mixed ICT Strategies

Data centers are a major investment for any business and present a series of unusual challenges due to their combination of real estate, engineering, and IT demands. In many ways, a data center is more like a factory or assembly plant than any normal business property or operation. The high power density, high cost of failure, and the disconnect between the 20+ year investment horizons on the building and major plant and the 2–5-year technology cycle on the IT equipment all serve to make data centers a complex and expensive proposition.

The large initial capital cost, long operational cost commitments, high cost of rectifying mistakes, and complex technology all serve to make data centers a relatively specialist, high risk area for most businesses. At the same time, as data centers are becoming more expensive and more complex to own, there is a growing market of specialist providers offering everything from outsourced management for your corporate data center to complete services rented by the user hour. This combination of pressures is driving a substantial change in the views of corporate CFOs, CIOs, and CEOs on how much of their IT estate they should own and control.

There is considerable discussion in the press of IT moving to a utility model like power or water in which IT services are all delivered by specialist operators from a "cloud" and no enterprise business needs to own any servers or employ any IT staff. One of the key requirements for this utility model is that the IT services are completely homogeneous and entirely substitutable for each other, which is clearly not presently the case. The reality is likely to be a more realistic mix of commercial models and technology.

Most businesses have identified that a substantial part of their IT activity is indeed commodity and represents little more than an overhead on their cost of operating; in many cases, choosing to obtain these services from a specialist service provider is a sensible choice. On the other hand, most businesses also have something that they believe differentiates them and forms part of their competitive advantage. In a world where the Internet is the majority of media for customer relationships and more services are delivered electronically, it is increasingly common to find that ICT is an important or even a fundamental part of that unique, competitive advantage. There are also substantial issues with application integration when many independent providers of individual specific service components are involved as well as security, legal, risk, and regulatory compliance concerns. Perhaps the biggest threat to cloud adoption is the same vendor lock-in problem businesses currently face with their internal applications where it is difficult or impossible to effectively move the data built up in one system to another.

In reality, most enterprise businesses are struggling to find the right balance of cost, control, compliance, security, and service integration. They will find their own mix of in-house data center capacity, owned IT equipment in colocation facilities, and IT purchased as a service from cloud providers.

Before any business can make an informed decision on whether to build a service in their own data center capacity or outsource it to a cloud provider, they must be able to assess the cost implications of each choice. A consistent and unbiased assessment of each option that includes the full costs over the life cycle is an essential basis for this decision that may then be considered along with the deployment time, financial commitment, risk, and any expected revenue increase from the project.

6.1.2 Common Decisions

For many organizations, there is a substantial, and ever growing, range of options for their data center capacity against which any option or investment may be tested by the business:

- Building a new data center
- Capacity expansion of an existing data center
- Efficiency improvement retrofit of an existing data center
- Sale and lease-back of an existing data center
- Long-term lease of private capacity in the form of wholesale colocation (8+ years)
- Short-term lease of shared capacity in the form of retail colocation
- Medium-term purchase of a customized service on dedicated IT equipment

- Medium-term purchase of a commodity service on dedicated IT equipment
- Short-term purchase of a commodity service on provider-owned equipment

For each project, the relative costs of delivery internally will increasingly need to be compared with the costs of partial or complete external delivery. Where a project requires additional capital investment to private data center capacity, it will be particularly hard to justify that investment against the individually lower capital costs of external services.

6.1.3 Cost Owners and Fragmented Responsibility

ICT and, particularly, data center cost is subject to an increasing level of scrutiny in business, largely due to the increased fraction of the total business budget which is absorbed by the data center. As this proportion of cost has increased, the way in which businesses treat IT and data center cost has also started to change. In many organizations, the IT costs were sufficiently small to be treated as part of the shared operating overhead and allocated across consuming parts of the business in the same way that the legal or tax accounts department costs would be spread out. This treatment of costs failed to recognize any difference in the cost of IT services supporting each function and allowed a range of suboptimal behaviors to develop.

A common issue is for the responsibility and budget for the data center and IT to be spread across a number of separate departments that do not communicate effectively. It is not uncommon for the building to be owned and the power bill paid by the corporate real estate (CRE) group, a facilities group to own and manage the data center mechanical and electrical infrastructure, while another owns the IT hardware, and individual business units are responsible for the line of business software. In these situations, it is very common for perverse incentives[1] to develop and for decisions to be made which optimize that individual department's objectives or cost at the expense of the overall cost to the business.

A further pressure is that the distribution of cost in the data center is also changing, though in many organizations the financial models have not changed to reflect this. In the past, the data center infrastructure was substantially more expensive than the total power cost over the data center lifetime, while both of these costs were small compared to the IT equipment that was typically purchased from the end user department budget. In the past few years, IT equipment capital cost has fallen rapidly while the performance yield from each piece of IT equipment has increased rapidly. Unfortunately, the power efficiency of IT equipment has not improved at the same rate that capital cost has fallen, while the cost of energy has also risen and for many may continue on its upward path. This has resulted in the major cost shifting away from the IT hardware and into the data center infrastructure and power. Many businesses have planned their strategy based on the apparently rapidly falling cost of the server, not realizing the huge hidden costs they were also driving.

In response to this growth and redistribution of data center costs, many organizations are now either merging responsibility and strategy for the data center, power, and IT equipment into a single department or presenting a direct cross-charge for large items such as data center power to the IT departments. For many organizations, this, coupled with increasing granularity of cost from external providers, is the start of a more detailed and effective chargeback model for data center services.

Fragmented responsibility presents a significant hurdle for many otherwise strong ROI cases for data center investment which may need to overcome in order to obtain the budget approval for a project. It is common to find issues, both within a single organization and between organizations, where the holder of the capital budget does not suffer the operational cost responsibility and vice versa. For example:

- The IT department does not benefit from the changes to air flow management practices and environmental control ranges, which would reduce energy cost because the power cost is owned by CRE.
- A wholesale colocation provider has little incentive to invest or reinvest in mechanical and electrical equipment, which would reduce the operational cost of the data center as this is borne by the lease-holding tenant who, due to accounting restrictions, probably cannot invest in capital infrastructure owned by a supplier.

To resolve these cases of fragmented responsibility, it is first necessary to make realistic and high confidence assessments of the cost and other impacts of proposed changes to provide the basis for a negotiation between the parties. This may be a matter of internal budget holders taking a joint case to the CFO, which is deemed to be in the business overall interests, or it may be a complex customer–supplier contract and SLA issue that requires commercial negotiations. This aspect will be explored in more detail under the section "Energy Service Contracts."

6.1.4 What Is TCO?

Total cost of ownership is a management accounting concept that seeks to include as many of the costs involved in a device, product, service, or system as possible to provide the best available decision-making information. TCO is frequently used to select one from a range of similar products or services, each of which would meet the business needs, and in order to minimize the overall cost. For example, the 3-year TCO of a server may be used as the basis for a service provider pricing a managed server, or for cross-charge to consuming business units within the same organization.

[1]A perverse incentive occurs when a target or reward program, instead of having the desired effect on behavior, instead produces unintended and undesirable results contrary to the goals of those establishing the target or reward.

As a simple example, we may consider a choice between two different models of server that we wish to compare for our data center, one is more expensive but requires less power and cooling than the other; the sample costs are shown in Table 6.1.

On the basis of this simplistic TCO analysis, it would appear that the more expensive server A is actually cheaper to own than the initially cheaper server B. There are, however, other factors to consider when we look at the time value of money and net present value (NPV), which are likely to change this outcome.

When considering TCO, it is normal to include at least the first capital cost of purchase and some element of the operational costs, but there is no standard definition of which costs you should include in a TCO analysis. This lack of definition is one of the reasons to be careful with TCO and ROI analyses provided by other parties; the choices made regarding the inclusion or exclusion of specific items can have a substantial effect on the outcome, and it is as important to understand the motivation of the creator as their method.

6.1.5 What Is ROI?

In contrast to TCO (which is cost-focused and tends to be used for service costing or where there is an already defined need to make a purchase such as a new server), a ROI analysis looks at both costs and incomes and is commonly used to inform the decision whether to make a purchase at all, for example, whether it makes sense to upgrade an existing device with a newer, more efficient device.

In the case of an ROI analysis, the goal is, as for TCO, to attempt to include all of the relevant costs, but there are some substantial differences:

- The output of TCO analysis is frequently used as an input to an ROI analysis.
- ROI analysis is typically focused on the difference between the costs of alternative actions, generally "what is the difference in my financial position if I make or do not make this investment?"
- Where a specific cost is the same over time between all assessed options, omission of this cost has little impact and may simplify the ROI analysis, for example, a hard-to-determine staff cost for support and maintenance of the device.
- Incomes due to the investment are a key part of ROI analysis; for example, if the purchased server is to be used to deliver charged services to customers, then differences in capacity which result in differences in the per server income are important.

We may consider an example of whether to replace an existing old UPS system with a newer device, which will both reduce the operational cost and address a constraint on data center capacity, allowing a potential increase in customer revenue, as shown in Table 6.2.

In this case, we can see that the balance is tipped by the estimate of the potential increase in customer revenue available after the upgrade. Note that both the trade-in rebate of the new UPS from the vendor and the estimate of increased customer revenue are of the opposite sign to the costs. In this case, we have shown the costs as negative and the income as positive. This is a common feature of ROI analysis; we treat all costs and income as cash-flows in or out of our analysis,

TABLE 6.1 Simple TCO example, not including time. Please visit the companion website for an editable example Excel spreadsheet for this table.

Cost	Server A	Server B
Capital purchase	$2000	$1500
3-year maintenance contract	$900	$700
installation and cabling	$300	$300
3-year data center power and cooling capacity	$1500	$2000
3-year data center energy consumption	$1700	$2200
3-year monitoring, patches, and backup	$1500	$1500
TCO	$7900	$8200

TABLE 6.2 Simple ROI example, not including time. Please visit the companion website for an editable example Excel spreadsheet for this table.

Income received or cost incurred	Existing UPS	UPS upgrade	Difference
New UPS purchase	$0	−$100,000	−$100,000
New UPS installation	$0	−$10,000	−$10,000
Competitive trade-in rebate for old UPS	$0	$10,000	$10,000
UPS battery costs (old UPS also requires replacement batteries)	−$75,000	−$75,000	$0
10 years UPS service and maintenance contract	−$10,000	−$5,000	$5,000
Cost of power lost in UPS inefficiency	−$125,000	−$50,000	$75,000
Additional customer revenue estimate	$0	$80,000	$80,000
Total	−$210,000	−$150,000	$60,000

whether costs are signed positive or negative only makes a difference to how we explain and present our output, but they should be of the opposite sign to incomes. In this case, we present the answer: "The ROI of the $100,000 new UPS upgrade is $60,000 over 10 years."

As for the simple TCO analysis, this answer is by no means complete as we have yet to consider how the values change over time and is thus unlikely to earn us much credit with the CFO.

6.1.6 Time Value of Money

While it may initially seem sensible to do what is presented earlier in the simple TCO and ROI tables and simply add up the total cost of a project and then subtract the total cost saving or additional revenue growth, this approach does not take into account what economists and business finance people call the "time value of money."

At a simple level, it is relatively easy to see that the value of a certain amount of money, say $100, depends on when you have it; if you had $100 in 1900, this would be considerably more valuable than $100 now. There are a number of factors to consider when we need to think about money over a time frame.

The first factor is inflation; in the earlier example, the $100 had greater purchasing power in 1900 than now due to inflation, the rise in costs of materials, energy, goods, and services between then and now. In the context of a data center evaluation, we are concerned with how much more expensive a physical device or energy may become over the lifetime of our investment.

The second factor is the interest rate that could be earned on the money, the $100 placed in a deposit account with 5% annual interest would become $105 at the end of year 1, $110.25 in year 2, $115.76 in year 3, and so on. If $100 was invested in a fixed interest account with 5% annual interest in 1912, when RMS Titanic departed from Southampton, the account would have increased to $13,150 by 2012 and in a further 100 years in 2112 would have become $1,729,258 (not including taxes or banking fees). This nonlinear impact of compound interest is frequently the key factor in ROI analysis.

The third factor, one that is harder to obtain a defined number, or even the method agreed for, is risk. If we invest the $100 in April on children's toys that we expect to sell from a toy shop in December, we may get lucky and be selling the must-have toy; alternatively, we may find ourselves selling most of them off at half price in January. In a data center project, the risk could be an uncertain engineering outcome affecting operational cost savings, uncertainty in the future cost of energy, or potential variations in the customer revenue received as an outcome of the investment.

6.1.7 Cost of Capital

When we calculate the Present Value (PV) of an investment option, the key number we will need for our calculation is the discount rate. In simple examples, the current interest rate is used as the discount rate, but many organizations use other methods to determine their discount rate, and these are commonly based on their cost of capital; you may see this referred to as the Weighted Average Cost of Capital (WACC).

The cost of capital is generally given in the same form as an interest rate and expresses the rate of return that the organization must achieve from any investment in order to satisfy its investors and creditors. This may be based on the interest rate the organization will pay on loans or on the expected return on other investments for the organization. It is common for the rate of ROIs in the normal line of business to be used for this expected return value. For example, an investment in a data center for a pharmaceuticals company might well be evaluated against the return on investing in new drug development.

There are various approaches to the calculation of cost of capital for an organization, all of which are outside the scope of this book. You should ask the finance department of the organization to whom you are providing the analysis what discount rate or cost of capital to use.

6.1.8 ROI Period

Given that the analysis of an investment is sensitive to the time frame over which it is evaluated, we must consider this time frame. When we are evaluating a year one capital cost against the total savings over a number of years, both the number of years' savings we can include and the discount rate have a significant impact on the outcome. The ROI period will depend on both the type of project and the accounting practices in use by the organization whose investment you are assessing.

The first aspect to consider is what realistic lifetime the investment has. In the case of a reinvestment in a data center which is due to be decommissioned in 5 years, we have a fairly clear outer limit over which it is reasonable to evaluate savings. Where the data center has a longer or undefined lifetime, we can consider the effective working life of the devices affected by our investment. For major elements of data center infrastructure such as transformers, generators, or chillers, this can be 20 years or longer, while for other elements such as computer room air conditioning/computer room air handling (CRAC/CRAH) units, the service lifetime may be shorter, perhaps 10–15 years. Where the devices have substantial periodic maintenance costs such as UPS battery-refresh, these should be included in your analysis if they occur within the time horizon.

One key consideration in the assessment of device lifetime is proximity to the IT equipment. There are a range of devices such as rear door and in-row coolers that are installed very close to the IT equipment, in comparison to traditional devices such as perimeter CRAC units or air handling units (AHUs). A major limiting factor on the service lifetime of data center infrastructure is the rate of change in the demands of the IT equipment. Many data centers today face cooling problems due to the increase in IT power density. The closer coupled an infrastructure device is to the IT equipment, the more susceptible it is likely to be to changes in IT equipment power density or other demands. You may choose to adjust estimates of device lifetimes to account for this known factor.

In the case of reinvestment, particularly those designed to reduce operational costs by improving energy efficiency, the allowed time frame for a return is likely to be substantially shorter, NPV analysis durations as short as 3 years are not uncommon, while others may calculate their Internal Rate of Return (IRR) with savings "to infinity."

Whatever your assessment of the service lifetime of an investment, you will need to determine the management accounting practices in place for the organization and whether there are defined ROI evaluation periods; and if so, which of these is applicable for the investment you are assessing. These defined ROI assessment periods are frequently shorter than the device working lifetimes and are set based on business, not technical criteria.

6.1.9 Components of TCO and ROI

When we are considering the TCO or ROI of some planned project in our data center, there are a range of both costs and incomes which we are likely to need to take into account. While TCO focuses on costs, this does not necessarily exclude certain types of income; in an ROI analysis, we are likely to include a broader range of incomes as we are looking for the overall financial outcome of the decision.

It is useful when identifying these costs to determine which costs are capital and which are operational, as these two types of cost are likely to be treated quite differently by the finance group. Capital costs not only include purchase costs but also frequently include capitalized costs occurring at the time of purchase of other actions related to the acquisition of a capital asset.

6.1.9.1 Initial Capital Investment The initial capital investment is likely to be the first value in an analysis. This cost will include not only the capital costs of equipment purchased but frequently some capitalized costs associated with the purchase. These might include the cost of preparing the site, installation of the new device(s), and the removal and disposal of any existing devices being replaced. Supporting items such as software licenses for the devices and any cost of integration to existing systems are also sometimes capitalized.

You should consult the finance department to determine the policies in place within the organization for which you are performing the analysis, but there are some general guidelines for which costs should be capitalized.

Costs are capitalized where they are incurred on an asset that has a useful life of more than one accounting period; this is usually one financial year. For assets that last more than one period, the costs are amortized or depreciated over what is considered to be the useful life of the asset. Again, it is important to note that the accounting lifetime; and therefore, depreciation period of an asset may well be shorter than the actual working life you expect to achieve based on accounting practice or tax law.

The rules on capitalization and depreciation vary with local law and accounting standards; but as a conceptual guide, the European Financial Reporting Standard guidance indicates that the costs of fixed assets should initially be "directly attributable to bringing the asset into working condition for its intended use."

Initial capitalized investment costs for a UPS replacement project might include the following:

- Preparation of the room
- Purchase and delivery
- Physical installation
- Wiring and safety testing
- Commissioning and load testing
- Installation and configuration of monitoring software
- Training of staff to operate the new UPS and software
- Decommissioning of the existing UPS devices
- Removal and disposal of the existing UPS devices

Note that disposal does not always cost money; there may be a scrap value or rebate payment; this is addressed in the additional incomes section that follows.

6.1.9.2 Reinvestment and Upgrade Costs There are two circumstances in which you would need to consider this second category of capital cost.

The first is where your project does not purchase completely new equipment but instead carries out remedial work or an upgrade to existing equipment to reduce the operating cost, increase the working capacity, or extend the lifetime of the device, the goal being "enhances the economic benefits of the asset in excess of its previously assessed standard of performance." An example of this might be reconditioning a cooling tower by replacing corroded components and replacing the old fixed speed fan assembly with a new variable frequency drive (VFD)-controlled motor and fan. This both extends the service life and reduces the operating cost and, therefore, is likely to qualify as a capitalized cost.

The second is where your project will require additional capital purchases within the lifetime of the device such as a UPS system that is expected to require one or more complete replacements of the batteries within the working life in order to maintain design performance. These would be represented in your assessment at the time the cost occurs. In financial terminology, these costs "relate to a major inspection or overhaul that restores the economic benefits of the asset which have been consumed by the entity."

6.1.9.3 Operating Costs The next major group of costs relate to the operation of the equipment. When considering the operational cost of the equipment, you may include any cost attributable to the ownership and operation of that equipment including staffing, service and maintenance contracts, consumables such as fuel or chemical supplies, operating licenses, and water and energy consumption.

Operating costs for a cooling tower might include the following:

- Annual maintenance contract including inspection and cleaning
- Cost of metered potable water
- Cost of electrical energy for fan operation
- Cost of electrical energy for basin heaters in cold weather
- Cost of the doping chemicals for tower water

All operating costs should be represented in the accounting period in which they occur.

6.1.9.4 Additional Income It is possible that your project may yield additional income, which could be recognized in the TCO or ROI analysis. These incomes may be in the form of rebates, trade-in programs, salvage values for old equipment, or additional revenue enabled by the project. If you are performing a TCO analysis to determine the cost at which a product or service may be delivered, then the revenue would generally be excluded from this analysis. Note that these additional incomes should be recognized in your assessment in the accounting period in which they occur.

Additional income from a UPS replacement project might include the following:

- Salvage value of the existing UPS and cabling
- Trade-in value of the existing UPS from the vendor of the new UPS devices
- Utility, state, or government energy efficiency rebate programs where project produces an energy saving that can realistically be shown to meet the rebate program criteria

6.1.9.5 Taxes and Other Costs One element that varies greatly with both location and the precise nature of the project is taxation. The tax impact of a project should be at least scoped to determine if there may be a significant risk or saving. Additional taxes may apply when increasing capacity in the form of emissions permits for diesel generators or carbon allowances if your site is in an area where a cap-and-trade scheme is in force, particularly if the upgrade takes the site through a threshold. There may also be substantial tax savings available for a project due to tax rebates, for example, rebates on corporate tax for investing or creating employment in a specific area. In many cases, corporate tax may be reduced through the accounting depreciation of any capital assets purchased. This is discussed further in the section "Accounting for Taxes."

6.1.9.6 End-of-Life Costs In the case of some equipment, there may be end-of-life decommissioning and disposal costs that are expected and predictable. These costs should be included in the TCO or ROI analysis at the point at which they occur. In a replacement project, there may be disposal costs for the existing equipment that you would include in the first capital cost as it occurs in the same period as the initial investment. Disposal costs for the new or modified equipment at the end of service life should be included and valued as at the expected end of life.

6.1.9.7 Environmental, Brand Value, and Reputational Costs Costs in this category for a data center project will vary substantially depending on the organization and legislation in the operating region but may also include the following:

- Taxation or allowances for water use
- Taxation or allowances for electricity use
- Taxation or allowances for other fuels such as gas or oil
- Additional energy costs from "green tariffs"
- Renewable energy certificates or offset credits
- Internal cost of carbon (or equivalent)

6.2 FINANCIAL MEASURES OF COST AND RETURN

When the changing value over time is included in our assessment of project costs and returns, it can substantially affect the outcome and viability of projects. This section provides an introduction and examples for the basic measures of PV and IRR, followed by a short discussion of the relative strengths and weaknesses.

6.2.1 Common Business Metrics and Project Approval Tests

There are a variety of relatively standard financial methods used and specified by management accountants to analyze investments and determine their suitability. It is likely that the finance department in your organization has a preferred metric that you will be expected to use—in many larger enterprises, a template spreadsheet or document is provided which must be completed as part of the submission. It is not unusual for there to be a standard "hurdle" for any investment expressed in terms of this standard calculation or metric such as "all projects must exceed a 30% Internal Rate of Return."

The measures you are most likely to encounter are as follows:

- TCO: Total Cost of Ownership
- NPV: the Net Present Value of an option
- IRR: the Internal Rate of Return of an investment

Both the NPV and IRR are forms of ROI analysis and are described later.

While the essence of these economic hurdles may easily be misread as "We should do any project which exceeds the hurdle" or "We should find the project with the highest ROI metric and do that," there is, unsurprisingly, more to consider than which project scores best on one specific metric. Each has its own strengths and weaknesses, and making good decisions is as much about understanding the relative strengths of the metric as how to calculate them.

6.2.1.1 Formulae and Spreadsheet Functions In this section, there are several formulae presented; in most cases where you are calculating PV or IRR, there are spreadsheet functions for these calculations that you can use directly without needing to know the formula. In each case, the relevant Microsoft Office Excel function will be described in addition to the formula for the calculation.

6.2.2 Present Value

The first step in calculating the PV of all the costs and savings of an investment is to determine the PV of a single cost or payment. As discussed under time value of money, we need to discount any savings or costs that occur in the future to obtain an equivalent value in the present. The basic formula for the PV of a single payment a at time n accounting periods into the future at discount rate i per period is given by the following relation:

$$\mathrm{PV}_n = \frac{a}{(1+i)^n}$$

In Microsoft Office Excel, you would use the PV function:

$$= \mathrm{PV}(\text{rate,nper,pmt,fv}) = \mathrm{PV}(i,n,0,-a)$$

We can use this formula or spreadsheet function to calculate the PV (i.e., the value today) of a single income we receive or cost we incur in the future. Taking an income of \$1000 and an interest rate of 10% per annum, we obtain the following:

$$\text{End of year 1 PV} = 1000.\frac{1}{(1+0.1)^1} = 1000.\frac{1}{(1.1)^1}$$
$$= 1000.\frac{1}{1.1} = 909.\dot{0}\dot{9}$$

$$\text{End of year 2}\,PV = 1000.\frac{1}{(1+0.1)^2} = 1000.\frac{1}{(1.1)^2}$$
$$= 1000.\frac{1}{1.21} = 826.45$$

$$\text{End of year 3}\,PV = 1000.\frac{1}{(1+0.1)^3} = 1000.\frac{1}{(1.1)^3}$$
$$= 1000.\frac{1}{1.331} = 751.31$$

If we consider an annual income of \$1000 over 10 years, with the first payment at the end of this year then, we obtain the series of PVs shown in Table 6.3 for our \$1000 per year income stream.

The values of this series of individual \$1000 incomes over a 20-year period are shown in Figure 6.1.

Figure 6.1 shows that the PVs of the incomes reduce rapidly at our 10% discount rate toward a negligible value. If we plot the total of the annual income PVs over a 50-year period, we see that the total tends toward \$10,000 as shown in Figure 6.2.

This characteristic of the PV is important when assessing the total value of savings against an initial capital investment; at higher discount rates, increasing the number of years considered for return on the investment has little impact. How varying the interest rate impacts the PVs of the income stream is shown in Figure 6.3.

As the PV of a series of payments of the same value is a geometric series, it is easy to use the standard formulae for the sum to n terms and to infinity, to determine the total value of the number of payments PV_A or the sum of a perpetual series of payments PV_P which never stops:

TABLE 6.3 PV of \$1000 over 10 years at 10% discount rate. Please visit the companion website for an editable example Excel spreadsheet for this table.

Year	1	2	3	4	5	6	7	8	9	10
Income	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000
Scalar	0.91	0.83	0.75	0.68	0.62	0.56	0.51	0.47	0.42	0.39
PV at 10%	\$909.09	\$826.45	\$751.31	\$683.01	\$620.92	\$564.47	\$513.16	\$466.51	\$424.10	\$385.54

$$PV_A = \frac{a}{i} \cdot \left(1 - \frac{1}{(1+i)^n}\right) \qquad = PV(\text{rate}, \text{nper}, \text{pmt}) = PV(i, n, -a)$$

$$PV_P = \frac{a}{i}$$

Note: In Excel, the PV function uses payments, not incomes; to obtain a positive value from the PV function, we must enter incomes as negative payments.

Using these formulae, we can determine the value of a perpetual series of $1000 incomes over any period for any interest rate as shown in Table 6.4.

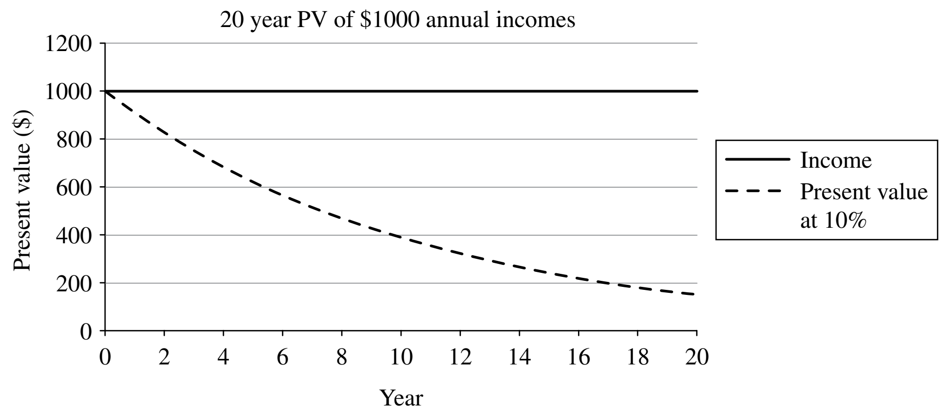


FIGURE 6.1 PV of $1000 annual incomes at 10% interest rate. Please visit the companion website for an editable example Excel spreadsheet for this figure.

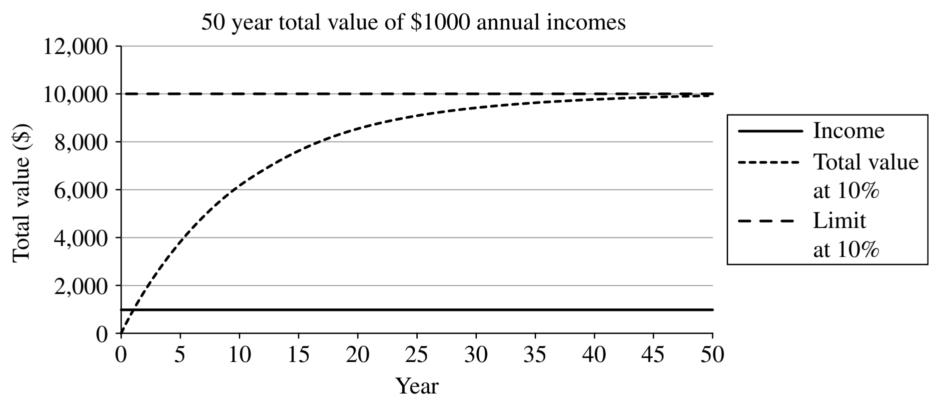


FIGURE 6.2 Total value of $1000 incomes at 10% interest rate. Please visit the companion website for an editable example Excel spreadsheet for this figure.

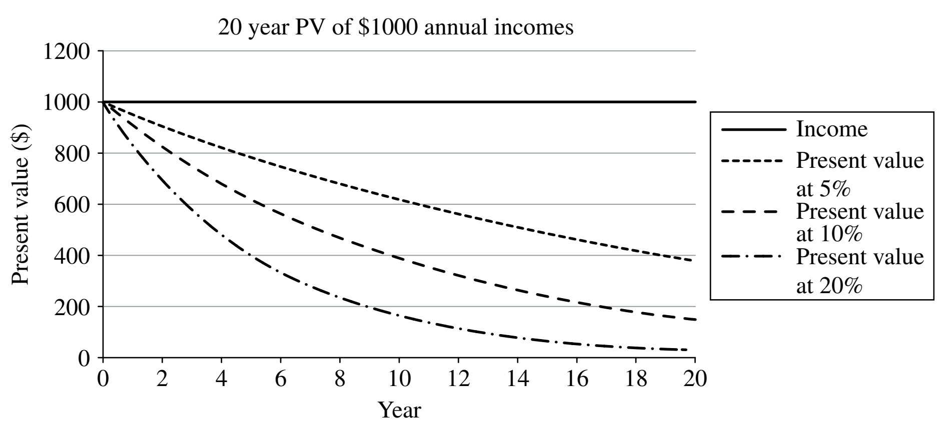


FIGURE 6.3 PV of $1000 annual incomes at varied interest rates. Please visit the companion website for an editable example Excel spreadsheet for this figure.

TABLE 6.4 Value of $1000 incomes over varying periods and discount rates. Please visit the companion website for an editable example Excel spreadsheet for this table.

Discount rate (%)	5 years	10 years	20 years	Perpetual
1	$4853	$9471	$18,046	$100,000
5	$4329	$7722	$12,462	$20,000
10	$3791	$6145	$8,514	$10,000
15	$3352	$5019	$6,259	$6,667
20	$2991	$4192	$4,870	$5,000
30	$2436	$3092	$3,316	$3,333

These values may be easily calculated using the financial functions in most spreadsheets; in Microsoft Office Excel, the PV function takes the arguments PV (Interest Rate, Number of Periods, Payment Amount).

To calculate the value to 10 years of the $1000 annual payments at 5% discount rate in a spreadsheet, we can use

$$=\mathbf{PV}\left(\mathbf{0.05},\mathbf{10},\mathbf{-1000}\right) \qquad \text{which returns \$7721.73.}$$

6.2.3 Net Present Value

To calculate the NPV of an investment, we need to consider more than just a single, fixed value, saving over the period, we must include the costs and savings, in whichever accounting period they occur, to obtain the overall value of the investment.

6.2.3.1 Simple Investment NPV Example As an example, if an energy savings project has a $7000 implementation cost, yields $1000 savings per year, and is to be assessed over 10 years, we can calculate the income and resulting PV in each year as shown in Table 6.5.

The table shows one way to assess this investment. Our initial investment of $7000 is shown in year zero as this money is spent up front; and therefore, the PV is – $7000. We then have a $1000 accrued saving at the end of each year for which we calculate the PV based on the 5% annual discount rate. Totaling these PVs gives the overall NPV of the investment, $722.

Alternatively, we can calculate the PV of each element and then combine the individual PVs to obtain our NPV as shown in Table 6.6; this is an equivalent method, and choice depends on which is easier in your particular case.

The general formula for NPV is as follows:

$$NPV(i,N)=\sum_{n=0}^{N}\frac{R_n}{(1+i)^n} \qquad \begin{aligned}&=\text{NPV(rate, value 1, value 2, …)}\\&=\text{NPV}(i,R_1,R_2,\ldots)\end{aligned}$$

where R_t is the cost incurred or income received in period t, i is the discount rate (interest rate), N is the number of costs or income periods, and n is the time period over which to evaluate NPV. In the Excel formula, R_1, R_2, etc. are the individual costs or incomes. Note that in the Excel, the first cost or income is R_1 and not R_0 and therefore one period's discount rate is applied to the first value; we must handle the year zero capital cost separately.

6.2.3.2 Calculating Break-Even Time Another common request when forecasting ROI is to find the time (if any) at which the project investment is equaled by the incomes or savings of the project to determine the break-even time of the project. If we simply use the cash-flows, then the break-even point is at 7 years where the total incomes of $7000 match the initial cost. The calculation becomes more complex when we include the PV of the project incomes as shown in Figure 6.4.

Including the impact of discount rate, our break-even points are shown in Table 6.7.

As shown in the graph and table, the break-even point for a project depends heavily on the discount rate applied to the analysis. Due to the impact of discount rate on the total PV of the savings, it is not uncommon to find that a project fails to achieve break even over any time frame despite providing ongoing returns that appear to substantially exceed the implementation cost.

As for the NPV, spreadsheets have functions to help us calculate the break-even point; in Microsoft Office Excel, we can use the NPER (discount rate, payment, present value) function but only for constant incomes. Once you consider any aspect of a project that changes over time, such as the energy tariff or planned changes in IT load, you are more likely to have to calculate the annual values and look for the break-even point manually.

6.2.4 Profitability Index

One of the weaknesses of NPV as an evaluation tool is that it gives no direct indication of the scale of return compared to the initial investment. To address this, some organizations use a simple variation of the NPV called profitability index, which simply divides the PV of the incomes by the initial investment.

The general formula for profitability index is

$$\text{Profitability Index}=\frac{\text{PV}\left(\text{Future Incomes}\right)}{\text{Initial Investment}} \qquad \begin{aligned}&=\text{NPV(rate, value 1, value 2, …)/Investment}\\&=\text{NPV}(i,N_1,N_2,\ldots)\text{/Investment}\end{aligned}$$

where i is the discount rate (interest rate) and N_1 and N_2 are the individual costs or incomes.

For our simple investment example presented earlier, the Profitability Indexes would be as shown in Table 6.8.

6.2.5 NPV of the Simple ROI Case

Returning to the simple ROI case used previously of a UPS replacement, we can now re-calculate the ROI including the discount rate and assess whether our project actually provides an overall return and, if so, how much. In our

TABLE 6.5 Simple investment example as NPV

Year	0	1	2	3	4	5	6	7	8	9	10
Cost	$7000										
Saving		$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000
Annual cost	−$7000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000
PV	−$7000	$952	$907	$864	$823	$784	$746	$711	$677	$645	$614
NPV											$722

TABLE 6.6 Calculate the PV of each element and combine.

	Amount	Periods	Discount rate	PV
Cost	$7000			−$7000
Saving	−$1000	10	5%	$7722
NPV				$722

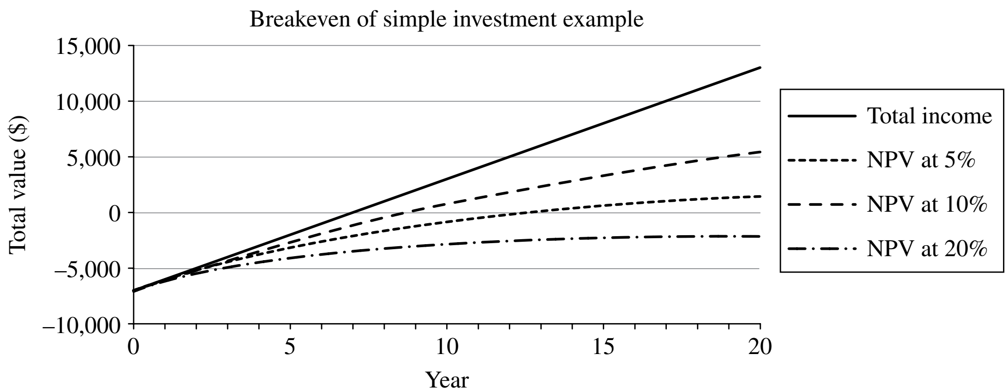


FIGURE 6.4 Breakeven of simple investment example.

TABLE 6.7 Break-even point of simple investment example under varying discount rates

Case	Break-even years	Formula
Simple payback	7.0	=NPER(0, −1000, 7000)
NPV = 0 at 5%	8.8	=NPER(0.05, −1000, 7000)
NPV = 0 at 10%	12.6	=NPER(0.1, −1000, 7000)
NPV = 0 at 20%	#NUM!	=NPER(0.2, −1000, 7000)

TABLE 6.8 Profitability index of simple investment example

Discount rate	Profitability index	Formula
0% simple payback	2.86	=$20,000/$7000
5%	1.78	=$12,462/$7000
10%	1.22	=$8,514/$7000
20%	0.70	=$4,869/$7000

simple addition previously, the project outcome was a saving of $60,000, for this analysis we will assume that the finance department has requested the NPV over 10 years with a 10% discount rate as shown in Table 6.9.

With the impact of our discount rate reducing the PV of our future savings at 10% per annum, our UPS upgrade project now evaluates as showing a small loss over the 10-year period.

The total NPV may be calculated either by summing the individual PVs for each year or by using the annual total costs or incomes to calculate the NPV. In Microsoft Office Excel, we can use the NPV worksheet function that takes the arguments; NPV (Discount Rate, Future Income 1, Future Income 2, …). It is important to treat each cost or income in the correct period. Our first cost occurs at the beginning of the first year, but our payments occur at the end of the year; this must be separately added to the output of the NPV function. The other note is that the NPV function takes incomes rather than payments, so the signs are reversed as compared to the PV function.

To calculate our total NPV in the cells already mentioned, we would use the formula **=B9+NPV(0.1, C9:L9)** which takes the initial cost and adds the PV of the savings over the 10-year period.

6.2.6 Internal Rate of Return

The Internal Rate of Return (IRR) is closely linked to the NPV calculation. In the NPV calculation, we use a discount rate to reduce the PV of costs or incomes in the future to determine the overall, net, value of an investment. To obtain the IRR of an investment, we simply reverse this process to find the discount rate at which the NPV of the investment is zero.

TABLE 6.9 Calculating the NPV of the simple ROI example. Please visit the companion website for an editable example Excel spreadsheet for this table.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	0	1	2	3	4	5	6	7	8	9	10
2	New UPS purchase	−$100,000										
3	New UPS installation	−$10,000										
4	Competitive trade in rebate	$10,000										
5	UPS battery costs	$0										
6	UPS maintenance contract		$500	$500	$500	$500	$500	$500	$500	$500	$500	$500
7	UPS power costs		$7,500	$7,500	$7,500	$7,500	$7,500	$7,500	$7,500	$7,500	$7,500	$7,500
8	Additional revenue		$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000
9	Annual total	−$100,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000
10	PV	−$100,000	$14,545	$13,223	$12,021	$10,928	$9,935	$9,032	$8,211	$7,464	$6,786	$6,169
11	NPV											−$1,687

TABLE 6.10 Calculating IRR for the simple investment example. Please visit the companion website for an editable example Excel spreadsheet for this table.

Year	0	1	2	3	4	5	6	7	8	9	10
Cost	$7000										
Saving		$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000
Annual cost	−$7000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000
IRR											7.07%

TABLE 6.11 NPV of the simple investment example with a discount rate equal to the IRR. Please visit the companion website for an editable example Excel spreadsheet for this table.

Year	0	1	2	3	4	5	6	7	8	9	10
Cost	$7000										
Saving		$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000
Annual cost	−$7000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000	$1000
PV	−$7000	$934	$872	$815	$761	$711	$664	$620	$579	$541	$505
NPV											$0

To find the IRR in Microsoft Office Excel, you can use the IRR function:

$$= \text{IRR}\left(\text{values,}\left[\text{guess}\right]\right)$$

6.2.6.1 Simple Investment IRR Example We will find the IRR of the simple investment example from NPV given earlier of a $7000 investment that produced $1000 per annum operating cost savings. We tested this project to yield an NPV of $722 at a 5% discount rate over 10 years. The IRR calculation is shown in Table 6.10.

The IRR was calculated using the formula = IRR(B4:L4), which uses the values in the “annual cost” row from the initial − $7000 to the final $1000.

In this case, we see that the IRR is just over 7%; if we use this as the discount rate in the NPV calculation, then our NPV evaluates to zero as shown in Table 6.11.

6.2.6.2 IRR Over Time As observed with the PV, incomes later in the project lifetime have progressively less impact on the IRR of a project; in this case, Figure 6.5 shows the IRR of the simple example given earlier up to 30 years project lifetime. The IRR value initially increases rapidly with project lifetime but can be seen to be tending toward approximately 14.3%.

6.2.7 Choosing NPV or IRR

In many cases, you will be required to present either an NPV or an IRR case, based on corporate policy and sometimes within a standard form, without which finance will not consider your proposal. In other cases, you may need to choose whether to use an IRR or NPV analysis to best present the investment case. In either case, it is worth understanding what the relative strengths and weaknesses of NPV and IRR analysis are, to select the appropriate tool, and to properly manage the weaknesses of the selected analysis method.

At a high level, the difference is that NPV provides a total money value without indication of how large the return is in comparison to the first investment, while IRR provides a rate of return with no indication of the scale. There are, of course, methods of dealing with both of these issues, but perhaps the simplest is to lay out the key numbers for investment, NPV, and IRR to allow the reader to compare the projects in their own context.

To illustrate some of the potential issues with NPV and IRR, we have four simple example projects in Table 6.12, each of which has a constant annual return over 5 years, evaluated at a discount rate of 15%.

6.2.7.1 Ranking Projects The first issue is how to rank these projects. If we use NPV to rank the projects, then we would select project D with the highest NPV when, despite requiring twice the initial investment of project C, the return is less than 1% larger. If we rank the projects using only the Profitability Index or IRR, then projects A and C would appear to be the same despite C being five times larger in both investment and return than A. If we are seeking maximum total return, then C would be preferable; conversely, if there is substantial risk in the projects, we may choose to take five projects like A rather than C alone.

A further complication in data center projects is that in many cases the project options are mutually exclusive, either because there is limited total budget available, or because the projects cannot both be implemented such as options to upgrade or replace the same piece of equipment. If we had $1 million to invest and these four projects to choose from, we might well choose B and C; however, if these two projects are an either or option, then A and C would be our selection, and we would not invest $400k of our available budget.

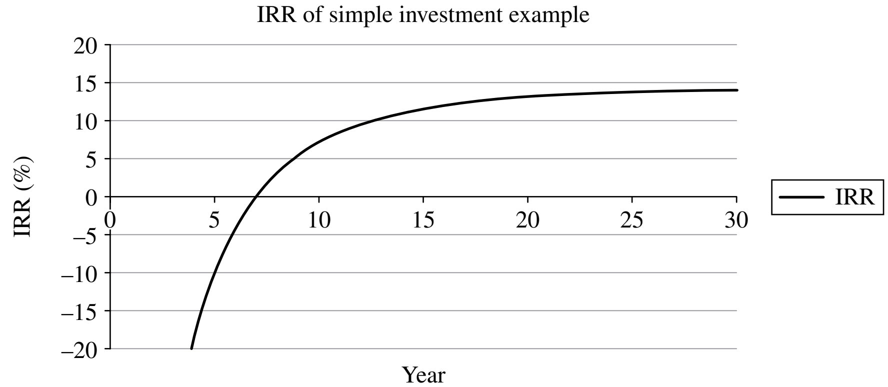


FIGURE 6.5 IRR of simple investment example. Please visit the companion website for an editable example Excel spreadsheet for this figure.

TABLE 6.12 NPV and IRR of four simple projects. Please visit the companion website for an editable example Excel spreadsheet for this table.

Project	Capital cost	Annual return	NPV	Profitability index	IRR (%)
A	–$100,000	$50,000	$67,608	1.68	41
B	–$500,000	$200,000	$170,431	1.34	29
C	–$500,000	$250,000	$338,039	1.68	41
D	–$1,000,000	$400,000	$340,862	1.34	29

Clearly, neither NPV nor IRR alone is suitable for ranking projects; this can be a particular issue in organizations where the finance group set a minimum IRR for any project, and it may be appropriate to present options that are near to the minimum IRR but have larger available returns than those which exceed the target IRR.

6.2.7.2 Other Issues IRR should not be used to compare projects of different durations; your finance department will typically have a standard number of years over which an IRR calculation is to be performed.

IRR requires both costs and savings; you can't use IRR to compare purchasing or leasing a piece of equipment.

In a project with costs at more than one time, such as modular build of capacity, there may be more than one IRR at different times in the project.

6.3 COMPLICATIONS AND COMMON PROBLEMS

All of the examples so far have been relatively simple with clear predictions of the impact of the changes to allow us to clearly assess the NPV or IRR of the project. In the real world, things are rarely this easy, and there will be many factors that are unknown, variable, or simply complicated, which will make the ROI analysis less easy. This section will discuss some of the complications as well as some common misunderstandings in data center financial analysis.

6.3.1 ROI Analysis is about Optimization, Not Just Meeting a Target Value

When assessing the financial viability of data center projects, there will generally be a range of options in how the projects are delivered, which will affect the overall cost and overall return. The art of an effective financial analysis is to break down the components of each project and understand how each of these contributes to the overall ROI outcome. Once you have this breakdown of benefit elements, these may be weighed against the other constraints that you must work within. In any organization with more than one data center, it will also be necessary to balance the available resources across the different sites.

A good ROI analysis will find an effective overall balance considering the following:

- Available internal resource to evaluate, plan, and implement or manage projects
- Projects that are mutually exclusive for engineering or practical reasons
- The total available budget and how it is distributed between projects

6.3.2 Sensitivity Analysis

As already stated, analysis of a project requires that we make a number of assumptions and estimations of future events. These assumptions may be the performance of devices once installed or upgraded, the changing cost of electricity over the next 5 years, or the increase in customer revenue due to a capacity expansion. While the estimated ROI of a project is important, it is just as vital to understand and communicate the sensitivity of this outcome to the various assumptions and estimations.

At a simple level, this may be achieved by providing the base analysis, accompanied by an identification of the impact on ROI of each variable. To do this, you can state the estimate and the minimum and maximum you would reasonably expect for each of variable and then show the resulting ROI under each change.

As a simple example, a project may have an estimated ROI of $100,000 at a power cost of $0.10 per kWh, but your estimate of power cost ranges from $0.08 to $0.12 per kWh, which result in ROIs of $50,000 and $150,000, respectively. It is clearly important for the decision maker to understand the impact of this variability, particularly if the company has other investments that are subject to variation in energy cost.

There are, of course, more complex methods of assessing the impact of variability on a project; one of the more popular, Monte Carlo analysis, is introduced later in this chapter.

6.3.2.1 Project Benefits Are Generally Not Cumulative One very common mistake is to independently assess more than one data center project and then to assume that the results may be added together to give a total capacity release or energy savings for the combined projects if implemented together.

The issue with combining multiple projects is that the data center infrastructure is a system and not a set of individual components. In some cases, the combined savings of two projects can exceed the sum of the individual savings; for example, the implementation of air flow containment with VFD fan upgrades to the CRAC units coupled with the addition of a water side economizer. Either project would save energy, but the air flow containment allows the chilled water system temperature to be raised, which will allow the economizer to further decrease the compressor cooling requirement.

More frequently, some or all of the savings of two projects rely on reducing the same overheads in the data center. The same overhead can't be eliminated twice; and therefore, the total savings will not be the sum of the individual projects. A simple example might be the implementation of raised supply temperature set-points and adiabatic intake air cooling in a data center with direct outside air economizing AHUs. These two projects would probably be complementary, but the increase in set points seeks to reduce the same compressor cooling energy as the adiabatic cooling; and therefore, the total will almost certainly not be the sum of the parts.

6.3.3 Accounting for Taxes

In many organizations, there may be an additional potential income stream to take account of in your ROI analysis in the form of reduced tax liabilities. In most cases, when a capital asset is purchased by a company, the cost of the asset is not dealt for tax purposes as one lump at the time of purchase. Normal practice is to depreciate the asset over some time frame at a given rate; this is normally set by local tax laws. This means that, for tax purposes, some or all of the capitalized cost of the project will be spread out over a number of years, this depreciation cost may then be used to reduce tax liability in each year. This reduced tax liability may then be included in each year of the project ROI analysis and counted toward the overall NPV or IRR. Note that for the ROI analysis, you should still show the actual capital costs occurring in the accounting periods in which they occur, it is only the tax calculation that uses the depreciation logic.

The discussion of regional tax laws and accounting practices related to asset depreciation and taxation is clearly outside of the scope of this book, but you should consult the finance department in the organization for whom you are producing the analysis to determine whether and how they wish you to include tax impacts.

6.3.4 Costs Change over Time—Real and Nominal Discount Rates

As already discussed, the value of money changes over time; however, the cost of goods, energy, and services also changes over time, and this is generally indicated for an economy by an annual percentage inflation or deflation. When performing financial analysis of data center investments, it may be necessary to consider how costs or incomes may change independently of a common inflation rate.

The simpler method of NPV analysis uses the real cash flows. These are cash flows that have been adjusted to the current value, or more frequently, simply estimated at their current value. This method then applies what is called the real discount rate that includes both the nominal interest rate and a reduction to account for the inflation rate. The relationship between the real and nominal rates is shown as follows:

$$\text{Real} = \left(\frac{1+\text{nominal}}{1+\text{inflation}}\right) - 1$$

The second method of NPV analysis allows you to make appropriate estimates for the changes in both costs and revenues over time. This is important where you expect changes in goods or energy costs which are not well aligned with inflation or each other. In this case, the actual (nominal) cash flows are used and the full nominal discount rate is applied.

As an example, consider a project with a $100,000 initial capital investment, which we expect to produce a $50,000 income in today's money across each of 3 years. For this project, the nominal discount rate is 10%, but we expect inflation over the period to be 2.5% which gives a real discount rate of 7.3%.

TABLE 6.13 NPV of real cash flows at the real discount rate. Please visit the companion website for an editable example Excel spreadsheet for this table.

Capital	1	2	3	NPV	Notes
£100,000	£50,000	£50,000	£50,000		Real cash flows
	£46,591	£43,414	£40,454	£30,459	Real discount rate

TABLE 6.14 NPV of nominal cash flows at the nominal discount rate. Please visit the companion website for an editable example Excel spreadsheet for this table.

Capital	1	2	3	NPV	Notes
£100,000	£51,250	£52,531	£53,845		Nominal cash flows
	£46,591	£43,414	£40,454	£30,459	Nominal discount rate

We can perform an NPV analysis using real cash flows and the real discount rate as in Table 6.13.

Alternatively, we can include the effect of our expected inflation in the cash flows and then discount them at the nominal discount rate as in Table 6.14.

The important thing to note here is that both NPV calculations return the same result. Where the future costs and revenues all increase at the same rate as our inflation factor, the two calculations are equivalent. Where we expect any of the future cash flows to increase or decrease at any rate other than in line with inflation, it is better to use the nominal cash flows and nominal discount rate to allow us to account for these changes. Expected changes in the future cost of energy are the most likely example in a data center NPV analysis. This latter approach is used in both the Monte Carlo and main realistic example analysis later in this chapter.

6.3.5 Multiple Solutions for IRR

One of the issues in using IRR is that there is no simple formula to give an IRR; instead, you or the spreadsheet you are using must seek a value of discount rate for which the NPV evaluates to zero. When you use the IRR function in a spreadsheet such as Microsoft Office Excel, there is an option in the formula to allow you to provide a guess to assist the spreadsheet in determining the IRR you seek.

$$= \mathrm{IRR}\left(\mathrm{values},\left[\mathrm{guess}\right]\right)$$

This is not because the spreadsheet has trouble iterating through different values of discount rate; but because there is not always a single, unique solution to the IRR for a series of cash flows. If we consider the series of cash flows in Table 6.15, we can see that our cash flows change sign more than once; that is, they start with a capital investment, negative, then change between incomes, positive, and further costs, negative.

TABLE 6.15 Example cash-flow with multiple IRR solutions

Year	0	1	2	3	4
Income	−$10,000	$27,000	−$15,000	−$7000	$4500

The chart in Figure 6.6 plots the NPV over the 4 years against the applied discount rate. It is evident that the NPV is zero twice due to the shape of the curve, in fact, the IRR solves to both 11 and 60% for this series of cash-flows.

There are a number of methods for dealing with this issue, from supplying an appropriate guess to the spreadsheet IRR function to assisting it in converging on the value you are looking for, through to using alternative methods such as the Modified Internal Rate of Return (MIRR), which is provided in most spreadsheet packages but is outside the scope of this chapter.

6.3.6 Broken and Misused Rules of Thumb

In the data center industry, there are many standard practices and rules of thumb; some of these have been developed over many years of operational experience, while others have taken root on thin evidence due to a lack of available information to disprove them. It is generally best to make an individual assessment; where only a rule of thumb is available this is unlikely to be an effective assumption in the ROI case.

Some of the most persistent of these are related to the cooling system and environmental controls in the data center. Some common examples are as follows:

- It is best to operate required capacity +1 of the installed CRAC/AHU; this stems from systems operating constant speed fans with flow dampers where energy was relatively linear with air flow and operating hours meant wear-out maintenance costs. In modern VFD controlled systems, the large savings of fan speed reduction dictate that, subject to minimum speed requirements, more units should operate in parallel and at the same speed.
- We achieve X% saving in cooling energy for every degree increase in supply air or water temperature. This may have been a good rule of thumb for entirely compressor cooled systems; but in any system with free cooling, the response is very nonlinear.
- The “optimum” IT equipment supply temperature is 25°C, above this IT equipment fan energy increases faster than cooling system energy. The minimum overall power point does, of course, depend upon not just the changing fan power profile of the IT equipment but also the response of the cooling system and, therefore, varies for each data center as well as between data centers.
- Applying a VFD to a fan or pump will allow the energy to reduce as the cube of flow; this is close to the truth for a system with no fixed head and the ability to turn down to any speed, but in the case of pumps which are controlled to a constant pressure such as secondary distribution water pumps, the behavior is very different.

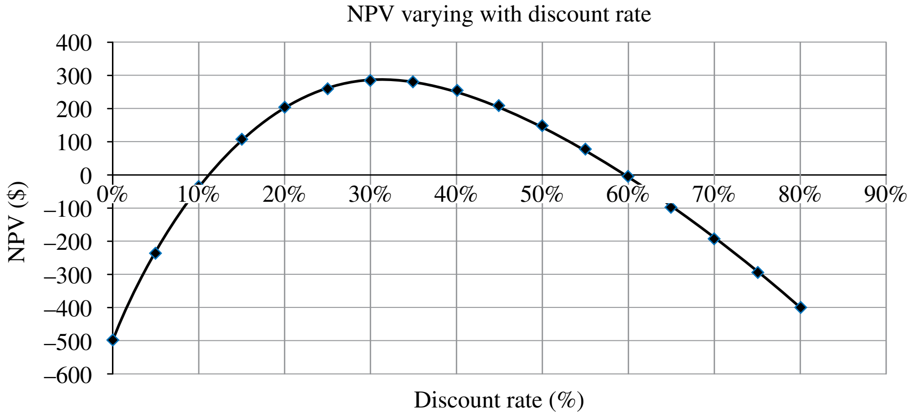


FIGURE 6.6 Varying NPV with discount rate. Please visit the companion website for an editable example Excel spreadsheet for this figure.

6.3.7 Standardized Upgrade Programs

In many end user and consulting organizations, there is a strong tendency to implement data center projects based on a single strategy that is believed to be tested and proven. This approach is generally flawed for two major reasons.

First, each data center has a set of opportunities and constraints defined by its physical building, design, and history. You should not expect a data center with split DX CRAC units to respond in the same way to an air flow management upgrade as a data center with central AHUs and overhead distribution ducts.

Second, where the data centers are distributed across different climates or power tariffs, the same investment that delivered excellent ROI in Manhattan may well be a waste of money in St Louis even when applied to a building identical in cooling design and operation.

There may well be standard elements, commonly those recognized as best practice by programs such as the EU Code of Conduct, which should be on a list of standard options to be applied to your estate of data centers. These standard elements should then be evaluated on a per-opportunity basis in the context of each site to determine the selection of which projects to apply based on a tailored ROI analysis rather than habit.

6.3.7.1 Climate Data Climate data is available in a range of formats, each of which is more or less useful for specific types of analysis. There are a range of sources for climate data, many of which are regional and have more detailed data for their region of operation.

While the majority of the climate data available to you will be taken from quite detailed observations of the actual climate over a substantial time period, this is generally processed before publication and the data you receive will be some sort of summary. The common formats you are likely to come across are as follows.

6.3.7.2 Design Conditions The design conditions for a site are generally given as the minimum and maximum temperature expected over a specified number of years. These values are useful only for ensuring the design is able to operate at the climate extremes it will encounter.

6.3.7.3 Heating/Cooling Hours It is common to find heating and cooling hours in the same data sets as design conditions; these are of no realistic use for data center analysis.

6.3.7.4 Temperature Binned Hours It is common to see analysis of traditional cooling components such as chillers carried out using data that sorts the hours of the year into temperature "bins," for example, "2316 annual hours between 10 and 15°C Dry Bulb." The size of the temperature bin varies with the data source. A major issue with this type of data is that the correlation between temperature and humidity is destroyed in the binning process. This data may be useful if no less processed data is available, but only where the data center cooling load does not vary with the time of day, humidity control is not considered (i.e., no direct air economizer systems), and the utility energy tariff does not have off-peak/peak periods or peak demand charges.

6.3.7.5 Hourly Average Conditions Another common processed form of data is the hourly average; in this format, there are 24 hourly records for each month of the year, each of which contains an average value for dry bulb temperature, humidity, and frequently other aspects such as solar radiation or wind speed and direction. This format can be more useful than binned hours where the energy tariff has peak/off-peak hours but is of limited use for humidity sensitive designs and may give false indications of performance for economized cooling systems with sharp transitions.

6.3.7.6 Typical Meteorological Year The preferred data type for cooling system analysis is Typical Meteorological Year (TMY). This data contains a set of values for each hour of the year, generally including dry bulb temperature, dew point, humidity, atmospheric pressure, solar radiation, precipitation, wind speed, and direction. This data is generally drawn from recorded observations but is carefully processed to represent a "typical" year.

6.3.7.7 Recorded Data You may have actual recorded data from a Building Management System for the site you are analyzing or another nearby site in the same climate region. This data can be useful for historical analysis; but in most cases, correctly processed TMY data is preferred for predictive analysis.

6.3.7.8 Sources of Climate Data Some good sources of climate data are the following:

- ASHRAE[2] and equivalent organizations outside the United States such as ISHRAE[3]
- The US National Renewable Energy Laboratories and Department Of Energy publish an excellent set of TMY climate data for use in energy simulations and converter tools between common file formats on the DOE website
- Weather Underground[4] where many contributors upload data recorded from weather stations which is then made freely available

6.3.8 Location Sensitivity

It is easy to see how even the same data center design may have a different cooling overhead in Finland than in Arizona and also how utility electricity may be cheaper in North Carolina than in Manhattan or Singapore. As an example, we may consider a relatively common 1 MW water-cooled data center design. The data center uses water-cooled chillers and cooling towers to supply chilled water to the CRAC units in the IT and plant areas. The data center has plate heat exchangers between the condenser water and chilled water circuits to provide free cooling when the external climate allows.

For the first part of the analysis, the data center was modeled[5] in four configurations, representing four different chilled water supply temperatures, all of the major variables in the cooling system are captured. The purpose of the evaluation is to determine the available savings from the cooling plant if the chilled water temperature is increased. Once these savings are known, it can be determined whether the associated work in air flow management or increase in IT equipment air supply temperature are worthwhile.

The analysis will be broken into two parts, first the PUE response to the local climate and then the impact of the local power tariff.

6.3.8.1 Climate Sensitivity The first part of the analysis is to determine the impact on the annual PUE for the four set-points:

- 7°C/45°F chilled water supply with cooling towers set to 5°C/41°F in free cooling mode
- 11°C/52°F chilled water supply with cooling towers set to 9°C/48°F in free cooling mode and chiller Coefficient of Performance (CoP) increased based on higher evaporator temperature
- 15°C/59°F chilled water supply with cooling towers set to 13°C/55°F in free cooling mode and chiller CoP increased based on higher evaporator temperature
- 19°C/66°F chilled water supply with cooling towers set to 17°C/63°F in free cooling mode, chiller CoP as per the 15°C/59°F variant and summer mode cooling tower return set-point increased by 5°C/9°F

The output of the analysis is shown in Figure 6.7 for four different TMY climates selected to show how the response of even this simple change depends on the location and does not follow a rule of thumb for savings. The PUE improvement for Singapore is less than 0.1 as the economizer is never active in this climate and the only benefit is improved mechanical chiller efficiency. St Louis Missouri shows a slightly stronger response, but still only 0.15, as the climate is strongly modal between summer and winter with few hours in the analyzed economizer transition region. Sao Paulo shows a stronger response above 15°C, where the site transitions from mostly mechanical cooling to mostly partial or full economizer. The largest saving is shown in San Jose California with a 0.24 reduction in PUE, which is substantially larger than the 0.1 for Singapore.

6.3.8.2 Energy Cost Both the cost and the charge structure for energy vary greatly across the world. It is common to think of electricity as having a unit kWh cost; but when purchased at data center scale, the costs are frequently more complex; this is particularly true in the US market, where byzantine tariffs with multiple consumption bands and demand charges are common.

To demonstrate the impact of these variations in both energy cost and type of tariff, the earlier analysis for climate sensitivity also includes power tariff data every hour for the climate year:

- Singapore has a relatively high cost of power with peak/off-peak bands and a contracted capacity charge which is unaffected by the economizer implementation as no reduction in peak draw is achieved
- Sao Paulo also has a relatively high cost of power but in this instance on a negotiated flat kWh tariff
- St Louis, Missouri, has a very low kWh charge as it is in the "coal belt" with an additional small capacity charge
- San Jose, California, has a unit kWh charge twice that of St Louis

[2]American Society of Heating Refrigeration and Air Conditioning Engineers.

[3]Indian Society of Heating Refrigerating and Air Conditioning Engineers.

[4]www.weatherundergound.com

[5]Using Romonet Software Suite to perform analysis of the entire data center mechanical and electrical infrastructure with full typical meteorological year climate data.

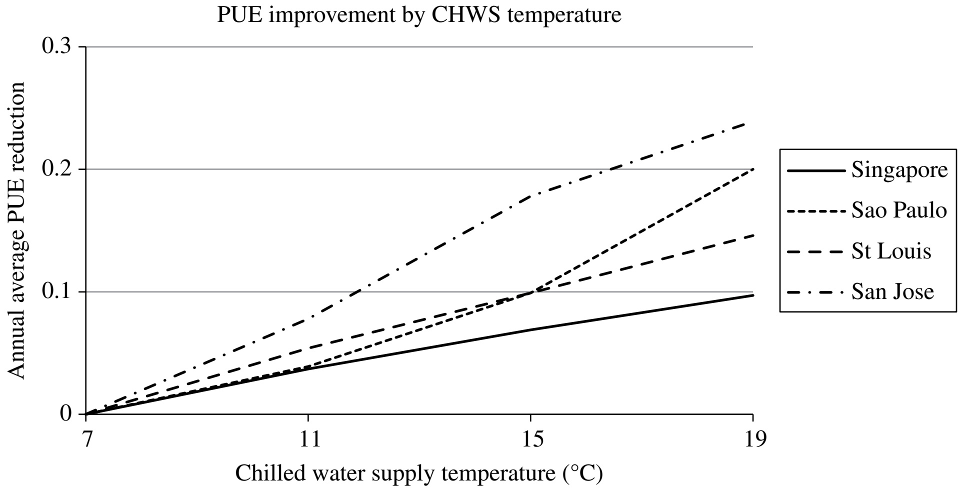


FIGURE 6.7 Climate sensitivity analysis—PUE variation with chilled water supply temperature.

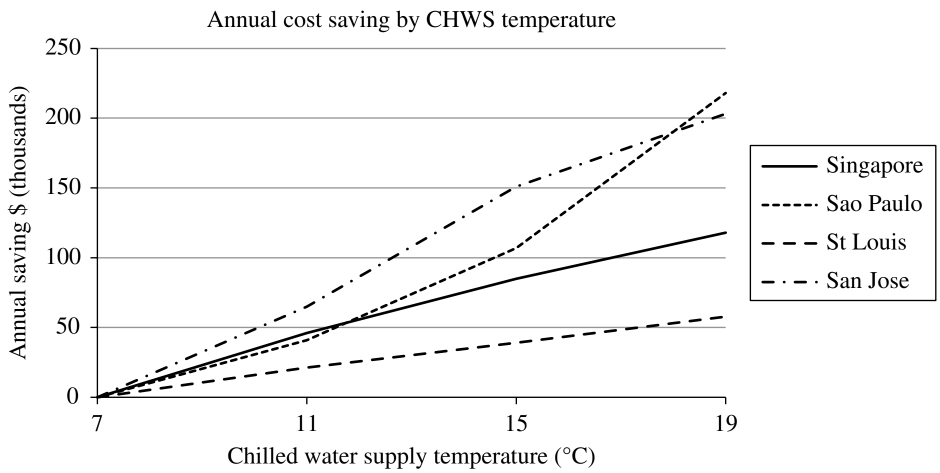


FIGURE 6.8 Energy cost sensitivity analysis—annual cost saving by chilled water supply (CHWS) temperature.

Note that the free cooling energy savings will tend to be larger during off-peak tariff hours and so, to be accurate, the evaluation must evaluate power cost for each hour and not as an average over the period.

The impact of these charge structures is shown in the graph in Figure 6.8. Singapore, despite having only two-third of the PUE improvement of St Louis, achieves more than twice the energy cost saving due to the high cost of power, particularly in peak demand periods. Sao Paolo and San Jose both show large savings but are again in inverse order of their PUE savings.

The cost outcomes shown here show us that we should consider the chilled water system upgrade very differently in St Louis than in San Jose or Sao Paulo.

As with any part of our ROI analysis, these regional energy cost and tariff structure differences are based on the current situation and may well change over time.

No Chiller Data Centers In recent years, the concept of a data center with no compressor-based cooling at all has been popularized with a number of operators building such facilities and claiming financial or environmental benefits due to this elimination of chillers.

While there are some benefits to eliminating the chillers from data centers, the financial benefit is primarily first capital cost, as neither energy efficiency nor energy cost are improved significantly. Depending on the climate the data center operates in, these benefits may come at the cost of the requirement of a substantial expansion of the working environmental range of the IT equipment.

As discussed in the section on free cooling that follows, the additional operational energy efficiency and energy cost benefits of reducing chiller use from a few months per year to never are minimal. There may be substantial first capital cost benefits, however, not just in the purchase and

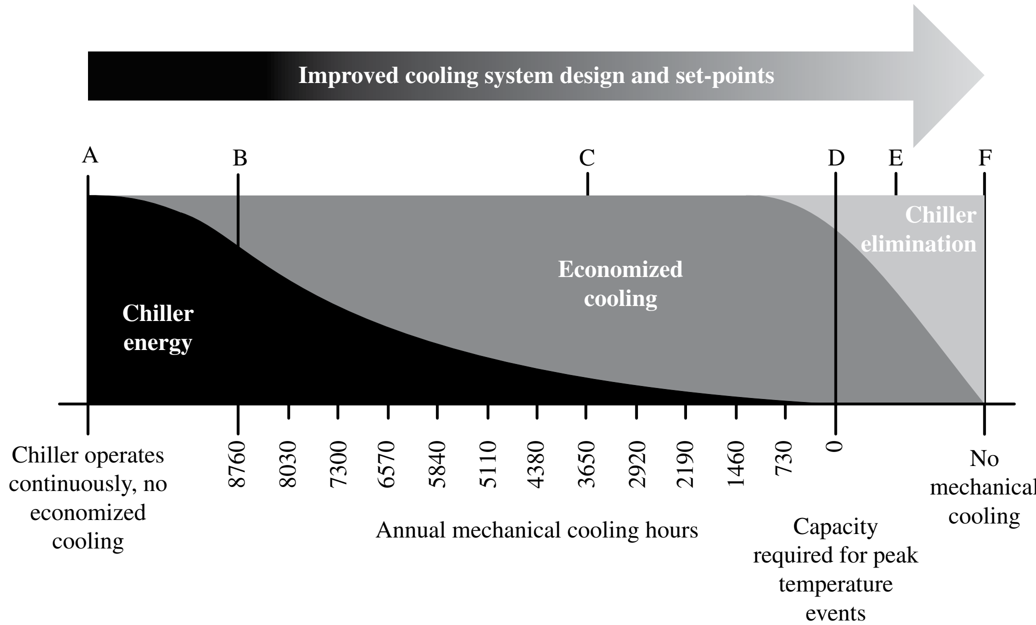


FIGURE 6.9 Chiller energy by economizer hours.

installation cost of the cooling plant but also in the elimination of upstream electrical equipment capacity otherwise required to meet compressor load. Additional operational cost benefits may be accrued through the reduction of peak demand or power availability charges as these peaks will no longer include compressor power.

The balancing factor against the cost benefits of no-chiller designs is the expansion in environmental conditions the IT equipment must operate in. This may be in the form of increased temperature, humidity range, or both. Commonly direct outside air systems will use adiabatic humidifiers to maintain temperature at the expense of high humidity. Other economizer designs are more likely to subject the IT equipment to high temperature peaks during extreme external conditions. The additional concern with no-chiller direct outside air systems is that they cannot revert to air recirculation in the event of an external air pollution event such as dust, smoke, or pollen, which may necessitate an unplanned shutdown of the data center.

Free Cooling, Economizer Hours, and Energy Cost Where a free cooling system is in use, it is quite common to see the performance of the free cooling expressed in terms of "economiser hours," usually meaning the number of hours during which the system requires mechanical compressor cooling. While the type of economizer may vary, from direct external air through to plate heat exchangers for the chilled water loop, the objective of cooling economizers is to reduce the energy consumed to reject the heat from the IT equipment.

As the cooling system design and set-points are improved, it is usual to expect some energy saving. As described earlier in the section on climate sensitivity, the level of energy saving is not linear with the changes in air or water set point temperature; this is not just due to the number of hours in each temperature band in the climate profile but also due to the behavior of the free cooling system.

Figure 6.9 shows a simplified overview of the relationship between mechanical cooling energy, economizer hours, and chiller elimination.

At the far left (A) is a system that relies entirely on mechanical cooling with zero economizer hours—the mechanical cooling energy is highest at this point. Moving to the right (B), the cooling set-points are increased, and this allows for some of the cooling to be performed by the economizer system. Initially, the economizer is only able to reduce the mechanical cooling load, and the mechanical cooling must still run for the full year. As the set points increase further (C), the number of hours per year that the mechanical cooling is required for reduces, and the system moves to primarily economized cooling. When the system reaches zero hours of mechanical cooling (D) in a typical year, it may still require mechanical cooling to deal with peak hot or humid conditions,[6] even though these do not regularly occur. Beyond this point (E), it is common to install mechanical cooling of reduced capacity to supplement the free cooling system. At the far right (F) is a system that is able to meet all of the heat rejection needs even at peak conditions without installing any mechanical cooling at all.

The area marked "Chiller energy" in the chart indicates (approximately, dependent on the system design and detailed climate profile) the amount of energy consumed in

[6]Commonly referred to as the design conditions.

mechanical cooling over the year. This initially falls sharply and then tails off, as the mechanical cooling energy is a function of several variables. As the economized cooling capacity increases,

- The mechanical cooling is run for fewer hours, thus directly using less energy;
- The mechanical cooling operates at part load for many of the hours it is run, as the free cooling system takes part of the load, thus using less energy;
- The mechanical cooling system is likely to work across a smaller temperature differential, thus allowing a reduction in compressor energy, either directly or through the selection of a unit designed to work at a lower temperature differential.

These three factors combine to present a sharp reduction in energy and cost initially as the economizer hours start to increase; this allows for quite substantial cost savings even where only one or two thousand economizer hours are achieved and substantial additional savings for small increases in set-points. As the economized cooling takes over, by point (C), there is very little mechanical cooling energy consumption left to be saved and the operational cost benefits of further increases in set-point are minimal. Once the system is close to zero mechanical cooling hours (D), additional benefit in capital cost may be obtained by reducing or completely eliminating the mechanical cooling capacity installed.

Why the Vendor Case Study Probably Doesn't Apply to You It is normal for vendor case studies to compare the best reasonably credible outcome for their product, service, or technology with a "base case" which is carefully chosen to present the value of their offering in the most positive light possible. In many cases, it is easy to establish that the claimed savings are in fact larger than the energy losses of those parts of your data center which are to be improved and, therefore, quite impossible for you to achieve.

Your data center will have a different climate, energy tariff, existing set of constraints, and opportunities to the site selected for the case study. You can probably also achieve some proportion of the savings with lower investment and disruption; to do so, break down the elements of the savings promised and how else they may be achieved to determine how much of the claimed benefit is actually down to the product or service being sold.

The major elements to consider when determining how representative a case study may be of your situation are as follows:

- Do the climate or IT environmental conditions impact the case study? If so, are these stated and how close to your data center are the values?
- Are there physical constraints of the building or regulatory constraints such as noise which would restrict the applicability?
- What energy tariff was used in the analysis? Does this usefully represent your tariff including peak/off-peak, seasonal, peak demand and availability charge elements?
- How much better than the "before" condition of the case study is your data center already?
- What other, cheaper, faster, or simpler measures could you take in your existing environment to produce some or all of the savings in the case study?
- Was any discount rate included in the financial analysis of the case study? If not, are the full implementation cost and savings shown for you to estimate an NPV or IRR using your internal procedures?

The process shown in the section "A Realistic Example" is a good example of examining how much of the available savings are due to the proposed project and how much may be achieved for less disruption or cost.

6.3.9 IT Power Savings and Multiplying by PUE

If the project you are assessing contains an element of IT power draw reduction, it is common to include the energy cost savings of this in the project analysis. Assuming that your data center is not perfectly efficient and has a PUE greater than 1.0, you may expect some infrastructure overhead energy savings in addition to the direct IT energy savings.

It is common to see justifications for programs such as IT virtualization or server refresh using the predicted IT energy saving and multiplying these by the PUE to estimate the total energy savings. This is fundamentally misconceived; it is well recognized that PUE varies with IT load and will generally increase as the IT load decreases. This is particularly severe in older data centers where the infrastructure overhead is largely fixed and, therefore, responds very little to IT load.

IT power draw multiplied by PUE is not suitable for estimating savings or for chargeback of data center cost. Unless you are able to effectively predict the response of the data center to the expected change in IT load, the predicted change in utility load should be no greater than the IT load reduction.

6.3.10 Converting Other Factors into Cost

When building an ROI case, one of the more difficult elements to deal with is probability and risk. While there is a risk element in creating any forecast into the future, there are some revenues or costs that are more obviously at risk and should be handled more carefully. For example, an upgrade reinvestment business case may improve reliability at the same time as reducing operational costs requiring us to put a value on the reliability improvement. Alternatively, for a service provider, an investment to create additional capacity may rely on additional customer revenue for business justification; there can be no guarantee of the amount or timing of this additional revenue, so some estimate must be used.

6.3.10.1 Attempt to Quantify Costs and Risks For each of the external factors that could affect the outcome of your analysis, make a reasonable attempt to quantify the variables so that you may include them in your assessment. In reality, there are many bad things that may happen to a data center that could cost a lot of money, but it is not always worth investing money to reduce those risks. There are some relatively obvious examples, the cost of adding armor to withstand explosives is unlikely to be an effective investment for a civilian data center but may be considered worthwhile for a military facility.

The evaluation of risk cost can be quite complex and is outside the scope of this chapter. For example, where the cost of an event may vary dependent on the severity of the event, modeling the resultant cost of the risk requires some statistical analysis.

At a simplistic level, if a reasonable cost estimate can be assigned to an event, the simplest way to include the risk in your ROI analysis is to multiply the estimated cost of the event by the probability of it occurring. For example, your project may replace end-of-life equipment with the goal of reducing the risk of a power outage from 5 to 0.1%/year. If the expected cost of the power outage is $500,000 in service credit and lost revenue, then the risk cost would be

- Without the project, 0.05 × $500,000 = $25,000 per annum
- With the project, 0.001 × $500,000 = $500 per annum

Thus, you could include $24,500 per annum cost saving in your project ROI analysis for this mitigated risk. Again, this is a very simplistic analysis, and many organizations will use more effective tools for risk quantification and management, from which you may be able to obtain more effective values.

6.3.10.2 Create a Parameterized Model Where your investment is subject to external variations such as the cost of power over the evaluation time frame, it may be necessary to evaluate how your proposed project performs under a range of values for each external factor. In these cases, it is common to construct a model of the investment in a spreadsheet which responds to the variable external factors and so allows you to evaluate the range of outcomes and sensitivity of the project to changes in these input values.

The complexity of the model may vary from a control cell in a spreadsheet to allow you to test the ROI outcome at $0.08, $0.10, and $0.12 per kWh power cost through to a complex model with many external variables and driven by a Monte Carlo analysis[7] package.

6.3.10.3 A Project That Increases Revenue Example It is not uncommon to carry out a data center project to increase (or release) capacity. The outcome of this is that there is more data center power and cooling capacity to be sold to customers, or cross charged to internal users. It is common in capacity upgrade projects to actually increase the operational costs of the data center by investing capital to allow more power to be drawn and the operational cost to increase. In this case, the NPV or IRR will be negative unless we consider the additional business value or revenue available.

As an example of this approach, a simple example model will be shown which evaluates the ROI of a capacity release project that includes both the possible variance in how long it takes to utilize the additional capacity and the power cost over the project evaluation time frame.

For this project we have

- $100,000 capital cost in year 0
- 75 kW increase in usable IT capacity
- Discount rate of 5%
- Customer power multiplier of 2.0 (customer pays metered kWh × power cost × 2.0)
- Customer kW capacity charge of $500 per annum
- Customer power utilization approximately 70% of contracted
- Estimated PUE of 1.5 (but we expect PUE to fall from this value with increasing load)
- Starting power cost of $0.12 per kWh

From these parameters, we can calculate in any year of the project the additional cost and additional revenue for each extra 1 kW of the released capacity we sell to customers.

We construct our simple spreadsheet model such that we can vary the number of years it takes to sell the additional capacity and the annual change in power cost.

We calculate the NPV as before, at the beginning of our project, year zero, we have the capital cost of the upgrade, $100,000. Then, in each year, we determine the average additional customer kW contracted and drawn based on the number of years, it takes to sell the full capacity. In Table 6.16 is a worked example where it takes 4 years to sell the additional capacity.

The spreadsheet uses a mean and variance parameter to estimate the increase in power cost each year; in this case, the average increase is 3% with a standard deviation of ±1.5%.

[7]A numerical analysis method developed in the 1940s during the Manhattan Project which is useful for modeling phenomena with significant uncertainty in inputs which may be modeled as random variables.

TABLE 6.16 Calculating the NPV for a single trial. Please visit the companion website for an editable example Excel spreadsheet for this table.

Year	0	1	2	3	4	5	6
Annual power cost		$0.120	$0.124	$0.126	$0.131	$0.132	$0.139
Additional kW sold		9	28	47	66	75	75
Additional kW draw		7	20	33	46	53	53
Additional revenue	$0	$18,485	$56,992	$96,115	$138,192	$159,155	$165,548
Additional cost	$100,000	$10,348	$32,197	$54,508	$79,035	$91,241	$96,036
Annual PV	−$100,000	$7,749	$22,490	$35,942	$48,669	$53,212	$51,871
Total PV	−$100,000	−$92,251	−$69,761	−$33,819	$14,850	$68,062	$119,933

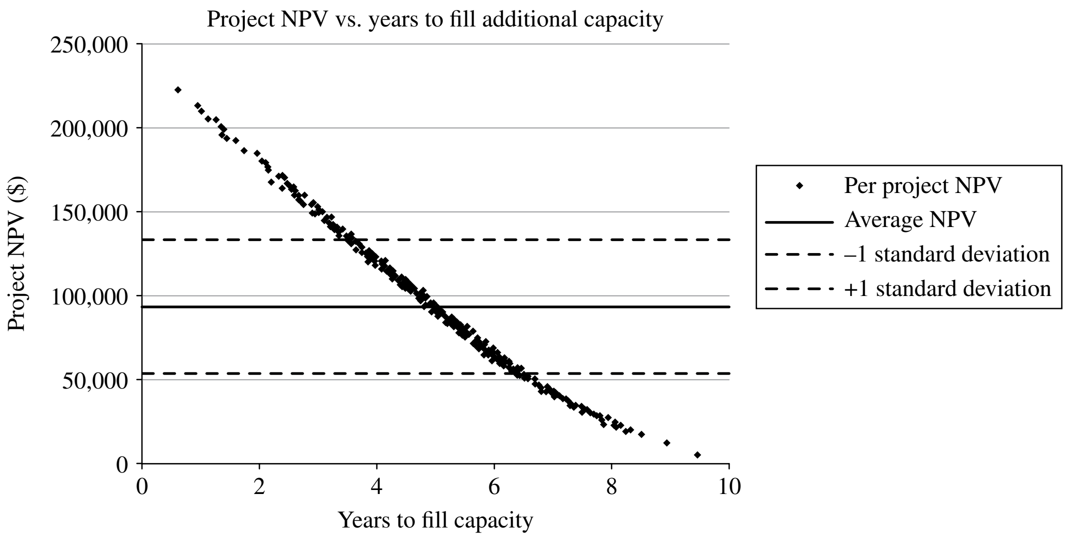


FIGURE 6.10 Simple Monte Carlo analysis of capacity upgrade project. Please visit the companion website for an editable example Excel spreadsheet for this figure.

From the values derived for power cost, contracted and drawn kW, we are able to determine the annual additional revenue and additional cost. Subtracting the cost from the revenue and applying the formula for PV, we can obtain the PV for each year. Summing these provides the total PV across the lifetime—in this case, $119,933, as shown in Table 6.16.

We can use this model in a spreadsheet for a simple Monte Carlo analysis by using some simple statistical functions to generate for each trial:

- The annual power cost increase based on the specified mean and standard deviation of the increase (In this example, I used the NORM.INV[RAND()], mean, standard deviation] function in Microsoft Office Excel to provide the annual increase assuming a normal distribution).
- The number of years before the additional capacity is fully sold (In this example the NORM.INV[RAND()], expected fill out years, standard deviation] function is used, again assuming a normal distribution).

By setting up a reasonably large number of these trials in a spreadsheet, it is possible to evaluate the likely range of financial outcomes and the sensitivity to changes in the external parameters. The outcome of this for 500 trials is shown in Figure 6.10; the dots are the individual trials plotted as years to fill capacity versus achieved NPV; the horizontal lines show the average project NPV across all trials and the boundaries of ± 1 standard deviation.

There are a number of things apparent from the chart:

- Even in the unlikely case of it taking 10 years to sell all of the additional capacity, the overall outcome is still likely to be a small positive return.
- The average NPV is just under $100,000, which against an investment of $100,000 for the capacity release is a reasonable return over the 6-year project assessment time frame.

An alternative way to present the output of the analysis is to perform more trials and then count the achieved NPV of each trial into a bin to determine the estimated probability

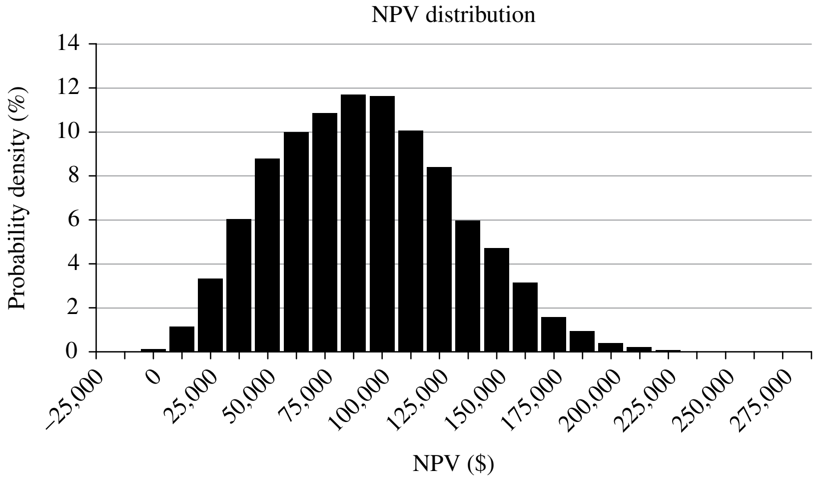


FIGURE 6.11 Probability density plot of simple Monte Carlo analysis.

of an NPV in each range. To illustrate this, 5,000 trials of the earlier example are binned into NPV bands of $25,000 and plotted in Figure 6.11.

6.3.10.4 Your Own Analysis The earlier example is a single, simplistic example of how you might assess the ROI of a project that is subject to one or more external factors. There are likely to be other plots and analyses of the output data that provide insight for your situation, those shown are merely examples. Most spreadsheet packages are capable of Monte Carlo analysis, and there are many worked examples available in the application help and online. If you come to use this sort of analysis regularly, then it may be worth investing in one of the commercial software packages[8] that provide additional tools and capability in this sort of analysis.

6.4 A REALISTIC EXAMPLE

To bring together some of the elements presented in this chapter, an example ROI analysis will be performed for a common reinvestment project. The suggested project is to implement cooling improvements in an existing data center. The example data center

- Has a 1 MW design total IT load
- Uses chilled water CRAC units supplied by a water cooled chiller with cooling towers
- Has a plate heat exchanger for free cooling when external conditions permit with a chilled water supply temperature of 9°C/48°F
- Is located in Atlanta, Georgia, the United States

[8]Such as Palisade @Risk or Oracle Crystal Ball.

The ROI analysis is to be carried out over 6 years using a discount rate of 8% at the request of the finance group.

6.4.1 Air Flow Upgrade Project

There are two proposals provided for the site:

- In-row cooling upgrade with full Hot Aisle Containment (HAC)
- Air flow management and sensor network improvements, upgrade of the existing CRAC units with Electronically Commutated (EC) variable speed fans combined with a distributed temperature sensor network that optimizes CRAC behavior based on measured temperatures

6.4.2 Break Down the Options

While one choice is to simply compare the two options presented with the existing state of the data center, this is unlikely to locate the most effective investment option for our site. In order to choose the best option, we need to break down which changes are responsible for the project savings and in what proportion.

In this example, the proposed cost savings are due to improved energy efficiency in the cooling system. In both options, the energy savings come from the following:

- A reduction in CRAC fan motor power through the use of Variable Speed Drives enabled by reducing or eliminating the mixing of hot return air from the IT equipment with cold supply air from the CRAC unit. This air flow management improvement reduces the volume required to maintain the required environmental conditions at the IT equipment intake.

- A reduction in chilled water system energy consumption through an increase in supply water temperature, also enabled by reducing or eliminating the mixing of hot and cold air. This allows for a small increase in compressor efficiency but more significantly, an increase in the free cooling available to the system.

To evaluate our project ROI, the following upgrade options will be considered.

6.4.2.1 Existing State We will assume that the site does not have existing issues that are not related to the upgrade such as humidity overcontrol or conflicting set-points. If there are any such issues, they should be remediated independently and not confused with the project savings as this would present a false and misleading impression of the project ROI.

6.4.2.2 Proposed Option One—In-Row Cooling The in-row cooling upgrade eliminates 13 of the 15 current perimeter CRAC units and replaces the majority of the data hall cooling with 48 in-row cooling units. The in-row CRAC units use EC variable speed fans operated on differential pressure to reduce CRAC fan power consumption. The HAC allows for an increase in supply air and, therefore, chilled water loop temperature to 15°C/59°F. The increased chilled water supply temperature (CHWS) temperature allows for an increase in achieved free cooling hours as well as a small improvement in operating chiller efficiency. The remaining two perimeter CRAC units are upgraded with a VFD and set to 80% minimum air flow.

6.4.2.3 Proposed Option* Two—*Air Flow Management and Sensor Network The more complex proposal is to implement a basic air flow management program that stops short of air flow containment and is an upgrade of the existing fixed speed fans in the CRAC units to EC variable speed fans. This is coupled with a distributed sensor network, which monitors the supply temperature to the IT equipment. There is no direct saving from the sensor network, but it offers the ability to reduce CRAC fan power and increase in the chilled water supply temperature to allow for more free cooling hours. This option is also evaluated at 15°C/59°F chilled water supply temperature.

6.4.2.4 Air Flow Management and VFD Upgrade Given that much of the saving is from reduced CRAC fan power, we should also evaluate a lower capital cost and complexity option. In this case, the same basic air flow management retrofit as in the sensor network option will be deployed but without the sensor network, a less aggressive improvement in fan speed and chilled water temperature will be achieved. In this case, a less expensive VFD upgrade to the existing CRAC fans will be implemented with a minimum air flow of 80% and fan speed controlled on return air temperature. The site has $N+20\%$ CRAC units, so the 80% airflow will be sufficient even without major reductions in hot/cold remix. The chilled water loop temperature will only be increased to 12°C/54°F.

6.4.2.5 EC Fan Upgrade with Cold Aisle Containment As the in-row upgrade requires the rack layout to be adjusted to allow for HAC, it is worth evaluating a similar option. As the existing CRAC units feed supply air under the raised floor, in this case, Cold Aisle Containment (CAC) will be evaluated with the same EC fan upgrade to the existing CRAC units as in the sensor network option but in this case controlled on differential pressure to meet IT air demand. The contained air flow allows for the same increase in chilled water supply temperature to 15°C/59°F.

6.4.3 Capital Costs

The first step in evaluation is to determine the capitalized costs of the implementation options. This will include capital purchases, installation costs, and other costs directly related to the upgrade project. The costs provided in this analysis are, of course, only examples and, as for any case study, the outcome may or may not apply to your data center.

- The air flow management and HAC/CAC include costs for both air flow management equipment and installation labor.
- The In-Row CRAC unit costs are estimated to cost 48 units × $10,000 each.
- The In-Row system also requires four Coolant Distribution Units and pipework at a total of $80,000.
- The 15 CRAC units require $7000 upgrades of fans and motors for the two EC fan options.
- The distributed temperature sensor network equipment, installation, and software license are $100,000.
- Each of the options requires a $20,000 Computational Fluid Dynamic analysis; prior to implementation, this cost is also capitalized.

The total capitalized costs of the options are shown in Table 6.17.

6.4.4 Operational Costs

The other part of the ROI assessment is the operational cost impact of each option. The costs of all options are affected by both the local climate and the power cost. The local climate is represented by a Typical Meteorological Year climate data set in this analysis.

The energy tariff for the site varies peak and off-peak as well as summer to winter, averaging $0.078 in the first year.

TABLE 6.17 Capitalized costs of project options. Please visit the companion website for an editable example Excel spreadsheet for this table.

	Existing state	Air flow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Air flow management		$100,000			$100,000
HAC/CAC			$250,000	$250,000	
In-row CRAC			$480,000		
CDU and pipework			$80,000		
EC fan (brushless direct current fan) upgrade				$105,000	$105,000
VFD fan upgrade		$60,000	$8,000		
Sensor network					$100,000
CFD analysis		$20,000	$20,000	$20,000	$20,000
Total capital	$0	$180,000	$838,000	$375,000	$325,000

TABLE 6.18 Analyzed annual PUE of the upgrade options. Please visit the companion website for an editable example Excel spreadsheet for this table.

Option	PUE
Existing state	1.92
Air flow management and VFD fan	1.72
In-row cooling	1.65
EC fan upgrade and CAC	1.63
AFM, EC fan, and sensor network	1.64

This is then subject to a 3% annual growth rate to represent an expected increase in European energy costs.

6.4.4.1 Efficiency Improvements Analysis[9] of the data center under the existing state and upgrade conditions yields the achieved annual PUE results shown in Table 6.18.

These efficiency improvements do not translate directly to energy cost savings as there is an interaction between the peak/off-peak, summer/winter variability in the energy tariff and the external temperature, which means that more free cooling hours occur at lower energy tariff rates. The annual total energy costs of each option are shown in Table 6.19.

6.4.4.2 Other Operational Costs As an example of other cost changes due to a project, the cost of quarterly CFD air flow analysis has been included in the operational costs. The use of CFD analysis to adjust air flow may continue under the non-contained air flow options, but CFD becomes unnecessary once either HAC or CAC is implemented, and this cost becomes a saving of the contained air flow options. The 6-year operational costs are shown in Table 6.19.

6.4.5 NPV Analysis

To determine the NPV of each option, we first need to determine the PV of the future operational costs at the specified discount rate of 8%. This is shown in Table 6.20.

[9]The analysis was performed using Romonet Software Suite simulating the complete mechanical and electrical infrastructure of the data center using full typical meteorological year climate data.

The capitalized costs do not need adjusting as they occur at the beginning of the project. Adding together the capitalized costs and the total of the operational PVs provides a total PV for each option. The NPV of each upgrade option is the difference between the total PV for the existing state and the total PV for that option as shown in Table 6.21.

6.4.6 IRR Analysis

The IRR analysis is performed with the same capitalized and operational costs but without the application of the discount rate. To set out the costs so that they are easy to supply to the IRR function in a spreadsheet package, we will subtract the annual operational costs of each upgrade option from the baseline costs to give the annual saving as shown in Table 6.22.

From this list of the first capital cost shown as a negative number and the annual incomes (savings) shown as positive numbers, we can use the IRR function in the spreadsheet to determine the IRR for each upgrade option.

6.4.7 Return Analysis

We now have the expected change in PUE, the NPV, and the IRR for each of the upgrade options. The NPV and IRR of the existing state are zero, as this is the baseline against which the other options are measured. The analysis summary is shown in Table 6.23.

It is perhaps counter intuitive that there is little connection between the PUE improvement and the ROI for the upgrade options.

The air flow management and VFD fan upgrade option has the highest IRR and the highest ratio of NPV to invested capital. The additional $145,000 capital investment for the EC fans and distributed sensor network yields only a $73,000 increase in the PV, thus the lower IRR of only 43% for this option. The base air flow management has already provided a substantial part of the savings and the incremental improvement of the EC fan and sensor network is small. If we have

TABLE 6.19 Annual operational costs of project options. Please visit the companion website for an editable example Excel spreadsheet for this table.

	Existing state	Air flow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Annual CFD analysis	$40,000	$40,000			$40,000
Year 1 energy	$1,065,158	$957,020	$915,394	$906,647	$912,898
Year 2 energy	$1,094,501	$983,437	$940,682	$931,691	$938,117
Year 3 energy	$1,127,336	$1,012,940	$968,903	$959,642	$966,260
Year 4 energy	$1,161,157	$1,043,328	$997,970	$988,432	$995,248
Year 5 energy	$1,198,845	$1,077,134	$1,030,284	$1,020,439	$1,027,474
Year 6 energy	$1,231,871	$1,106,866	$1,058,746	$1,048,627	$1,055,858

TABLE 6.20 NPV analysis of project options at 8% discount rate. Please visit the companion website for an editable example Excel spreadsheet for this table.

	Existing state	Air flow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
6 years CFD analysis PV	$184,915	$184,915	$0	$0	$184,915
Year 1 energy PV	$986,258	$886,129	$847,587	$839,488	$845,276
Year 2 energy PV	$938,359	$843,138	$806,483	$798,775	$804,284
Year 3 energy PV	$894,916	$804,104	$769,146	$761,795	$767,048
Year 4 energy PV	$853,485	$766,877	$733,537	$726,527	$731,537
Year 5 energy PV	$815,914	$733,079	$701,194	$694,493	$699,282
Year 6 energy PV	$776,288	$697,514	$667,190	$660,813	$665,370

TABLE 6.21 NPV of upgrade options. Please visit the companion website for an editable example Excel spreadsheet for this table.

	Existing state	Air flow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Capital	$0	$180,000	$838,000	$375,000	$325,000
PV Opex	$5,450,134	$4,915,757	$4,525,136	$4,481,891	$4,697,712
Total PV	$5,450,134	$5,095,757	$5,363,136	$4,856,891	$5,022,712
NPV	$0	$354,377	$86,997	$593,243	$427,422

TABLE 6.22 IRR analysis of project options. Please visit the companion website for an editable example Excel spreadsheet for this table.

Option	Existing state	Air flow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Capital cost	$0	−$180,000	−$838,000	−$375,000	−$325,000
Year 1 savings	$0	$108,139	$189,765	$198,512	$152,261
Year 2 savings	$0	$111,065	$193,820	$202,810	$156,385
Year 3 savings	$0	$114,397	$198,434	$207,694	$161,076
Year 4 savings	$0	$117,829	$203,187	$212,725	$165,909
Year 5 savings	$0	$121,711	$208,561	$218,406	$171,371
Year 6 savings	$0	$125,005	$213,125	$223,244	$176,013

TABLE 6.23 Overall return analysis of project options. Please visit the companion website for an editable example Excel spreadsheet for this table.

	Existing state	Air flow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Capital	$0	$180,000	$838,000	$375,000	$325,000
PUE	1.92	1.72	1.65	1.63	1.64
NPV	$0	$354,377	$86,997	$593,243	$427,422
IRR	0%	58%	11%	50%	43%
Profitability index		2.97	1.10	2.58	2.32

other projects with a similar return to the base air flow management and VFD fan upgrade on which we could spend the additional capital of the EC fans and sensor network, these would be better investments. The IRR of the sensor network in addition to the air flow management is only 23%, which would be unlikely to meet approval as an individual project.

The two air flow containment options have very similar achieved PUE and operational costs; they are both quite efficient and neither requires CFD or movement of floor tiles. There is, however, a substantial difference in the implementation cost; so despite the large energy saving, the in-row cooling option has the lowest return of all the options while the EC fan upgrade and CAC has the highest NPV.

It is interesting to note that there is no one "best" option here as the air flow management and VFD fan has the highest IRR and highest NPV per unit capital, while the EC fan upgrade and CAC has the highest overall NPV.

6.4.8 Break-Even Point

We are also likely to be asked to identify the break-even point for our selected investments; we can do this by taking the PV in each year and summing these over time. We start with a negative value for the year 0 capitalized costs and then add the PV of each year's operational cost saving over the 6-year period. The results are shown in Figure 6.12.

The break-even point is where the cumulative NPV of each option crosses zero. Three of the options have a break-even point of between 1.5 and 2.5 years, while the in-row cooling requires 5.5 years to break even.

6.4.8.1 Future Trends This section examines the impact of the technological and financial changes on the data center market, and how these may impact the way you run your data center or even dispose of it entirely. Most of the future trends affecting data centers revolve around the commoditization of data center capacity and the change in focus from technical performance criteria to business financial criteria. Within this is the impact of cloud, consumerization of ICT, and the move toward post-PUE financial metrics of data center performance.

6.4.8.2 The Threat of Cloud and Commoditization At the time of writing, there is a great deal of hype about cloud computing and how it will turn IT services into utilities such as water or gas. This is a significant claim that the changes of cloud will erase all distinctions between IT services and that any IT service may be transparently substituted with any other IT service. If this were to come true, then IT would be subject to competition on price alone with no other differentiation between services or providers.

Underneath the hype, there is little real definition of what actually constitutes "cloud" computing, with everything from free webmail to colocation services, branding itself as cloud. The clear trend underneath the hype, however, is the commoditization of data center and IT resources. This is facilitated by a number of technology changes including:

- Server, storage, and network virtualization at the IT layer have substantially reduced the time, risk, effort, and cost of moving services from one data center to another. The physical location and ownership of IT equipment is of rapidly decreasing importance.
- High-speed Internet access is allowing the large scale deployment of network dependent end user computing devices; these devices tend to be served by centralized platform vendors such as Apple, Microsoft, or Amazon rather than corporate data centers.
- Web-based application technology is replacing many of the applications or service components that were previously run by enterprise users. Many organizations now

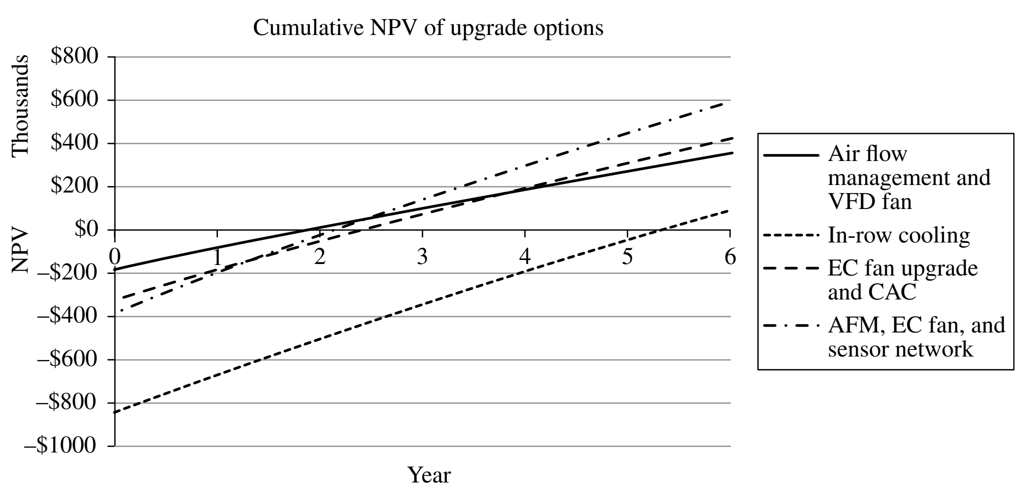


FIGURE 6.12 Break-even points of upgrade options. Please visit the companion website for an editable example Excel spreadsheet for this figure.

select externally operated platforms such as SalesForce because of their integration with other web-based applications instead of requiring integration with internal enterprise systems.

6.4.8.3 Data Center Commoditization Data centers are commonly called the factories of Information Technology; unfortunately, they are not generally treated with the same financial rigor as factories. While the PUE of new data centers may be going down (at least in marketing materials), the data center market is still quite inefficient. Evidence of this can be seen in the large gross margins made by some operators and the large differences in price for comparable products and services at both M&E device and data center levels.

The process of commoditization will make the market more efficient, to quote one head of data center strategy "this is a race to the bottom and the first one there wins." This recognition that data centers are a commodity will have significant impacts not just on the design and construction of data centers but also on the component suppliers who will find it increasingly hard to justify premium prices for heavily marketed but nonetheless commodity products.

In general, commoditization of a product is the process of the distinguishing factors becoming less relevant to the purchaser and thereby becoming simple commodities. In the data center case, commoditization comes about through several areas of change:

- Increased portability—it is becoming faster, cheaper, and easier for customers of data center capacity or services delivered from data centers to change supplier and move to another location or provider. This prevents "lock-in," and so increases the impact of price competition among suppliers.
- Reductions in differentiating value—well-presented facilities with high levels of power and cooling resilience or availability certifications are of little value in a world where customers neither know nor care which data center their services are physically located in, and service availability is handled at the network and software level.
- Broadening availability of the specific knowledge and skills required to build and operate a financially efficient data center; while this used to be the domain of a few very well informed experts, resources such as the EU Code of Conduct on Data Centers and effective predictive financial and operational modeling of the data center are making these capabilities generally available.
- Factory assembly of components through to entire data centers being delivered as modules, so reducing the capital cost of delivering new data center capacity compared to traditional on-site construction.
- Business focus on financial over technical performance metrics.

While there are many barriers obstructing IT services or data centers from becoming truly undifferentiated utility commodities, such as we see with water or oil, much of the differentiation, segmentation, and price premium that the market has so far enjoyed is disappearing. There will remain some users for whom there are important factors such as physical proximity to, or distance from, other locations, but even in these cases it is likely that only the minimum possible amount of expensive capacity will be deployed to meet the specific business issue and the remainder of the requirement will be deployed across suitable commodity facilities or providers.

6.4.8.4 Driving Down Cost in the Data Center Market Despite the issues that are likely to prevent IT from ever becoming a completely undifferentiated commodity such as electricity or gas, it is clear that the current market inefficiencies will be eroded and the cost of everything from M&E equipment to managed application services will fall. As this occurs, both enterprise and service provider data centers will have to substantially reduce cost in order to stay competitive.

Enterprise data centers may

- Improve both their cost and flexibility closer to that offered by cloud providers to reduce the erosion of internal capacity and investment by low capital and short commitment external services.
- Target their limited financial resource and data center capacity to services with differentiating business value or high business impact of failure, while exporting commodity services which may be cheaply and effectively delivered by other providers.
- Deliver multiple grades of data center at multiple cost levels to meet business demands and facilitate a functioning internal market.

Cloud providers are likely to be even more vulnerable than enterprise data centers as their applications are, almost by definition, commodity, fast and easy to replace with a cheaper service. It is already evident that user data is now the portability issue and that some service providers resist competition by making data portability for use in competitive services as difficult as possible.

6.4.8.5 Time Sensitivity One of the key issues in the market for electricity is our present inability to economically store any large quantity of it once generated. The first impact of this is that sufficient generating capacity to meet peak demand must be constructed at high capital cost but not necessarily full utilization. The second is the substantial price fluctuation over short time frames with high prices at demand peaks and low prices when there is insufficient demand to meet the available generating capacity.

For many data centers, the same issue exists, the workload varies due to external factors and the data center must be sized to meet peak demand. Some organizations are able to schedule some part of their data center workload to take place during low load periods, for example, web crawling and construction of the search index when not serving search results. For both operators purchasing capacity and cloud providers selling it through markets and brokers, price fluctuation and methods of modifying demand schedules are likely to be an important issue.

6.4.8.6 Energy Service Contracts Many data center operators are subject to a combination of capital budget reductions and pressure to reduce operational cost or improve energy efficiency. While these two pressures may seem to be contradictory, there is a financial mechanism which is increasingly used to address this problem.

In the case where there are demonstrable operational cost savings available from a capital upgrade to a data center, it is possible to fund the capital reinvestment now from the later operational savings. While energy service contracts take many forms, they are in concept relatively simple:

1. The expected energy cost savings over the period are assessed.
2. The capitalized cost of the energy saving actions including equipment and implementation are assessed.
3. A contract is agreed and a loan is provided or obtained for the capitalized costs of the implementation; this loan funds some or all of the project implementation costs and deals with the capital investment hurdle.
4. The project is implemented and the repayments for the loan are serviced from some or all of the energy cost savings over the repayment period.

Energy service contracts are a popular tool for data center facilities management outsourcing companies. While the arrangement provides a mechanism to reduce the up-front cost of an energy performance improvement for the operator, there are a number of issues to consider:

- The service contract tends to commit the customer to the provider for an extended period; this may be good for the provider and reduce direct price competition for their services.
- There is an inherent risk in the process for both the provider and customer; the cost savings on which the loan repayments rely may either not be delivered or it may not be possible to prove that they have been delivered due to other changes, in which case responsibility for servicing the loan will still fall to one of the parties.
- There may be a perverse incentive for outsource facilities management operators to "sandbag" on operational changes, which would reduce energy in order to use these easy savings in energy service contract-funded projects.

6.4.8.7 Guaranteed Performance and Cost The change in focus from technical to financial criteria for data centers coupled with the increasing brand value importance of being seen to be energy efficient is driving a potentially significant change in data center procurement. It is now increasingly common for data center customers to require their design or build provider to state the achieved PUE or total energy consumption of their design under a set of IT load fill out conditions. This allows the customer to make a more effective TCO optimization when considering different design strategies, locations, or vendors.

The logical extension of this practice is to make the energy and PUE performance of the delivered data center part of the contractual terms. In these cases, if the data center fails to meet the stated PUE or energy consumption, then the provider is required to pay a penalty. Contracts are now appearing, which provide a guarantee that if the data center fails to meet a set of PUE and IT load conditions the supplier will cover the additional energy cost of the site.

The form of these guarantees varies from a relatively simple, above a certain kW load the PUE, when measured as defined, will be at or below the guaranteed performance through to more complex definitions of performance at varying IT load points or climate conditions.

A significant issue for some purchasers of data centers is the split incentive inherent in many of the build or lease contracts currently popular. It is common for the provider of the data center to pay the capital costs of construction but to have no financial interest in the operational cost or efficiency. In these cases, it is not unusual for capital cost savings to be made directly at the expense of the ongoing operational cost of the data center, which results in a substantial increase in the total TCO and poor overall performance. When purchasing or leasing a data center, it is essential to ensure that the provider constructing the data center has a financial interest in the operational performance and cost to mitigate these incentives. This is increasingly taking the form of energy performance guarantees that share the impact of poor performance with the supplier.

6.4.8.8 Charging for the Data Center—Activity-Based Costing With data centers representing an increasing proportion of the total business operating cost and more business activity becoming critically reliant upon those data centers, a change is being forced in the way in which finance departments treat data centers. It is becoming increasingly unacceptable for the cost of the data center to be treated as a centralized operating overhead or to be distributed across business units with a fixed finance "allocation formula" which is often out of date and has little basis in reality. Many businesses are attempting to institute some level of chargeback model to apply the costs of their data center resources to the (hopefully value-generating) business units that demand and consume them.

These chargeback models vary a great deal in their complexity and accuracy all the way from square feet, through to detailed and realistic activity-based costing models. For many enterprises, this is further complicated by a mix of data center capacity that is likely to be made up of the following:

- One or more of their own data centers, possibly in different regions with different utility power tariffs and at different points in their capital amortization and depreciation
- One or more areas of colocation capacity, possibly with different charging models as well as different prices, dependent upon the type and location of facility
- One or more suppliers of cloud compute capacity, again with varying charging mechanisms, length of commitment, and price

Given this mix of supply, it is inevitable that there will be tension and price competition between the various sources of data center capacity to any organization. Where an external colo or cloud provider is perceived to be cheaper, there will be a pressure to outsource capacity requirements. A failure to accurately and effectively cost internal resources for useful comparison with outsourced capacity may lead to the majority of services being outsourced, irrespective of whether it makes financial or business sense to do so.

6.4.8.9 The Service Monoculture Perhaps the most significant issue facing data center owners and operators is the service monoculture that has been allowed to develop and remains persistent by a failure to properly understand and manage data center cost. The symptoms of this issue are visible across most types of organization, from large enterprise operators with legacy estates through colocation to new build cloud data centers. The major symptoms are a single level of data center availability, security, and cost with the only real variation being due to local property and energy costs. It is common to see significant data center capacity built to meet the availability, environmental, and security demands of a small subset of the services to be supported within it.

This service monoculture leads to a series of problems which, if not addressed, will cause substantial financial stress for all types of operator as the data center market commoditizes, margins reduce, and price pressure takes effect.

As an example of this issue, we may consider a fictional financial services organization that owns a data center housing a mainframe which processes customer transactions in real time. A common position for this type of operator when challenged on data center cost efficiency is that they don't really care what the data center housing the mainframe costs, as any disruption to the service would cost millions of dollars per minute and the risk cost massively outweighs any possible cost efficiencies. This position fails to address the reality that the operator is likely to be spending too much money on the data center for no defined business benefit while simultaneously under-investing in the critical business activity. Although the mainframe is indeed business critical, the other 90% plus of the IT equipment in the data center is likely to range from internal applications through to development servers with little or no real impact of downtime. The problem for the operator is that the data center design, planning, and operations staff are unlikely to have any idea which servers in which racks could destroy the business and which have not been used for a year and are expensive fan heaters.

This approach to owning and managing data center resources may usefully be compared to Soviet Union era planned economies. A central planning group determines the amount of capacity that is expected to be required, provides investment for, and orders the delivery of this capacity. Business units then consume the capacity for any requirement they can justify and, if charged at all, pay a single fixed internal rate. Attempts to offer multiple grades and costs of capacity are likely to fail as there is no incentive for business units to choose anything but the highest grade of capacity unless there is a direct impact on their budget. The outcomes in the data center or the planned economy commonly include insufficient provision of key resources, surplus of others, suboptimal allocation, slow reaction of the planning cycle to demand changes, and centrally dictated resource pricing.

6.4.8.10 Internal Markets—Moving Away from the Planned Economy The increasing use of data center service chargeback within organizations is a key step toward addressing the service monoculture problem. To develop a functioning market within the organization, a mixture of internal and external services, each of which has a cost associated with acquisition and use, is required. Part of the current momentum toward use of cloud services is arguably not due to any inherent efficiency advantages of cloud but simply due to the ineffective internal market and high apparent cost of capacity within the organization, allowing external providers to undercut the internal resources.

As organizations increasingly distribute their data center spend across internal, colocation, and cloud resources and the cost of service is compared with the availability, security, and cost of each consumed resource, there is a direct opportunity for the organization to better match the real business needs by operating different levels and costs of internal capacity.

6.4.8.11 Chargeback Models and Cross Subsidies The requirement to account or charge for data center resources within both enterprise and service provider organizations has led to the development of a number of approaches to determining the cost of capacity and utilization. In many cases, the early mechanisms have focused on data gathering and measurement precision at the expense of the accuracy of the cost allocation method itself.

Each of the popular chargeback models, some of which are introduced in the following, has its own balance of strengths and weaknesses and creates specific perverse incentives. Many of these weaknesses stem from the difficulty in dealing with the mixture of fixed and variable costs in the data center. There are some data center costs that are clearly fixed, that is, they do not vary with the IT energy consumption, such as the capital cost of construction, staffing, rent and property taxes. Others, such as the energy consumption at the IT equipment, are obviously variable cost elements.

6.4.8.12 Metered IT Power Within the enterprise, it is common to see metering of the IT equipment power consumption used as the basis for chargeback. This metered IT equipment energy is then multiplied by a measured PUE and the nominal energy tariff to arrive at an estimate of total energy cost for the IT loads. This frequently requires expensive installation of metering equipment coupled with significant data gathering and maintenance requirements to identify which power cords are related to which delivered service. The increasing use of virtualization and the portability of virtual machines across the physical infrastructure present even more difficulties for this approach.

Metered IT power × PUE × tariff is a common element of the cost in colocation services where it is seen by both the operator and client as being a reasonably fair mechanism for determining a variable element of cost. The metering and data overheads are also lower as it is generally easier to identify the metering boundaries of colo customer areas than IT services. In the case of colocation, however, the metered power is generally only part of the contract cost.

The major weakness of metered IT power is that it fails to capture the fixed costs of the data center capacity occupied by each platform or customer. Platforms or customers with a significant amount of allocated capacity but relatively low draw are effectively subsidized by others which use a larger part of their allocated capacity.

6.4.8.13 Space Historically, data center capacity was expressed in terms of square feet or square meters, and therefore, costs and pricing models were based on the use of space while the power and cooling capacity was generally given in kW per square meter or foot. Since that time, the power density of the IT equipment has risen, transferring the dominant constraint to the power and cooling capacity. Most operators charging for space were forced to apply power density limits, effectively changing their charging proxy to kW capacity. This charging mechanism captures the fixed costs of the data center very effectively but is forced to allocate the variable costs as if they were fixed and not in relation to energy consumption.

Given that the majority of the capital and operational costs for most modern data centers are related to the kW capacity and applied kW load, the use of space as a weak proxy for cost is rapidly dying out.

6.4.8.14 Kilowatt Capacity or Per Circuit In this case, the cost is applied per kilowatt capacity or per defined capacity circuit provided. This charge mechanism is largely being replaced by a combination of metered IT power and capacity charge for colocation providers, as the market becomes more efficient and customers better understand what they are purchasing. This charging mechanism is still popular in parts of North America and some European countries where local law makes it difficult to resell energy.

This mechanism has a similar weakness and, therefore, exploitation opportunity to metered IT power. As occupiers pay for the capacity allocated irrespective of whether they use it, those who consume the most power from each provided circuit are effectively subsidized by those who consume a lower percentage of their allocated capacity.

6.4.8.15 Mixed kW Capacity and Metered IT Power Of the top–down charge models, this is perhaps the best representation of the fixed and variable costs. The operator raises a fixed contract charge for the kilowatt capacity (or circuits, or space as a proxy for kilowatt capacity) and a variable charge based on the metered IT power consumption. In the case of colocation providers, the charge for metered power is increasingly "open book" in that the utility power cost is disclosed and the PUE multiplier stated in the contract allowing the customer to understand some of the provider margin. The charge for allocated kW power and cooling capacity is based on the cost of the facility and amortizing this over the period over which this cost is required to be recovered. In the case of colocation providers, these costs are frequently subject to significant market pressures, and there is limited flexibility for the provider.

This method is by no means perfect; there is no real method of separating fixed from variable energy costs, and it is also difficult to deal with any variation in the class and, therefore, cost of service delivered within a single data center facility.

6.4.8.16 Activity-Based Costing As already described, two of the most difficult challenges for chargeback models are separating the fixed from variable costs of delivery and differentially costing grades of service within a single facility or campus. None of the top–down cost approaches discussed so far is able to properly meet these two criteria, except in the extreme case of completely homogenous environments with equal utilization of all equipment.

An approach popular in other industries such as manufacturing is to cost the output product as a supply chain, considering all of the resources used in the production of the product including raw materials, energy, labor, and licensing. This methodology, called Activity-Based Costing, may be applied to the data center quite effectively to produce not just effective costing of resources but to allow for the simultaneous delivery of multiple service levels with properly understood differences in cost. Instead of using fixed allocation percentages for different elements, ABC works by identifying relationships in the supply chain to objectively assign costs.

By taking an ABC approach to the data center, the costs of each identifiable element, from the land and building, through mechanical and electrical infrastructure to staffing and power costs, are identified and allocated to the IT resources that they support. This process starts at the initial resources, the incoming energy feed, and the building and passes costs down a supply chain until they arrive at the IT devices, platforms, or customers supported by the data center.

Examples of how ABC may result in differential costs are as follows:

- If one group of servers in a data hall has single-corded feed from a single $N+1$ UPS room, while another is dual-corded and fed from two UPS rooms giving $2(N+1)$ power, the additional capital and operational cost of the second UPS room would only be borne by the servers using dual-corded power.
- If two data halls sharing the same power infrastructure operate at different temperature and humidity control ranges to achieve different free cooling performance and cost, this is applied effectively to IT equipment in the two halls.

For the data center operator, the most important outcomes of ABC are as follows:

- The ability to have a functioning internal and external market for data center capacity, and thereby invest in and consume the appropriate resources.
- The ability to understand whether existing or new business activities are good investments. Specifically, where business activities require data center resources, the true cost of these resources should be reflected in the cost of the business activity.

For service providers, this takes the form of per customer margin assessment and management. It is not unusual to find that through cross subsidy between customers; frequently, the largest customers (usually perceived as the most valuable) are in fact among the lowest margin and being subsidized by others, to whom less effort is devoted to retaining their business.

6.4.8.17 Unit Cost of Delivery: \$/kWh The change in focus from technical to financial performance metrics for the data center is also likely to change focus from the current engineering-focused metrics such as PUE to more financial metrics for the data center. PUE has gained mind share through being both simple to understand and being an indicator of cost efficiency. The use of activity-based costing to determine the true cost of delivery of data center loads provides the opportunity to develop metrics that capture the financial equivalent of the PUE, the unit cost of each IT kWh, or \$/kWh.

This metric is able to capture a much broader range of factors for each data center, such as a hall within a data center or individual load, than PUE can ever do. The capital or lease cost of the data center, staffing, local taxes, energy tariff, and all other costs may be included to understand the fully loaded unit cost. This may then be used to understand how different data centers within the estate compare with each other and how internal capacity compares for cost with outsourced colocation or cloud capacity.

When investment decisions are being considered, the use of full unit cost metrics frequently produces what are initially counter-intuitive results. As an example, consider an old data center for which the major capital cost is considered to be amortized, operating in an area where utility power is relatively cheap, but with a poor PUE; we may determine the unit delivery cost to be 0.20 \$/kWh, including staffing and utility energy. It is not uncommon to find that the cost of a planned replacement data center which, despite having a very good PUE, once the burden of the amortizing capital cost is applied, cannot compete with the old data center. Frequently, relatively minor reinvestments in existing capacity are able to produce lower unit costs of delivery than even a PUE = 1 new build.

An enterprise operator may use the unit cost of delivery to compare multiple data centers owned by the organization and to establish which services should be delivered from internal versus external resources, including allocating the appropriate resilience, cost, and location of resource to services.

A service provider may use unit cost to meet customer price negotiation by delivering more than one quality of service at different price points while properly understanding the per deal margin.

6.5 CHOOSING TO BUILD, REINVEST, LEASE, OR RENT

A major decision for many organizations is whether to invest building new data center capacity, reinvest in existing, lease capacity, colocate, or use cloud services. There is, of course, no one answer to this; the correct answer for many organizations is neither to own all of their own capacity nor to dispose of all of it and trust blindly in the cloud. At the simplest level, colocation providers and cloud service providers need to make a profit and, therefore, must achieve improvements in delivery cost over that which you can achieve, which are at least equal to the required profit to even achieve price parity.

The choice of how and where to host each of your internal or customer-facing business services depends on a range of factors, and each option has strengths and weaknesses. For many operators, the outcome is likely to be a mix of the following:

- High failure impact services, high security requirement services, or real differentiating business value operated in owned or leased data centers that are run close to capacity to achieve low unit cost
- Other services that warrant ownership and control of the IT equipment or significant network connectivity operated in colocation data centers
- Specific niche and commodity services such as email which are easily outsourced, supplied by low cost cloud providers
- Short-term capacity demands and development platforms delivered via cloud broker platforms which auction for the current lowest cost provider.

As a guide, some of the major benefits and risks of each type of capacity are described in the following. This list is clearly neither exhaustive nor complete but should be considered a guide as to the questions to ask.

6.5.1 Owned Data Center Capacity

Data center capacity owned by the organization may be known to be located in the required legal jurisdiction, operated at the correct level of security, maintained to the required availability level, and operated to a high level of efficiency. It is no longer difficult to build and operate a data center with a good PUE. Many facilities management companies provide the technical skills to maintain the data center at competitive rates, eliminating another claimed economy of scale by the larger operators. In the event of an availability incident, the most business critical platforms may be preferentially maintained or restored to service. In short, the owner controls the data center.

The main downside of owning capacity is the substantial capital and ongoing operational cost commitment of building a data center although this risk is reduced if the ability to migrate out of the data center and sell it is included in the assessment.

The two most common mistakes are the service monoculture, building data center capacity at a single level of service, quality and cost, and failing to run those data centers at full capacity. The high fixed cost commitments of the data center require that high utilization be achieved to operate at an effective unit cost, while migrating services out of a data center you own into colo or cloud simply makes the remainder more expensive unless you can migrate completely and dispose of the asset.

6.5.2 Leased Data Center Capacity

Providers of wholesale or leased data center capacity claim that their experience, scale, and vendor price negotiation leverage allow them to build a workable design for a lower capital cost than the customer would achieve.

Leased data center capacity may be perceived as reducing the capital cost commitment and risk. However, in reality the capital cost has still been financed and a loan is being serviced. Furthermore, it is frequently as costly and difficult to get out of a lease as it is to sell a data center you own.

The risk defined in Section 6.4.8.6 may be mitigated by ensuring contractual commitments by the supplier to the ongoing operational cost and energy efficiency of the data center.

As for the owned capacity, once capacity is leased, it should generally be operated at high levels of utilization to keep the unit cost acceptable.

6.5.3 Colocation Capacity

Colocation capacity is frequently used in order to leverage the connectivity available at the carrier neutral data center operators. This is frequently of higher capacity and lower cost than may be obtained for your own data center; where your services require high-speed and reliable Internet connectivity, this is a strong argument in favor of colocation. There may also be other bandwidth-intensive services available within the colocation data center made available at lower network transit costs within the building than would be incurred if those services were to be used externally.

It is common for larger customers to carry out physical and process inspections of the power, cooling, and security at colocation facilities and to physically visit them reasonably frequently to attend to the IT equipment. This may provide the customer with a reasonable assurance of competent operation.

A common perception is that colocation is a much shorter financial commitment than owning or leasing data center capacity. In reality, many of the contracts for colocation are of quite long duration and when coupled with the time taken to establish a presence in the colo facility, install and connect network equipment, and then install the servers, storage, and service platforms, the overall financial commitment is of a similar length.

Many colocation facilities suffer from the service monoculture issue and are of high capital cost to meet the expectations of "enterprise colo" customers as well as being located in areas of high real estate or energy cost for customer convenience. These issues tend to cause the cost base of colocation to be high when compared with many cloud service providers.

6.5.4 Cloud Capacity

The major advantages of cloud capacity are the short commitment capability, sometimes as short as a few hours, relatively low unit cost, and the frequent integration of cloud services with other cloud services. Smart cloud operators build their data centers to minimal capital cost in cheap locations and negotiate for cheap energy. This allows them to operate at a very low basic unit cost, sometimes delivering complete managed services for a cost comparable to colocating your own equipment in traditional colo.

One of the most commonly discussed downsides of cloud is the issue of which jurisdiction your data is in and whether you are meeting legal requirements for data retention or privacy laws.

The less obvious downside of cloud is that, due to the price pressures, cloud facilities are built to low cost, and availability is generally provided at the software or network layer rather than spending money on a resilient data center infrastructure. While this concept is valid, the practical reality is that cloud platforms also fail, and when they do, thanks to the high levels of complexity, it tends to be due to human error, possibly combined with an external or hardware event. Failures due to operator misconfiguration or software problems are common and well reported.

The issue for the organization relying on the cloud when their provider has an incident is that they have absolutely no input to or control over the order in which services are restored.

FURTHER READING

Cooling analysis white paper (prepared for the EU CoC). Available at http://www.romonet.com/sites/default/files/document/Manchester%20to%20Madrid%20-%20chillers%20not%20needed%20-%20a%20summary%20of%20cooling%20analysis%20v1.1.pdf. Accessed on May 22, 2014.

With supporting detailed content

- http://dcsg.bcs.org/data-centre-cooling-analysis-world-it-environmental-range-analysis. Accessed on May 22, 2014.
- http://dcsg.bcs.org/data-centre-cooling-analysis-european-it-environmental-range-analysis. Accessed on May 22, 2014.
- http://dcsg.bcs.org/data-centre-cooling-analysis-analysis-data-centre-cooling-energy-efficiency. Accessed on May 22, 2014.

Drury C. *Management and Cost Accounting*. 7th Rev ed. Hampshire: Cengage Learning; 2007.

Energy re-use metrics paper. Available at http://dcsg.bcs.org/energy-re-use-metrics. Accessed on May 22, 2014.

EU, Data center code of conduct. Available at http://iet.jrc.ec.europa.eu/energyefficiency/ict-codes-conduct/data-centres-energy-efficiency. Accessed on May 22, 2014.

7

OVERVIEW OF DATA CENTERS IN CHINA

ZHE LIU, JINGYI HU, HONGRU SONG, YUTAO YANG, AND HAIBO LI
China Electronics Standardization Institute, Beijing, China

7.1 INTRODUCTION

7.1.1 Background

Along with the rapid development of information technologies (IT) and the increasing dependence on IT and products, the informationization (i.e., information based) development in China has been experiencing a continuous boom.

First, Internet popularizing rate is growing rapidly. According to the *34th Statistic Report on Internet Development in* China [1] published by China Internet Network Information Center (CNNIC), by the end of 2013, China' netizen (i.e., a person actively involved in online community) population reached 513 million, with an increase of 618 million year on year, and an Internet popularizing rate of 45.8%; the number of websites had stood at 3.2 million, up 19.4% compared with that at the end of 2012. Take Internet entertainment application, for example, at the end of 2013; the number of online video users had increased by 15.2% over the previous year to 428 million, with the utilization ratio up by 69.3%—online video watching has become the fifth Internet feature after instant messaging, online news, search engine, online music, microblog/personal space. Furthermore, the application of e-business, online music, and video for mobile devices had also increased sharply.

Second, the informationization level of enterprises has been growing with full speed. The investigation results released by the People's Posts and Telecommunications News [2] showed that China's 51.5% large-scale enterprises have applied informationization (i.e., information database) on 80% of the operation work, 25% of which realized 100% informationization; and 68.3% enterprises realized 60% informationization. Among such enterprises, those engaging in finance, telecommunication, and the Internet showed a high level in applying informationization. A total of 46.62% industrial manufacturing enterprises have realized online purchasing, with the online purchasing volume accounting for 25.03% of the total; 45.66% enterprises have realized online sales, with the online sales volume accounting for 29.99% of the total. Since only 10% small- and medium-sized enterprises (SMEs), which take up 99% of total enterprises in number and contribute 60% in social output value, have realized informationization management so far, it is believed that with the policy supports in informationization development, in the coming 5–10 years, the demands on informationization will increase obviously.

Third, the e-government has been booming. The number of registered domain names ending with gov.cn has exceeded over 5% of China's total, and almost 90% prefectural and municipal governments have set online offices. The platform construction maturity of China's e-government was low, and e-government is gaining increasing attention.

With the increasing demands on informationization, the key information data volume of China has been increasing at a rate of over 50%, leading to the optimization and integration of information resources such as hosts, data backup devices, data storage devices, high availability systems, data safety systems, database systems and infrastructure platforms, which have pushed forward the rapid development of data centers.

7.1.2 Development History

The development history of China's data centers falls into four stages according to relevant technologies, scales, and applications.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

Stage I (before 1980s): During this time period, the data center was actually an environment built to meet the operation needs of single large-scale computers, and the environment was called "computer field." Since the "data center" was built for the operation of a single computer, the hardware and software technologies applied were few in variety and large in size. The devices were cooled by air supplied through ducts, and no equipment and measures for keeping constant temperature and humidity were available. Limited by technologies at that time, the voltage regulator, instead of Uninterruptible Power Supply (UPS), was employed to control voltage, and manual 24-h monitoring, instead of electronic monitoring, was used for management. As a result, power failures and obvious changes in temperature and humidity would cause interruption of computer operation. Therefore, computer rooms during this time period were not real data centers and were mainly used for scientific research.

Stage II (between 1980s and 1990s): During this period of time, large-scale computers had been gradually replaced by microcomputers, the local area network (LAN) developed rapidly, and more and more researchers started to cast their eyes on technologies relating to computer fields. In 1982, the monograph *Computer Fields-Related Technologies* was published, specifying the requirements of temperature, humidity, and power source quality; in 1986, the *Computer Room Design Standard* was issued. At that time, constant temperature and humidity devices, air conditioners, UPS power sources had been adopted, and electronic monitoring on several parameter indexes had been conducted, which enabled a stable computer operation for dozens of hours and even several months. However, limited by transmission speed, the then data centers mainly served the LAN.

Stage III (from 1990s to 2000): Benefiting from the increasingly mature Internet technologies, informationization was more and more widely applied in the society, rack-mounted servers sprang up in large amounts, and technologies relating to design, hardware, and software of data center developed rapidly. In 1993, relevant standards such as the *Code for Design of Electronic Information System Room* were released, which means that data center standardization has caused more and more concerns. During this stage, the application of data centers had expanded to wider fields such as national defense, scientific research, and office work.

Stage IV (from 2000 to present): During this stage, the sharp growth of the worldwide web pushed forward the rapid Internet popularization in various aspects in China, which raised higher demands on the construction of data centers. The standardization and modularization-oriented development of data centers enabled flexible extension, maintenance, and management. As virtualization technologies are getting mature, the development of data centers showed a trend of diversification: special industries engaging in the design, construction, supporting devices supply, and services relating to data centers have formed. Today's data centers are able to serve various aspects of our society.

At present, the development of China's data centers is on the fast lane. The market size relating to data center application stood at RMB 83.9 billion (US$13.1 billion) in 2010, and the compound growth rate of which in the recent 4 years reached 18.6% (Fig. 7.1). The data center

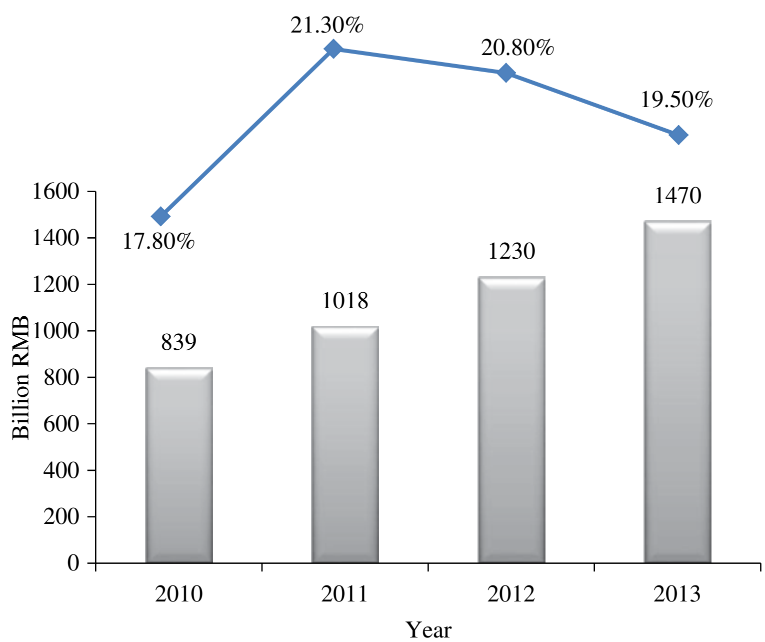


FIGURE 7.1 Market value of China's data center application.

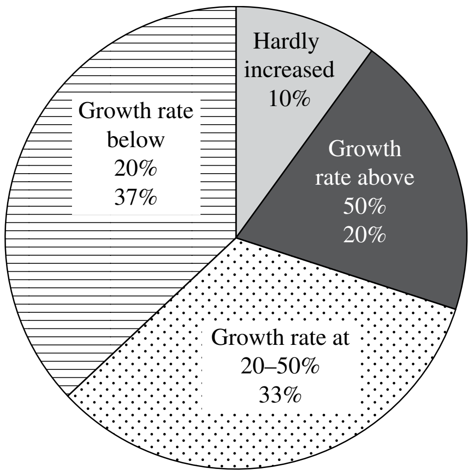


FIGURE 7.2 Distribution of server number increase of China's Internet data center in 2011.

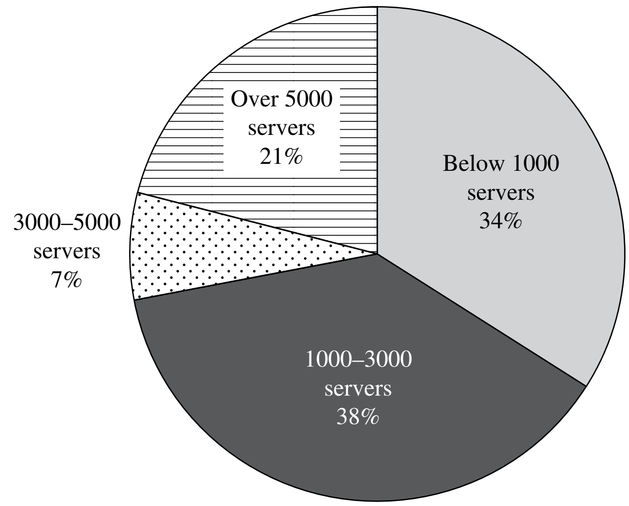


FIGURE 7.3 Scale distribution of Internet data centers in 2011.

development of China has begun to take shape. Market value of china's data center application (Fig. 7.1) shows that the number of various [3] data centers and computer centers had reached 519,990, and over 90% of which were small-sized data centers with an area of less than 400 m^2, which means that small-sized data centers still dominated the data center market. According to the prediction of IDC, the number of China's data centers will reach 540,777 in 2012, with an annual compound growth rate of 1.3%. Accordingly, Internet data centers experienced the fastest development: in 2011, the market size of China's Internet data centers reached RMB 17.08 billion (US$2.7 billion), up 67.1% year on year; 33% data center companies, the second largest proportion among all data center companies, experienced a growth of 20–50% in server number (Fig. 7.2); and among all Internet data center service providers being investigated, 21% had over 5000 servers. So far, most IDC service providers are small- and medium-sized ones, but they have been growing as a very strong force (Fig. 7.3).

7.2 POLICIES, LAWS, REGULATIONS, AND STANDARDS

7.2.1 Organization Structure of Data Center Industry

The standardization of Data Centers is governed by Standardization Administration of China (SAC), Ministry of Industry and Information Technologies (MIIT), Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD), and other related departments, and there are mainly four technical committees (Fig. 7.4).

The Structure of Governmental Regulatory Agency is shown in Figure 7.5. China Electronics Standardization

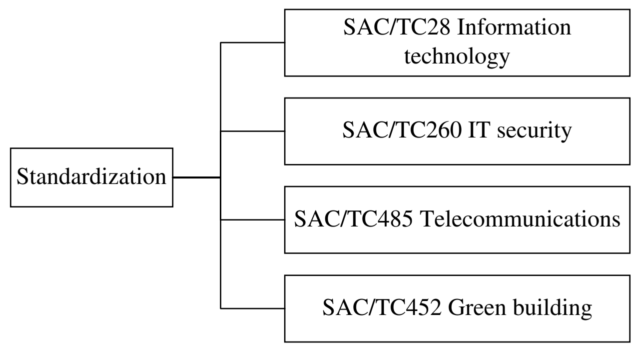


FIGURE 7.4 Structure of standardization organizations.

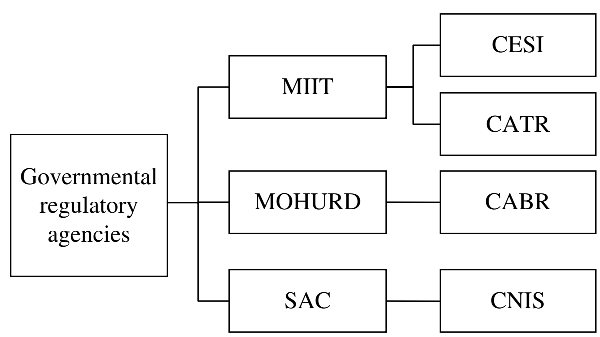


FIGURE 7.5 Structure of governmental regulatory agencies.

Institute (CESI) is a leading force in standard research on information technology, and is the mirror of ISO/IEC JTC1, and Secretariats of SAC/TC 28 and SAC/TC 260 and the affiliation of China Electronics Standardization Association. SAC has approved the establishment of SAC/TC 28 SC39 Sustainability for and by Information Technology, responsible for green data center standardization, as the mirror of ISO/IEC JTC1 SC39. China Academy of Telecommunication Research

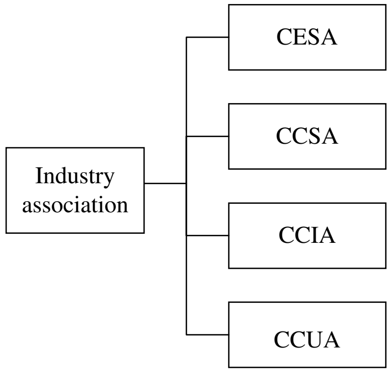


FIGURE 7.6 Structure of industry associations.

(CATR) is authorized telecommunication technology research institute, the achievement of which includes TD-SCDMA. China Academy of Building Research (CABR) is a comprehensive institution of building technology including green building. China National Institute of Standardization (CNIS) is a fundamental institution in Standard research.

Industry Associations, as shown in Figure 7.6, mainly include China Electronics Standardization Association (CESA), China Communications Standards Institute (CCSA), China Computer Industry Association (CCIA), and China Computer User Association (CCUA). Each of these associations has departments related to data center.

7.2.2 Policies

7.2.2.1 National Policies Development plans and key regulations released by China State Council are compiled for the purpose of defining the development directions, working focuses, and implementation schemes of various industries. To ensure the implement of the plans and key regulations, most departments in charge of the industries will develop corresponding supporting policies and promotion schemes, and advance the planned work focus via incentive measures such as fiscal and tax preferential policies during daily work; and governments of all levels will make medium- and long-term development plans and guides, develop corresponding supporting policies and implementation schemes, and provide policy supports to planned work focus during daily work.

In the *Decision of China State Council on Accelerating the Cultivation and Development of Strategic Emerging Industries* [4], it was clearly pointed out that efforts would be made to "accelerate the construction of broadband-based, ubiquitous, synergetic and safe information network; push forward the R&D and industrialization of the new generation's mobile communication, the next generation's internet core devices and intelligent terminals; speed up the integration of the three networks—the telecommunication network, broadcast and television network and internet; and promote the R&D and exemplary application of the internet of things and cloud computing." Data centers are the important and core infrastructures for the construction of next-generation Internet, combination of three networks, Internet of things, and cloud computing.

The *Outline of the 12th Five-Year National Economic Development Plan* [5] *(2011–2015)* clearly indicated the goals of "accelerating the construction of broadband-based, ubiquitous, synergetic and safe information network and promoting the information technology based development in various social and economic fields" and "reinforcing the construction of the cloud computing server platforms."

It is specified in the *Industrial Transformation and Upgrading Plan 2011—2015* [7] that efforts would be made to "build and improve the service system for realizing the information technology-based enterprise development; strengthen the service ability; implement the information technology-based industrial development project; push forward the close combination of the R&D of information-based technologies and industry application; develop a groups information service platforms facing various industries; cultivate a group of national level information-based technology promotion center, build a groups of national-level data centers serving for key industries; and establish a group of IT application demonstration enterprises."

In the "Complete Machine Value Chain Promotion Project" put forward in the *12th Five-Year Development Plan for Electronic Information Manufacturing Industries* [6], it was clearly pointed out that efforts would be made to "support the R&D and industrialization of key products such as high-end servers and network storage systems, and push forward the R&D and industrialization of key equipment employed in green and intelligent data centers and various end products."

In the *12th Five-Year Development Plan for Communication Industries* [8], it was clearly indicated that efforts would be made to "deepen the all-round application of internet technologies in various economic and social fields; make breakthroughs in mobile internet technology based business system innovation; accelerate the commercialization of the cloud computing; and realize the integrated arrangement of new type application platforms such as cloud computing centers, green data centers, CDN," "popularize the all-round application of technologies relating to energy saving and environmental protection; reduce the gross comprehensive energy consumption ratio per communication business by 10% in 2015 compared with that in 2010; and lower the PUE value of newly built cloud computing data centers to below 1.5," and "push forward the transformation of traditional internet data centers to cloud computing technology based data centers; and construct green large-scale internet data centers meeting the national requirements on energy saving and environmental protection."

In the *12th Five-Year Development Plan for Internet Industries* [9], it was clearly indicated that "factors such as network structures, market demands, supporting environment, geographical energies, and information safety will be comprehensively considered, and measures will be taken to strengthen technical standards and industrial policy guiding, optimize the construction layout of large-scale data centers, and guarantee the high-speed connection between large-scale data centers."

In the *Several Opinions of the State Council on Further Promoting the Development of Small and Medium-sized Enterprises* (SMEs) [10], it was clearly pointed out that measures would be taken to "speed up the informationization development of SMEs; keep implementing the SME informationization promotion project; guide SMEs to improve their R&D, management, manufacturing and service level and enhance their marketing and after-sale service ability by employing information technologies; and encourage information technology enterprises to develop and build industry application platforms to provide social services such as software and hardware tools, project outsourcing, industrial designs to SMEs." The *12th Five-Year Growth Plan for SMEs* put forward the following goals: "the proportion of SMEs applying information technologies in R&D, management and production control will reach 45%; the proportion of SMEs applying e-commerce in purchase, sales and other business will be up to 40%; and the information service platform for SMEs will be basically built." For example, data portals built by China's large network service platforms represented by Alibaba (an Internet company "to buy or sell online anywhere in the world") have built a free, open, and shared Internet data platform to provide convenient, diversified, and professional integrated data services for third parties, including SMEs and e-commerce practitioners by analyzing and mining the data relating to the e-commerce behaviors such as search, query, and transaction of 45 million SMEs.

The construction of data centers plays a fundamental role in carrying out the priorities in the policies mentioned already. The Chinese government is promoting the support for financial and preferential policies to accelerate the construction of data centers to achieve the goals already mentioned.

7.2.2.2 Local Development Planning With the advance of the work concerning the twelfth 5-year plans and strategic emerging of industries mentioned earlier, in 2011, the Ministry of Industry and Information Technology (MIIT), together with National Development and Reform Committee (NDRC), issued *Notice on Pilot Demonstration for Innovative Development of Cloud Computing Service*, and decided that Beijing, Shanghai, Shenzhen, Hangzhou, Wuxi, etc. started to carry out the pilot work (Fig. 7.7). In addition, under the circumstances that the state devotes major efforts to promote cloud computing demonstration project in the five cities, other areas also begin to make cloud computing development planning with support from the market and local government (Fig. 7.4).

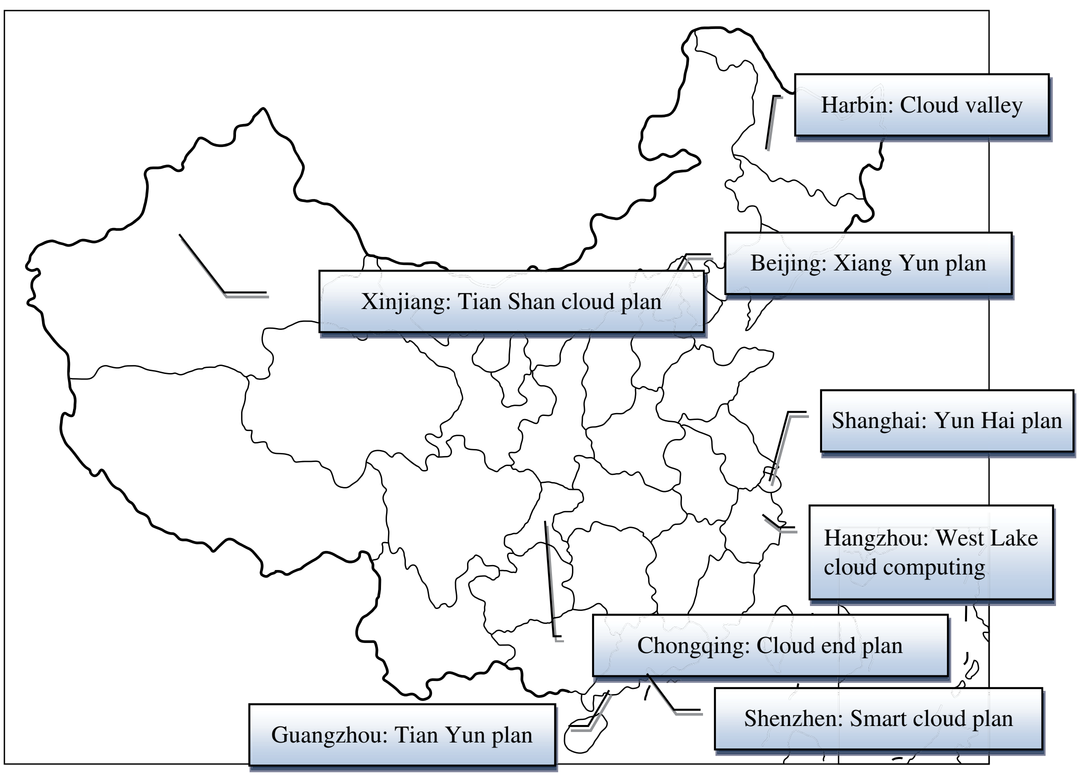


FIGURE 7.7 Cloud computing construction plans around China.

In 2010, Beijing Municipal Commission of Economy and Information Technology released *Action Plan for Beijing "Xiang Yun Cloud Project,"* which proposes the plan of making the three typical services of "cloud computing"—infrastructure service, platform service, and software service—form a RMB 50 billion (US$7.8 billion) industry scale in 2015 and thus driving the cloud computing industry chain to produce an output value of RMB 200 billion (US$31.3 billion). It also plans to strive to build a data center providing a 20,000-m^2 high-grade computer room and support over 100,000 servers to become a world-class cloud computing industry base. The government will provide all-round tracking service on major projects, dispatch and coordinate at regular time, give major supports to the excellent projects, and support project construction in such ways as capital fund injection, loan with discounted interest, and investment grant. In September 2011, China Cloud Industry Park was officially launched. Located in Yizhuang, Beijing. China Cloud Industry Park is planned to cover an area of 3 km^2 in the preliminary planning, and the reserved building area is 2 km^2. The projects of the first batch of cloud industry parks include KDDI data center project and Beijing Telecommunication data center project and so on, with a total investment of RMB 26.1 billion (US$4.1 billion).

In 2010, Shanghai launched "Yun Hai Plan," and is planning to establish "Yunhai Data Center," whose area is 320 mu (or 213,333 m^2) in the Phase I project and that can support hundreds of thousands of servers. According to the Yun Hai Plan, within 3 years from 2010 to 2012, Shanghai will be built into an Asia-Pacific level data center and will increase the operating income of the information service industry to over RMB 100 billion (US$15.6 billion). The Phase I project comprises three major areas—data center industry area, business operation area, and expansion area, of which, the data center industry area will mainly develop date center, disaster recovery backup center, cloud computing center, e-commerce, financial back-office, logistics back-office, accounting settlement center, and other industries that mainly focus on domestic users; the business operation area will mainly attract such industries as data value-added service industry, animation and online gaming, software testing and outsourcing, R&D and design industry, and so on, and provide business supporting services. The construction of Shanghai Yunhai Data Center will continue until 2020 and is a long-term investment project in the cloud computing field.

Harbin, the capital and largest city of Heilongjiang Province in Northeast China, issued *Development Planning for Cloud Computing Industry in Harbin (2011–2015)* in 2011. Though it is not listed as the first group of national pilots, Harbin has attracted much attention from the industry by advantages of its cool weather and location that are very suitable for the construction of data centers. First, Harbin is a provincial capital with the highest latitude in China, with its northern latitude of between 44° and 46°. The annual mean temperature is +4.25°C (39.7°F). The cold air there can be used for refrigeration. Second, it is not on a high seismic zone area, its geologic structure is stable, and it is free from threats from such natural disasters as tsunami, mountain torrents, and debris flow; third, Harbin is abundant in electric power resources, with 40–70 billion kilowatt hours being underused through the year; fourth, Harbin has land optical cables connecting to Russia and is currently applying for international port bureau and will then be rich in bandwidth resources; last but not least, Harbin's underground water temperature and air quality are very suitable for the requirements of a large free cooling data center. By 2015, Harbin will build a data center that can hold two million servers. Further, China Mobile will start the national cloud computing project with a total investment of RMB 1.5 billion (US$234 million) in this year and Harbin's "China Cloud Valley" will become an important data center in China.

In 2010, the construction of Chongqing, one of the five national cities in PRC, "Liangjiang International Cloud Computing Center" was officially started. During the twelfth 5-year period, it will be built into a data center base covering an area of 10 km^2 and holding 3 million servers, forming a large cloud computing center inland. The total building area is 2.07 million square meters and the total planned investment is RMB 40 billion (US$6.2 billion). Also in the core area, about 3-km^2 international-level data center is planned. By the end of this year, the infrastructure of Chongqing Liangjiang New Area will be capable of supporting 1 million servers. See Table 7.1 for the summary on China's data centers proposed to be completed in the near future.

TABLE 7.1 Summary of China's Data Centers Proposed to be Completed in the Near Future

Serial no.	Location	Project name	Estimated area (unit: m^2)	Estimated data center scale (unit: 10,000 servers)	Investment (unit: billion RMB/US$ billions)
1	Beijing	Xiang Yun plan	Building area: 20,000	10	26.1/4.1
2	Shanghai	Yun Hai plan	Floor space: 213,333 (Phase I project)	10–90	—
3	Chongqing	Cloud end plan	Floor space: 10,000,000	300	40/6.2
4	Harbin	Cloud valley plan	TBD	200	—
5	Xinjiang	Tian Shan cloud computing	TBD	25	—

7.3 STANDARDS

Currently, there hasn't been a unified coordination mechanism, so the standards for the data center mainly follows the standard systems for electronic information computer rooms (computer fields), key infrastructures, and building construction. In this section, national and industry standards with"/T" in the standard number are recommended to conform to, while the ones without "/T" are compulsory and must be conformed to.

7.3.1 Comprehensive Standards

GB 50174–2008 *Code for Design of Electronic Information System Room*[1] (GB is short for Guo Biao, which means national standard), which was first issued and implemented in 1993 and revised in 2008, is a basic compulsory national standard for data center infrastructure. It includes the location of computer room, equipment layout, environmental requirements, building structure, air conditioning, electrical technology, electromagnetic shielding, arrangement of wires in computer rooms, computer room monitoring and safety precautions, water supply and drainage, firefighting, etc. This standard classifies the reliability of the data center into highest (Level A), medium (Level B), and basic level (Level C). The standard in China's data centers is as important as that of ANSI/TIA 942-2005 in the United States. What's more, GB 50462-2008 *Code for Construction and Acceptance of Electronic Information System Room*, an acceptance standard that supports GB50174, is formulated to strengthen project quality management, unify construction and acceptance requirements, and ensure project quality.

GB/T 2887-2000 *General Specification for Computer Field*[2] (GB/T is short for Guo Biao Tui, which means recommended national standard) is a recommended national standard, which regulates the requirements on the composition, area, layout, noise, earthing, power supply and safety of the electronic computer filed, and specifies the test methods and acceptance specifications. It is highly operable and is applicable to all kinds of electronic computer system fields and can be also used as reference for electronic equipment system fields.

GB/T 9361-2011 *Safety Requirements for Computer Field*[3] is a recommended national standard, which puts forward the safety requirements that mainly concern computer fields. It includes safety requirements on site selection, seismic resistance, firefighting, water resistance, static resistance, lightning protection, anti-noise, rat-proof, anti-intrusion, decoration, power distribution and air conditioning, etc.

YD/T 5003-2005 *Specifications of Engineering Design for Telecommunication Private Premise*[4] (YD/T is short for You Dian Tui, which means recommended telecommunication industry standard) is a recommended standard of the post and telecommunications industry. It includes requirements on site selection, fire-resistance ratings of buildings, general layout, architectural design, structural design, heating, air conditioning, ventilation, water supply, water drainage and fire control design, electrical design, lightning protection, earthing, etc.

In addition, the standards such as SJ/T 10796-2001 *General Specification for Raised Access Floors for Electrostatic Protection* (SJ/T is short for Si Ji Tui, which means recommended electronic information industry standard) are also important standards concerning the construction of data centers.

JR/T 0011-2004 *Code for Centralized Data Center of Banks*[5] (JR/T is short for Jin Rong Tui, which means recommended financial standard) is a recommended standard in the financial industry of China. Focusing on the banking industry, it regulates the daily operation, system maintenance, application maintenance, safety management, data and file management, emergency treatment, and service quality management requirements. It integrates the generic specifications of data center with the service features of the banking industry, which makes it a standard more focused on the application layer.

Moreover, some enterprises and public institutions have formulated their own data center standards based on their own business demand according to national, industrial, and foreign standards. For example, China Telecom Corporation formulated DXJS1006-2006 *Code for Design of Power Supply and Air Conditioning Environment of Data Center of China Telecom* and DXJS1007-2005 *Code for Acceptance of Power Supply and Air Conditioning Environment of Data Center of China Telecom* (DXJS is short for China Telecom Corporation), which, based on computer room equipment power, equipment cabinet structure, and computer room equipment layout of the data center, classifies data center computer rooms into different classes, regulates the safety operation requirements and configuration principles for the power system of the computer rooms of different classes as well as the safety operation requirements, the configuration principles of refrigerating output and air output, and the layout principle of the airflow organization of the air conditioning system.

[1] GB 50174-2008 Code for Design of Electronic Information System Room, Ministry of Housing and Urban-Rural Development (MOHURD) and General Administration of Quality Supervision, Inspection and Quarantine (AQSIQ).

[2] GB/T 2887-2000 General Specification for Computer Field, MIIT and AQSIQ.

[3] GB/T 9361-2011 Safety Requirements for Computer Field, MIIT and AQSIQ.

[4] YD/T 5003-2005 Specifications of Engineering Design for Telecommunication Private Premise, MIIT and AQSIQ.

[5] JR/T 0011-2004 Code for Centralized Data Center of Banks, People's Bank of China and AQSIQ.

7.3.2 Standards for IT Equipment

Standards regarding servers include GB/T 21028-2007 *Information Security Technology—Security Techniques Requirement for Server*, GB/T 25063-2010 *Information Security Technology—Testing and Evaluation Requirement for Server Security*, YD/T 1659-2007 *Testing Methods of Security for Broadband Network Access Server*, and YD/T 1658-2007 *Technical Specification for Broadband Network Access Server Security*. GB/T 9813.4 *Generic Specification for Microcomputers Part 4: Server* is currently under the process of formulation.

The standards regarding network equipment include GB/T 21050-2007 Information security techniques—Security requirements for network switch (EAL3), YD/T 2042-2009 IPv6 Network Equipment Security Requirements—Ethernet Switch with Routing Capability, YD/T 2043-2009 IPv6 Network Equipment Security Testing Methods—Ethernet Switch with Routing Capability, YD/T 1941-2009 Testing Methods for Ethernet Switch Equipment with Content Exchange Capability, YD/T 1917-2009 IPv6 Network Equipment Testing Methods—Ethernet Switch with IPv6 Routing Capability, YD/T 1691-2007 Technical Requirements for Ethernet Switch Equipment with Content Exchange Capability, YD/T 1627-2007 Security Requirements of Ethernet Switch Equipment, etc.

With regard to energy conservation, the national standards *Requirements on Energy Consumption of Server* and *Requirements and Test Methods for Energy Efficiency Limit Value of Server* have been officially set up and are currently under the consultation procedures. GB/T 9813.4 *Generic specification for microcomputers Part 4: Server*, which is currently under the process of formulation, also includes the requirements on the energy efficiency of server. China also issued GB/T 26262-2010 *Guide for Classification of Telecommunication Equipment Energy Efficiency*, which gives guidance on the energy conservation evaluation of telecommunication products such as network switch. The Ministry of Environmental Protection issued HJ 2507-2011 *Technical Requirement for Environmental Labeling Products Servers* this year, which includes energy efficiency requirements of computing servers and storage servers.

7.3.3 Standards for Buildings

In addition to the special comprehensive standards of data centers, the design, construction, and operation of buildings in data centers shall also satisfy some general standards of the construction industry.

The basic standards include the following:

- GB 50015-2003 *Code for Design of Building Water Supply and Drainage*
- JGJ/T 16-2008 *Code for Electrical Design of Civil Buildings*

The standards for building security include the following:

- GB 50016-2006 *Code of Design on Building Fire Protection and Prevention*
- GB 50370-2005 *Code for Design of Gas Fire Extinguishing Systems*
- GB 50045-2005 *Code for Fire Protection Design of Tall Civil Buildings*
- GB 50348-2004 *Technical Code for Engineering of Security & Protection System*
- GB 50343-2004 *Technical Code for Protection against Lightning of Building Electronic Information System*
- GB 50084-2005 *Code of Design for Sprinkler Systems*
- GB 50166-1992 *Code for Installation and Acceptance of Fire Alarm System*
- GB 50034-2004 *Standard for Lighting Design of Buildings*

With the energy conservation and emission reduction being increasingly accepted by the public, more and more attention has been paid to the concepts of green buildings and green data centers. In China, the green building standard system has been preliminarily established, which provides requirements and reference for the development of green data centers. On the basis of summarizing the practical experience in green buildings and taking examples of international green building evaluation system, China issued GB/T 50378-2006 *Evaluation Standard for Green Building*,[6] which provides a comprehensive evaluation system focusing on multiple objectives and multi-layer green building for the purpose of comprehensive evaluation on such aspects as site selection, materials, energy saving, water saving, and operation management. This Standard focuses on energy saving and control during design. In addition, China also issued *Green Building Rating Labeling Management Method* and *Technical Rules for Evaluation of Green Buildings* as basis of design and evaluation. After being examined by experts and China Green Building Council, buildings will obtain "Green Building Rating Label" rated as 1 Star, 2 Stars and 3 Stars, in which 3 Stars is the highest level. In 2008, the *Technical Guideline for Building Energy Evaluation & Labeling* was on trial. By absorbing achievements and experience of international building energy labeling, according to China's current building energy design standard, and the current conditions and features of building energy work in China, the Guideline, which is applicable to the evaluation and labeling for newly built residential and public buildings and existing buildings that have undergone energy saving transformation, emphasizes the evaluation system for actual

[6] GB/T 50378-2006 Evaluation Standard for Green Building, MOHURD and AQSIQ.

energy consumption of buildings and energy effectiveness control. In 2011, GB/T 50668-2011 *Evaluation Standard for Energy-Saving Building* was also officially issued. The standards, codes, and technical guidelines like JGJ/T177-2009 *Energy Efficiency Test Standard for Public Buildings* and GB50411-2007 *Code for Acceptance of Energy Efficient Building Construction* serve as basis for acceptance and operation management of energy-saving work of buildings. In addition, in terms of energy-saving design of buildings, China has formulated the design standards for residential and public buildings covering three climate regions nationwide, including GB50189-2005 *Design Standard for Energy Efficiency of Public Building*, JGJ26-95 *Energy Conservation Design Standard for New Heating Residential Buildings*, *Design Standard for Energy Efficiency of Residential Buildings in Hot Summer and Cold Winter Zone* (JGJ134-2001, J116-2001), and *Design Standard for Energy Efficiency of Residential Buildings in Hot Summer and Warm Winter Zone* (JGJ75-2003, J275-2003). Furthermore, standards including GB/T 50314-2006 *Standard for Design of Intelligent Building and* GB 50339-2003 *Code for Acceptance of Quality of Intelligent Building Systems* can be used as references for the construction of data centers.

7.3.4 Standards for Power Supply

The basic standards for power supply mainly include GB 50052-95 *Code for Design of Power Supply and Distribution System*, GB 50054-95 *Code for Design of Low Voltage Electrical Installations*, GB 50060-2008 *Design Code for High Voltage Electrical Installation (3–110 kV)* and GB/T 12325-2008 *Power Quality-Admissible Deviation of Supply Voltage*; standards for wiring include GB 50311-2007 *Code for Engineering Design of Generic Cabling System*, GB 50312-2007 *Code for Engineering Acceptance of Generic Cabling System*, and GB/T 50312-2000 *Code for Engineering Acceptance of Generic Cabling System for Building and Campus*; standards for UPS include GB 7260-2003 *UPS Equipment and YD/T 1095-2008 Uninterruptible Power Systems for Communications*; standards for switches include GB/T 14048.11-2008 *Low-voltage Switchgear and Control Gear—Part 6–1: Multiple Function Equipment—Automatic Transfer Switching Equipment*; standards for emergency power supply include YD/T 799-2002 *Valve-Regulated Lead Acid Battery for Telecommunications* and YD/T 502-2000 *Technical Requirements of Diesel Generator Sets for Telecommunication.*

In addition, the standard GB/T 16664-1996 *Monitoring and Testing Method for Energy Saving of Power Supply Distribution System of Industrial Enterprise* also provides reference for monitoring loss of the enterprise's daily power supply system and guidance for improving power usage efficiency.

7.3.5 Standards for Air Conditioning Equipment

At present, the basic standards for air conditioning equipment used in data centers include GB/T 19413-2010 *Unitary Air-conditioners for Computer and Data Processing Room*. The construction of data centers shall also conform to other codes, including GB 50019-2003 *Code for Design of Heating Ventilation and Air Conditioning*, GB 50243-2002 *Code of Acceptance for Construction Quality of Ventilation and Air Conditioning Works*, and GB50365-2005 *Code for Operation and Management of Central Air Conditioning System*. Besides, GB/T 26759-2011 *Technical Specification for Energy-saving Control Device for Water System of Central Air conditioning* is China's first product technical standard of energy-saving and control for central air conditioning, which regulates the energy-saving control technology of central air conditioning water system. As a cold (heat) transmission and distribution system in the central air conditioning system with water (including saline and glycol) as a medium, the central air conditioning water system generally includes cold water (hot water) system and cooling water system. The energy-saving control device of the central air conditioning water system realizes the optimizing control through operation of the central air conditioning water system so as to improve the energy usage efficiency of air conditioning system. This technical specification is also a common standard for air conditioning equipment in data centers.

For the energy saving of air conditioning equipment, only energy-efficient evaluation standards for air conditioner are formulated at present. China has issued energy-efficient evaluation standards for common air conditioners like GB 21455-2008 *Minimum Allowable Values of the Energy Efficiency and Energy Efficiency Grades for Variable Speed Room Air Conditioners* and GB 12021.3-2010 *Minimum Allowable Values of the Energy Efficiency and Energy Efficiency Grades for Room Air Conditioners*. In addition, YD/T 2166-2010 *Adaptive Monitor System for Precision Air Conditioner for Telecommunication Stations/Sites* that regulates the control system of precision air conditioning can be used as reference for the formulation of energy-saving evaluation methods for precision air conditioning in data centers.

7.4 DEVELOPMENT STATUS OF CHINA'S DATA CENTERS

7.4.1 Development Status of Key Equipment

Recently, data center-related equipment industries have been developing rapidly and have made great progress in producing modular data centers, servers, UPS and air conditioners, becoming an important force in international market.

7.4.1.1 Modular Data Center With the combining progress of data centers and virtualization technologies, modular data centers with high power density, low energy consumption, flexible input, and easy deployment have been developed. Currently, global leading Internet enterprises and telecom operators have begun to launch pilot projects of new generation's modular data centers with less environmental effect step by step. Chinese electronic information equipment manufacturers such as ZTE, Huawei, and Inspur have finished the research and development of modular data center and realized commercialization, which are expected to be the main stream in the coming 5 years.

In April 2011, Inspur first released "Yunhai Container" products adopting modular and standardized design and integrating all modules such as power supply, refrigeration, cooling, and IT into one standard container. The container products are of two types, namely, 20 ft split and 40 ft, among which the 20 ft split can contain 7680 computing cores, being able to provide one hundred trillion times' the computing capacity. The design load reaches 30 tons, meeting the bearing demands under various configurations. Meanwhile, the containers' internal units have been subject to reinforcing and seismic design, and the equipment to double elastic fixation at both bottom and top, thereby reducing the impact of external disturbance on work in transportation and use process. What's more, the new containers are able to resist an earthquake of magnitude 7 and above. The power conversion efficiency has been upgraded from less than 80% to more than 93% through centralized power supply, centralized heat dissipation, and centralized management; there is no fan inside the server node while centralized fan wall heat dissipation method is adopted, reducing the heat dissipation and power consumption from 15 to 37% to less than 12% with a PUE value of only 1.22.

ZTE launched "Green Cloud Container" solution in October 2011, which includes indoor and outdoor types, realizing a higher computing density and lower energy consumption by modular design. Indoor data center module with a story height of 3 m can support 55U cabinet, while the outdoor 40 ft container can accommodate 2880 servers or 26 petabytes storage. About 30–70% electric power cost can be saved if analyzed from the perspective of cost: for example, as for a data center with 1400U rack space, 34.85% Capex, 44.88% Opex, 37.92% TCO can be saved for ZTE modular data center when compared with traditional data centers; as to covering area, the covering area under this solution is only one-fourth of that of the traditional data center; as to delivery and expandability, ZTE modular data center is of modular design with standardized components, which can provide flexible combination of expanding on demand, rapidly deploy by "piling up" and facilitate the customers to deploy by stages as required. This solution can be rapidly delivered within 90 days, and it can be installed and commissioning can be finished within 1 week, with a greatly improved efficiency.

In November, the same year, Huawei Technologies released modular data center solution, which includes container-type modular data center and building-type modular data center. The container-type IDS1000 applies to the outdoors scene without a computer room, the computer room site selection problem is solved. With construction period reduced to 2 months from 6 months, its energy consumption is only 50% of that of the traditional computer room. Therefore, it can be applied to the extreme environment in hot and arid regions in desert and polar cold regions; and, especially, it shows its unique advantages in emergency rescue, mass gathering, military, exploration, etc. The building-type modular computer IDS2000 adopts the advanced refrigeration technologies like separation between hot and cool air ducts, precise air supply, and outdoor cold source, to make sure the PUE <1.2, and so it can be applied to the modularized construction and expansion of large and medium-size data centers. With these products, the customers can save more than 30% Total Costs of Ownership.

7.4.1.2 Server China is relatively late in the development of servers. The first China's own server came into being in the early 1990s, a decade later than world early servers. Due to lack of server core technology, China's server market has seen restrained development for a long time, and the domestic server market has been always monopolized by such international, well-known brands as IBM, Dell, HP, and Sun. Although China's local servers are still lack of absolute advantages, with the development of core technology, this situation has been improved. With great progress in performance, reliability and service, the domestic servers have gradually developed from generic branded products to branded ones and are advantageous in price; they have been gradually accepted by domestic users and account for larger and larger domestic market share, so they are also the key point concerned by purchasing users.

At present, Inspur, Lenovo, Sugon, and Tsinghua Tongfang all have possessed stronger server design and manufacturing ability. Through several years of experience in server field, Inspur has been able to independently research, develop, and manufacture the important components such as chassis, cooling system, motherboard, RAID card and expansion card, and carry Inspur Ruijie Server Management Software with proprietary intellectual property rights. Sugon has played an important role in the development of domestic servers, making several breakthroughs in super computer field and applying the technology and experience of super computer in the design and manufacturing of x86 Server, and also Sugon has developed its own management system software Gridview with friendly interface and various functions. Originating from an enterprise established by the Institute of Computing Technology of Chinese Academy of Sciences, Lenovo Group is one of the largest computer manufacturers in the world, and the servers it developed have a strong competitiveness and are excellent in expansibility, manageability and safety. Meanwhile, the small-scale server producers represented by Chongqing Zhengrui Technology have gradually grown and will become a powerful competent force in the future.

7.4.1.3 Uninterruptible Power Supply At present, the UPS market of China's data centers is quite booming. APC, EKSI, MEG, and other world-famous branded UPS products dominate the market of medium and high-power UPS products (10 kVA above), with a market share reaching 80%, because of their advanced technology. Since the 1990s, some excellent domestic brands have made great achievements and become the driving force in the medium and low-power UPS market through the persistent pursue in technology and the advantage of local production and service. Among UPS manufacturers, there are five with a sales volume of RMB 200 million (US$31.2 M); five with 100–200 million (US$15.6–31.2 M), 15–20 with RMB 50–80 million (US$7.8–12.5 M), about 247 with RMB 25–50 million (US$3.9–7.8 M), more than 1380 with RMB 5–25 million (US$8–39 M), and 300 others.

7.4.1.4 Air Conditioner The air conditioner market of China's data centers is quite active, which is mainly dominated by international manufacturers with a growing market share by domestic products. According to ICT Research, the air conditioner market of China's data center will reach RMB 3.135 billion (US$489 million) in 2012. There is a gradual change to high-power product market. The air conditioning of greater than 30 kW accounts for 70%. In addition, with the development of chilled water air conditioning, over 100 kW products have a growing proportion in market year by year. The international manufacturers still dominate the market, and Liebert, Hiross, and Atlas under Emerson have a total market share of more than 40%. Domestic products have an obvious improvement, and the domestic manufacturers are advantageous in price and local channels. Besides, independent R&D of high-end products have been enhanced, for example, Midea, Gree, and other civil air conditioning brands all have increased R&D inputs and market promotion efforts.

7.4.2 Distribution of Data Centers

7.4.2.1 Industry-Based Distribution China's data centers are mainly distributed in finance and telecom industries with a total market share of greater than 50% [11]. Since 2009, under the background of telecom operator reform and 3G network construction accelerating, with fiercer competition among three major operators (i.e., China Mobile, China Telecom, and China Unicom), accelerated upgrade of telecom core system, explosive growth of Internet audio and video business, and expended application of e-business, the data center market development has been greatly promoted. In the meantime, with the development of finance business and online bank, large financial enterprises are in the construction of disaster recovery centers, traditional data centers have been in an accelerated upgrade, urban commercial banks, rural credit cooperatives, and joint-stock commercial banks are under a rapid development, and IT system construction investment is in a rapid growth, all of which are important factors to drive the data center market development of financial industry.

7.4.2.2 Scale Distribution According to the report *Analysis on Construction Market Conditions of China's Data Centers in 2009* issued by IDC 2010, the total sum of data centers and computer rooms in China reached 519,990 in 2009, in which small data centers with an area of less than 400 m^2 exceed 90% of the total ones, so that small data centers and mini-size computer rooms still are the main form in China's data center market. It is predicted that China's data centers will reach about 540,777 at a compounded annual growth rate of 1.3%.

7.5 ENERGY EFFICIENCY STATUS

With constant promotion of informationization in China, key data volume is increasing at compounded annual average growth rate of 52%. As the major carrying facility of data, the data center has stepped into a phase of rapid development with increasing energy consumption. According to statistics, the total power consumed by China's data centers accounts for 2.4% of the power consumed by the whole society, which is higher than annual energy output generated by the Three Gorges dam, and the area of data centers completed grows at the rate of 15% annually. At the end of the twelfth 5-year, the power consumption of data centers will increase by 56%. The problems on energy efficiency are mainly embodied in two aspects.

First, energy efficiency technology fails to be promoted and applied widely. China's data centers are still mainly based on medium and small-sized traditional computer rooms, so that energy efficiency management technology and virtualization technology application have been not widely developed, resulting in that the PUE of data centers is generally ranged between 2.2 and 3.0; the non-IT facilities have an excessively high power consumption, and the absolute value of total power consumption is higher than the world's advanced level.

Second, the utilization rate is quite low. According to statistics, in China, the average source utilization of data centers is 20–30% and the servers are idle under 4/5 cases. However, as for more traditional servers, their power consumption is 60% of the peak value even they are in idle state, so the power consumed by idle equipment and cooling devices in the data centers will be increased by 50–100%. Data center are mainly owned or operated by communication industry, finance industry and government, and there is excessive and repeated construction in the building of data centers. For example, communication companies and banks separately build data centers on the basis of different regions, and these centers cannot be integrated due to the problems of information security, business development, system operation, etc.

7.6 DEVELOPMENT TENDENCY

7.6.1 Market Trends

With the implementation of national policies and measures for emerging strategic industries, especially the active promotion of cloud computing and next-generation network, China's data center market will experience an unprecedented opportunity of development, mainly embodied in the following aspects.

First, the data centers will become more centralized and larger scale. With the development of the first national cloud computing service innovation pilots, Beijing, Shanghai, Shenzhen, Hangzhou, Wuxi, and other places will be the main cores of data center industry in China. In addition, some places that have advantages in terms of location and geographic environment will be the key nodes by joint drive of local authorities and commercial interests. For example, Chongqing is favored by China Telecom by virtue of its core location in southwest and abundant water resources; Heilongjiang, Inner Mongolia, and Xinjiang, due to their natural cooling resources, international location advantage, and cheap labor cost, will be the concentration areas for the commercial data centers.

Second, the application market of data centers will become more developed. Data centers have lot of development opportunities in cloud computing application, especially the application of extra-large public cloud data centers in the fields of Internet of things, public service, e-commerce, smart city, tele-medical service, distant education, smart agriculture, social network, content search, and so on. The market model of data centers with Chinese characteristics will be gradually formed.

Finally, green data centers will become the mainstream of development. The data center energy efficiency and green data center have been listed as the key points in the *12th Five-Year Plan of Industrial Energy Saving and Emission Reduction* and *Comprehensive Work Proposal for Energy Saving and Emission Reduction* of the 12th *Five-Year Plan*. In addition, operating cost is the key driver to promote green data centers. China's data centers are at a quite low specialization level, and there is huge optimization potential of energy efficiency. Energy efficiency technology in the field of IT equipment, air conditioning, UPS device, optimized distribution of computer rooms, renewable energy sources supply such as building data center at an area with plentiful solar energy and wind energy, and virtualization technology will be the major subjects of research and development, and there is a larger market space.

7.6.2 Policy Trends

At present, Chinese government has been supporting the construction of data centers by multiple encouraging policies and the future trends are as follows.

First, perfect the work coordination mechanism of data centers. As the data center industry is related to many industries and fields such as electronics, communications, electricity, construction, energy saving, and environmental protection, these industries play an important role in pushing the data center industry, but none of them can dominate the whole industry. Therefore, it is required to establish work coordination mechanism of data centers, in which communications and cooperation will be enhanced; the industries will be well positioned to make their advantages, speeding up the development of data center industry.

Second, enhance the support toward technology research and development of data centers. ZTE, Huawei, Inspur, Dawning, Great Wall, and Lenovo have engaged in research and development of key technology for data centers in China; however, compared with world's advanced technology level, there still is a significant gap. With the growing dependency on technologies, it is estimated that China will fund to support the key technology research and development of data centers by means of Strategic Emerging Industries Fund, Major Science and Technology Programs, National Science and Technology Supporting Plans, Electronic Information Industry Development Funds, and other channels.

Finally, push the major application markets. As China has recently focused on the informationization of medium-sized and small enterprises and broadband project, a series of policies on finance and taxation have come to being or are under preparation, and the demands of e-government affairs, e-commerce, and public service establishment for data centers have been found out, so a complex application market will come into being, which will further push the development of the data center industry.

7.6.3 Standardization Trends

At present, China is short of specific standards for the data centers; especially, with the development of technology and demands for data centers, data centers keep changing daily. The research institutes and government are engaging in formulating related standards. China has the following standardization trends of data centers:

First, enhance the cooperation with international high-level institutes and accelerate the research on integrated standardization of data centers. Data center standardization in major developed countries is growing quickly with rich experience in practice, has tightly integrated with industry development and national policies, and relatively complete systems have been established. Through communication, a foundation can be laid for study on the energy efficiency technology of data centers in China. China Electronics Standardization Institute and other major data center standardization research institutes in China have deep cooperation with Lawrence Berkeley National Laboratory, UL, Intel, and other major international data center standardization research institutes and industrial driving forces.

Second, promote the demonstration of data center standardization in large national and local projects. Research on data center standardization shall be carried out based on a full combination with China's actual situation; especially, the data center projects supported by national cloud computing fund are representative and can be taken as the object of study. Meanwhile, standardization work may provide technical guidance for these large projects to ensure high-level construction and operation.

REFERENCES

[1] China Internet Network Information Center (CNNIC). 29th Statistic Report on Internet Development in China; December 2012. p 12.

[2] Zhang K. Informationization Status of Major National Corporations. People's Posts and Telecommunications News, May 4, 2012.

[3] Zhou Z. China's market of data center construction is under steady growth. E-Business Journal, July 2010. p. 16.

[4] State of Council. Decision of the State Council on accelerating the cultivation and development of strategic emerging industries. State of Council; 2010.

[5] State of Council. Outline of the twelfth 5-year National Economic Development Plan. State of Council; 2010.

[6] MIIT. Twelfth 5-year development plan for electronic information manufacturing industries. Ministry of Industry and Information Technology; 2012.

[7] State of Council. Industrialtransformation and upgrading plan 2011–2015; 2012.

[8] MIIT. Twelfth 5-year development plan for communication industries; 2012.

[9] MIIT. Twelfth 5-year development plan for internet industries. MIIT; 2012.

[10] State of Council. Several opinions of the State of Council on further promoting the development of small and medium-sized enterprises. State of Council; 2009.

[11] Rebo Industry Research Center, Market Survey and Prediction Report on Internet Data Center Market of China (2011~2015), Nov 2011. Information Industry Research, CCID Group; April 2010. p 1.

8

OVERVIEW OF DATA CENTERS IN KOREA

Minseok Kwon, Mingoo Kim, and Hanwook Bae
Samsung SDS, Seoul, South Korea

8.1 INTRODUCTION

8.1.1 Data Center Market in Korea

The Republic of Korea ranked first out of 155 countries in the 2012 ICT Development Index (IDI) that was published by the International Telecommunication Union (ITU) and is continuously growing in economy based on the growth of ICT.

The Korean government established a plan for vitalizing the data center industry and is planning to begin its operation shortly. This is based on the idea that the ICT domain is the foundation of the nation's growth and the significance of data centers is magnified as the supporting infrastructure for the expanding big data and cloud computing.

According to the research data surveyed by the Korea IT Service industry Association (ITSA) in 2012, there are 81 major data centers in Korea, and within the 68 that responded, 22 data centers have whitespace of larger than 4000 m^2, and 17 data centers are between 1000 and 4000 m^2.

The data center market size has grown from $325 million in 2005 and $505 million in 2007 to $900 million in 2010, and it is expected in the industry to reach $2 billion in 2015.

In spite of the market size and the significance, data center is not classified as an industrial category nor does such an industry exist, so the problems are that an accurate size of the market is difficult to verify and that the regulations for general commercial buildings are evenly applied to data centers.

Thus, in order to promote data center industry, a data center council is organized in the ITSA, and it is making an effort to define data center as an industrial category and make improvements to legal regulations.

8.1.2 Characteristics of Data Centers in Korea

A unique characteristic of data centers in Korea is that most of the major data centers are located within the capital area around or in Seoul, as the economic activity in Korea is mainly focused in the capital city. Data centers located in the capital area are usually high-rise buildings because of the high land price, and due to the legal restrictions and limit for allowed area and height of the building above ground, the facilities are often placed in the basement despite the risk of being flooded.

Korean people, as well as corporates, have a very high requirement level for ICT services, and they expect an immediate response when a failure occurs. This requirement is one of the reasons that data centers are located in the capital area where there are relatively more advance skilled technicians and engineers. Furthermore, to station all the needed ICT staff and engineers in the data center, it is a distinct feature to have a data center and an office workspace combined in the same building to accommodate these employees.

At present, only three data centers in Korea are certified from the Uptime Institute and all of them are Tier 3. However, a lot of major data centers are autonomously evaluated to be equipped for Tier 3 level.

As of March 2013, the electricity rate has risen to 90 cents per kWh from 68 cents since 2010, which is about 32% increase, but this rate is still lower than the leading countries such as the United States, Japan, or Europe. Therefore, data centers were constructed with relatively more consideration on reliability perspective than financial efficiency, and their owners were conservative about the use of free cooling or high-efficiency power facilities. However, recent rapid

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

increase in electricity rate and legal regulations by the government on energy usage have increased much interest in energy efficiency.

Another interesting point of data centers in Korea is the design of water pipes. There were cases of data centers being flooded in the year 2000 by fault operation of fire protection sprinkler and also in 2010 by rupture of frozen water pipe. This made the data center owners and designers avoid water pipes installed through the server rooms and electric facility rooms at all costs.

Last, but not least, is that there is a limit on allowed utility power feed capacity for each incoming line to a building. The generally used 22.9 kV power for data centers is limited to a maximum feed capacity of 40 MVA.

8.2 KOREAN GOVERNMENT ORGANIZATIONS FOR DATA CENTER

8.2.1 Government Organizations

In Korea, there is no department or official organization that exclusively handles data centers. ICT policies and technology integration-based work are currently managed by the Ministry of Knowledge Economy (MKE), but from 2013, the government organization will be reformed and the work will be transferred to a newly organized Ministry of Future Creation and Science (MFCS). Details of each ICT domain inside the MFCS and their roles and organization are not confirmed at the time of this writing. The supervising departments and their roles could be changed or revised in the future.

The Korea Communications Commission is a presidential affiliate establishment that is responsible for broadcast policies and regulations, communications service policies, and regulatory functions. Its main function includes establishing convergence policies for broadcasting and telecommunications, activating convergence services and development of related technologies, making policy on radio signal and managing signal resources, promoting competition in communications market, enhancing broadcast and communications network, preventing adverse effects, developing protection policies for network users, investigating unfair acts of communications providers, and mediating conflicts. Its key goal is to realize a "Global ICT Hub Korea," create a smart service environment and new industry, implement an advanced digital broadcast system, and reinforce the welfare of telecommunications users [1].

The National Information Society Agency (NIA) develops the government policies related to the prosecution of national informatization, supports the development of healthy information culture, and alleviates information provision gap. Its services include e-government support, construction of national database, promotion of ubiquitous-based public services, improvement of the information usage environment of developing countries, and others.

The National IT Industry Promotion Agency (NIPA) is a government agency responsible for providing support to IT enterprises and professionals. The NIPA leads to national economy development and knowledge-based economic society by promoting competitiveness of overall industries through IT usage and advancing IT industries. According to the change of IT industry, the NIPA will lead the way to consolidate Korea's position as an IT powerhouse through the four key business objectives of advancing the IT industry, converging and fusing the IT industry with traditional industries, expanding the foundations of the IT industry, and fostering the SW industry [2].

8.2.2 Associations and Other Organizations

The ITSA, an affiliate of the MKE, is the voice of the Korean IT service industry by representing a diverse IT service community spanning SI, consulting, solution, outsourcing in IT industry, and the IT convergence sector with health, automation, e-government. The ITSA has more than 50 leading IT service companies as members. The main activities of the ITSA cover policy researches, global cooperation and export support, holding conferences and seminars, and counsel operations in accordance to analyzing publications from IT industry.

The Telecommunications Technology Association (TTA) is a leading nongovernmental ICT standardization organization that develops and establishes ICT standards and provides testing and certification services for ICT products all at once. The TTA seeks to strengthen mutual cooperation and collaboration among all ICT industry participants such as telecommunication service providers, industry entities, academia, research laboratories, and organizations and to contribute to the advancement of technology and the promotion of ICT industry as well as the development of the national economy by effectively driving tasks related to ICT standardization and testing and certification. This goal shall be met through the provision/utilization of information collected investigated/researched on the latest technologies, standards, and testing and certification in domestic and international ICT sectors [3].

8.3 CODES AND STANDARDS

In Korea, a "data center" is not yet defined as a separate building purpose according to the local construction law. Data centers acquire building code permits as an "Integrated Telecommunications Facility" or a general business building. As it is difficult to list up the building design codes, regulations, and standards in the limited pages, this chapter will only include the major standards and regulations to be considered when building a new data center in Korea.

8.3.1 Standards

The National Radio Research Agency (RRA) announced the Guideline for Establishment of Green Data Center (KCS. KO-09.0065) as a national standard in 2012, of which an international standard, ITU-T L.1300, is the basis [4]. Another national standard, the Green Data Center Maturity Model (KCS.KO-09.0082), is being developed as an international standard as ITU-T SG5 “L.measure_Infra” [5].

8.3.1.1 Comprehensive Standards Data centers for the finance industry need to follow the regulation of the Financial Supervisory Service (FSS), whose main role is the examination and supervision of financial institutions but can extend to other oversight and enforcement functions as charged by the Financial Services Commission (the former Financial Supervisory Commission) and the Securities and Futures Commission.

8.3.1.2 Standards for Buildings The building has to be certified of the building energy efficiency grade under the regulations “Certification of Building Energy Efficiency Grades” (Ministry of Land, Transport and Maritime Affairs (MLTM) notification no. 2009-1306, MKE notification no. 2009-329) and “Rational Energy Utilization Act” of funding and tax deduction support (MKE announcement no. 2011-81).

The Building Certification System provides objective information regarding a building’s energy performance based on Article 66-2 of the construction act (Certification of Building Energy Efficiency Grades), regulations regarding building energy efficiency certification (MLTM notification no. 2009-1306, MKE bulletin no. 2009-329), and guidelines for public organizations’ energy use rationalization (Prime Minister’s directive #2008-3: 2008.6.12) [6].

Furthermore, the building has to be designed according to the energy saving design criteria (MLTM notification no. 2010-1031), and the owner has to submit an energy saving plan to the government.

In case that the data center is constructed in Seoul, the building has to be designed and constructed according to the eco-friendly and energy building standards from the Seoul green building design guidelines, and the building has to be certified for green building and energy efficiency grade. There are acquisition and property tax incentive benefits for certified buildings following the acquired grades.

According to the Rational Energy Utilization Act enforcement ordinance, “article no. 20 Submission of Energy Utilization Plan,” any building or facility that uses more than 20 million kWh of power in a year has to submit the energy utilization plan to the MKE and acquire a deliberation. The review and consultation of the plan is delegated to Korea Energy Management Corporation, and it is required to apply usage of regeneration and renewable energy during this deliberation process.

8.3.1.3 Green Data Center Certification The ITSA announced the Green Data Center Certification Program in October 2012, with support from the MKE [7]. The program consists of qualitative and quantitative evaluation. The quantitative evaluation for 90 points mainly measures the data center’s annual PUE, while the qualitative evaluation for 10 points focuses on the efforts and innovations performed by the candidate data center for better energy efficiency. The sum of those two evaluation scores defines the certification level, from the highest level of A+++ with 90 points or above to the lowest level of A with 60 points or above.

The quantitative PUE measurement is currently based on category 1, considering existing data centers’ energy usage measurement facilities. PUE under 1.40 can get the full 90 quantitative points, 80 points to PUE 1.50, 70 points to PUE 1.60, etc., up to 50 points to PUE 1.80.

The qualitative points are given to PUE monitoring and management activities up to 3 points, to energy efficiency activities such as government awards up to 2 points, and to assessment activities based on the Guideline for Establishment of Green Data Center (KCS. KO-09.0082) up to 5 points.

The Green Data Center Certification Committee has sub-organizations such as the Green Data Center Technical Committee and the Green Data Center Evaluation Committee and Evaluation Group.

Six data centers are certified in 2012 based on this initial evaluation framework. The ITSA has a plan to weigh PUE less and to add evaluation items such as IT equipment efficiency, data center power and efficiency management facility, building eco-friendliness, etc., in second-phase evaluation framework to be announced.

8.4 DATA CENTER DESIGN AND CONSTRUCTION

8.4.1 Building Data Centers in Korea

Most of the major data centers in Korea are located in Seoul or in the capital area of Gyeonggi Province that surround Seoul, so they are in the form of high-rise building because of the high land price in these areas.

In addition, due to the regulations that limit the total area and the height of the building above ground, facilities for the data center such as generators, UPS, and chillers are often placed underground despite the risk of being flooded.

Major data centers in Korea have a tendency of being built with more safety and reliability factors considered than cost saving. This is because quite a number of large enterprises have IT service subsidiary companies and the expectation level of IT services is very high. The national expectation level of IT services itself is so high that even a short service outage news is spread across the press and the Internet, so the expectation for reliability is higher than any country. This is another reason that data centers are located in the capital area where it is relatively easier to find advance skilled engineers and has good accessibility.

At present, the primary configurations of a general data center in Korea are $N+1$ standby generator, $2N$ UPS, $N+1$ chillers/cooling towers, and $N+1$ precision air conditioner, which in overall would be above the Uptime Institute's Tier 3 level.

Furthermore, as mentioned previously, due to the circumstances that put reliability above cost efficiency, humid and high-temperature summer climate, and comparatively cheaper energy bills, data center owners were not actively applying air-side free cooling. Recently, however, as the electricity rates peaked rapidly (68 cents in 2010 → 90 cents in 2013, 32% increase) and energy utilization was limited by the government's regulations, free cooling application and high-efficiency power facilities are gaining interest and are being introduced quickly.

8.4.2 Electrical Considerations

8.4.2.1 Power System of Korea The Korea Electric Power Company (KEPCO) is a public enterprise established in 1898 to supply electric power in Korea. In 2001, the power generation division of the company was separated into six subsidiary companies, but the power delivery/transmission division remains as a sole provider in Korea.

Power transmission voltage in Korea is in four steps of 765, 345, 154, and 22.9 kV; and the main power system supplied to general household is 380 V three phase and 220 V single phase.

One of the things to be considered when designing a data center in Korea is the regulatory limitation of the maximum incoming power capacity. The maximum incoming capacity is 40 MW for 22.9 kV and 400 MW for 154 kV. Therefore, anyone who needs more than 40 MW has to receive power at 154 kV, and the 154 kV substation has to be constructed within the data center at its own cost. Recent increase in IT power density led to drastic increase of total power capacity for the data centers, but this regulation is an obstacle for enlarging a data center scale in Korea.

Due to the economic feasibility issue of substation installation cost, space, and operation, currently, there is only one data center that is constructing its own 154 kV substation, but if the power density increases at the current rate, there would be more data centers with self-equipped 154 kV substations under this regulation. However, it is difficult to build a power plant in the capital area, and as it is hard to construct additional power transmission lines leading into the capital area due to ecological issues, it would not be easy to get permission from the government to construct a large data center with 154 kV substation within the capital area. It is likely that in the future, large data centers would have to be constructed in non capital areas, especially outside of Seoul.

8.4.2.2 Electrical Configuration As mentioned in the previous section, most of the data centers receive power at the voltage of 22.9 kV. Table 8.1 [8] shows the utility power feed method of major 68 data centers, researched by the ITSA in 2012.

According to the table shown earlier, 25 data centers, which is 64% of 39 data centers larger than 1000 m^2, receive power from two substations; 15%, which is six data centers, receive power from the same substation through two different transmission grids; and only 10%, which is four data centers, receive power from a single line.

Recent, newly built large data centers are designed with two utility power feeds from separate substations as a default, but there are practical problems in actual construction cases. Some examples of the problems are that there are not enough reserved banks to supply high capacity, that it takes a long time to add or increase its capacity, or that there are enough reserved banks but the conduit is too small for additional supply.

In case that a new power transmission cable conduit has to be constructed, the regulation that prohibits reexcavation of roadside for a certain time after the previous dig work could act as a setback. Moreover, unlike some other countries, the expense for 22.9 kV utility feed construction from

TABLE 8.1 Utility power feed method of major data centers in Korea

Raised floor area	Total	Type A	Type B	Type C	Type D	Other	Not answered
Over 4000 m^2	22	13	4	2	0	3	0
1000–4000 m^2	17	12	2	2	0	1	0
<1000 m^2	29	9	7	7	4	1	1

Type A: Feed source from two separate substations.
Type B: Feed source from the same substation through two different transmission grids.
Type C: Feed source from a single substation through single line.
Type D: One source from the KEPCO, another source from own power plant or substation.

the substation up to the data center has to be covered by the user, not by the KEPCO.

The next step-down voltage following 22.9 kV is generally 6.6 or 3.3 kV. This voltage is determined by the voltage used by the chillers, and it's also the voltage for standby generators. IT equipment use 380 V three phase and 220 V single phase after another step-down through the transformer.

Since 1970, the government started the project to raise the household power voltage from 110 to 220 V in order to reduce the loss, and the project was finally completed in 1999. However, there are still some equipments that use 208 V three-phase power, and a separate transformer is additionally installed for 380–208 V step-down.

As for the UPS system used in data centers in Korea, although a detailed statistics data is not available, a majority of the data centers are equipped with static UPS, while a few data centers use dynamic UPS. Large data centers are usually configured for $2N$ configuration, some of which may be $N+1$ configured to reduce the UPS installation cost, but the delivery or distribution system is still mostly configured as $2N$.

The backup time for static UPS is generally about 30 min, but the standard is changing for recently built data centers within 1–2 years, which have 15 min backup time. An interesting point is that 2 V batteries are dominantly used. Some small data centers are equipped with 12 V batteries, but most of the large data centers use 2 V batteries. Active research and proof demonstrations on lithium-ion batteries are being made lately, and there is an applied case in one of the major data centers, as these batteries have advantages on environment, space, and weight.

Many large data centers in Korea are built in a high-rise building form, for the land price is very high. Thus, the mechanical and electrical facilities are placed in the basement floor, which makes the distance from the UPS to the actual load substantially far apart. The main power delivery system used in between is usually bus way type instead of cables.

Besides the major power distribution methods, there is a case where Korea Telecom (KT) demonstratively applied DC power distribution in one of their data centers, but this method is not getting much attention yet. Currently in Korea, there is a standardization work in progress to designate DC 300 V as a standard, which can be used by existing AC 220 V power supplies.

8.4.3 Mechanical Considerations

8.4.3.1 Heat Load and Cooling Capacity The IT heat load and cooling capacity would vary depending on the scale and size of the data center along with the industry, but as the cooling capacity is generally planned based on the IT power consumption and the mechanical facility installed accordingly, it is safe to say that the majority of data centers in Korea do not have difficulties in maintaining the cooling of the IT heat load. In fact, they are sometimes designed with excessive capacity, and the cooling facilities are installed with much safety margin overhead. Recently, however, the total power and cooling capacities of the data centers are estimated and designed based on the average power density of each IT rack, and there are quite a few data centers that failed to forecast the power usage and that exceeded the expected amount but the cooling is not enough due to the lack of space for cooling facilities. In order to accommodate the growing heat load, the number of cases for close-coupled cooling application is increasing, especially where there is a need for high- or ultradensity IT equipment.

Most of the data centers in Korea are using perimeter cooling as their main cooling solution. In contrast to using perimeter cooling, which has a limitation on cooling high-density heat load, the power density of IT equipment is growing rapidly and constantly. The heat dissipation of a single IT equipment has increased in multiples compared to the past, due to system consolidation and virtualization and cloud implementation. Therefore, the dependency on close-coupled cooling, which is usually an auxiliary solution, is in an increasing trend. Currently, the average power density of a data center in Korea is about 2.0–3.0 kW per rack, but the proportion of consolidated high-end systems and cloud systems with a power density of more than 10 kW per rack is continuously increasing. Recently designed and constructed data centers have about twice the capacity with an average power density of 5–6 kW, prepared for heavy power usage and heat load.

8.4.3.2 Free Cooling As the scale of cooling facilities increase proportionally to the heat load growth, the reduction of energy consumption on cooling is the key topic nowadays, followed by a widespread understanding of implementing free cooling solutions.

During earlier times when free cooling was first introduced, direct air-side free cooling solutions were mainly used in many data centers as they were easy to implement and have low on cost. But soon, they experienced higher expenses on humidity and air contamination control. Korea has a climatic characteristic of very high humidity during the summer and relatively low humidity during the winter (Table 8.2).[1] This made free cooling a nuisance and overcost solution instead of an energy saver. Also, during that time, ASHRAE's expanded allowable environmental condition range for ICT equipment was not publicized yet. So, the strict environmental condition and the minimal saving of operational cost made direct free cooling, or free cooling itself, an avoided solution. Later, new data centers

[1] Average climate data of Seoul (1981–2010), KMA (Korea Meteorological Administration).

TABLE 8.2 Average annual climate data of Seoul

Statistic	Units	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Average
Temperature (mean value)	C	−2.4	0.4	5.7	12.5	17.8	22.2	24.9	25.7	21.2	14.8	7.2	0.4	11.82
High temperature (mean daily value)	C	1.5	4.7	10.4	17.8	23.0	27.1	28.6	29.6	25.8	19.8	11.6	4.3	16.54
Low temperature (mean daily value)	C	−5.9	−3.4	1.6	7.8	13.2	18.2	21.9	22.4	17.2	10.3	3.2	−3.2	7.82
Precipitation (mean monthly value)	mm	20.8	25.0	47.2	64.5	105.9	133.2	394.7	364.2	169.3	51.8	52.5	21.5	114.21
Relative humidity (mean value)	%	59.8	57.9	57.8	56.2	62.7	68.1	78.3	75.6	69.2	64.0	62.0	60.6	68.48

implemented indirect free cooling or water-side free cooling using flat panel heat exchangers. But as there are still strong advantages of air-side economizers and increased number of applied cases worldwide, domestic application method and merits are actively being researched in Korea.

Free cooling operating condition is based on outside air wet bulb temperature of lower than 4°C, with about 32% of water-side economizer available time year-round. This provides the condition for condenser water to be 9 °C and so the temperature of chilled water to be 10°C, which is higher than the generally operated condition of 7°C for most of the data centers in Korea. There are efforts to save energy by raising the chilled water temperature, but these are not very common yet due to the conservative understanding of stability issues.

Free cooling application in data centers in Korea is still not in a mature state, and there are not many cases yet. Until recently, there were no official energy saving cases and samples gathered by the government or any organization, so a detailed case research data on energy saving native to Korea is not yet available, and it would seem to be a while before we get a meaningful statistical data.

8.4.4 Architectural Considerations

Korea is considered to be a seismic-safe area compared to other regions located on the Pacific Rim. Ever since the seismic activity was recorded, the most powerful earthquake that occurred was in 2004 at the seabed 80 km off the eastern coast of Uljin, and the magnitude was 5.2 Richter scale [9]. It is assumed and recommended that data center building structure should be designed utilizing Seismic Performance Level Design methodology and Site-Specific Geotechnical Soil Tests and Analysis results. Most data centers apply an Importance Factor of 1.5 per TIA-942 for Tier 3 level designation as this will provide a return period factor of over 1000 (year). However, since the 2011 Great East Japan Earthquake, the interest in seismic activity is rising and some companies are recently constructing data centers with seismic isolation technology and making effort to attract Japanese clients who are very much concerned about earthquakes.

8.4.5 Other Considerations: Service-Level Agreement, Office, and Fast Deployment

In the past, there were some clients who requested Service-Level Agreements (SLA) with expansive responsibility and unlimited liability for business loss caused by ICT service failures, but much of those practices have been purged recently.

As the expectation of IT service level is so high, not only for corporate clients but also for individuals who are the end users, corporate clients of data centers are battered with complaints from the end users whenever there is an IT service failure, and this would directly impact their business.

A significant number of large enterprises have their own subsidiary IT service companies, and these IT companies that rely most of their revenues on the affiliate companies have extreme repulsion to any fault from IT operation that has an influence on their clients' business. This makes them prefer reliability rather than efficiency for data center construction and operation.

Therefore, unlike foreign data centers that operate with a minimum number of staff for facility and network, data centers in Korea usually operate with all-internal ICT technicians and also the vendor engineers stationed locally. This is the reason for a very wide office area in large data centers compared to foreign ones. According to a survey report by the ITSA in 2012, data centers with whitespace of over 4000 m^2 have an average of 177 staff members and 56 people working for data centers of 1000–4000 m^2.

Another characteristic that differs from the data centers in United States or Europe is the construction period. This could be analyzed as a factor for modular data center not being popular in Korea, which was a worldwide trend since late 2000s. Small data centers between 100 and 200 m^2 are usually built within one or two months, which is similar or shorter than what most highly priced modular solutions claimed to be a "faster deployment."

In Korea, especially in capital areas, there are limited lands suitable for data centers, which lead to relatively high land price; hence, most of the data centers are built as a form of high-rise building. In spite of these building forms, it generally takes about 1 year without underground and around 2 years even with underground excavation for a data center to be "operation ready," and it is a prevalent expectation level in the market.

8.5 DATA CENTER MARKET

8.5.1 Current Data Center Locations

Most of the data centers in Korea are concentrated in Seoul or capital area, let alone some data centers for government or specific purpose in rural districts (Table 8.3). A fundamental reason is because 47.1% of Korea's gross regional product is from the capital area [10] and 72.1% of stock exchange listed corporates are located within this area [11].

Similarly, as it is easier to find ICT skilled labor in this area, data centers are preferred to be located where it is better for securing operation manpower and has good accessibility for failure recovery support. Specifically, as mentioned in the previous chapter, the high expectation level of the clients for ICT services and the significant impact of downtime cause even higher preference and business demand for the capital area, where ICT skilled personnel can be stationed in the data center, SLA for vendor response time at system failure can be requested within hours, and the clients' IT department staff can immediately have a visit at certain events.

The building cost of a data center in the capital area is high due to the high land price, but another point being considered is the network, as the network traffic is mainly distributed around the capital area and network cost would be an additional burden if the data center is at a distant location.

8.5.2 Market and Policy Trends

The central government is concerned with data centers being converged in the capital area due to high electric power consumption, thus tightening the related regulations. One example is that from 2008 data centers had benefit on electric cost by a special exemption law, but since 2012, the MKE excluded the data centers in the capital area from the exemption. In contrast, district governments' effort to attract data centers to revitalize the economy of provincial areas is continuously growing. MKE and Busan City designated a "Global Cloud Data Center Model District" under the plan to develop Busan into a cloud data center hub of Northeast Asia, and an enterprise finished construction of their data center in 2012. Busan is particularly good for access from Japan, so there are efforts to attract Japanese corporate data centers or disaster recovery centers since the earthquake in East Japan.

The Gangwon Province area at about 40 km east of Seoul is noted for its low average temperature, and data centers are being constructed to utilize the advantages of free cooling for cost saving.

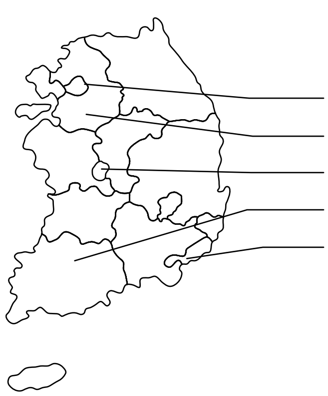
Pyeongchang, which is about 100 km east of Seoul, is also emerging as a point of attention for data centers, with much investment being made in communications networks and transportation means such as KTX express railways, in preparation for the 2018 Winter Olympic Games.

8.5.3 Case Studies

8.5.3.1 ***Water Phobia*** In Korea, there is a strong fear of bringing in water into the data center. There were some major incidents that this fear became reality, two of which were in financial company data centers, both with great post impact.

In 2000, a data center in a securities company suffered from flooding due to a fault operation of fire protection sprinkler on the upper floor. The impact was disastrous, with

TABLE 8.3 Current data center locations in Korea



Area	Data centers (%)	Large centers (>4000 m^2) (%)
Seoul	44.4	45.5
Gyeonggi	21.0	27.3
Daejeon	10.8	13.6
Gwangju	4.5	4.5
Busan	5.4	4.5
Others	13.9	4.5

the company's stock trading system becoming unoperational for days. The fault was due to negligence during fire protection system inspection.

Another data center flooding case was from a data center of a large banking company in winter of 2010. Cold air from outside continuously flowed into the building through a faulty air damper, freezing one of the water pipes. The water pipe broke and leaked water, eventually overflooding the data center. All banking systems were stopped for hours.

For such reasons, it is considered as a safety hazard to have water inside or near the data center, and the idea is pervasive that water leakage would lead to service being stopped with huge business impact. As these cases could occur to any company or data center, operators in Korea tend to have an animosity against chilled water close-coupled cooling system, which need to bring in water into the data center and also adjacent to the IT equipment. At the same time, water leakage detection system and water proof installations are considered as high-priority and significant facilities.

8.5.3.2 Water Storage Water, just like electricity, is an important factor in operating a data center.

In 2011, there was a city wide water supply cutoff in one of the cities in Korea. This was due to breakdown of the city's main supply pipeline. The cutoff lasted for over 3 days. Data centers in the city had to use the water for cooling system from their storage tanks. Large data centers in Korea usually have about three or more days of water supply stored locally, but smaller data centers, especially those located in general office buildings, would have great impact on cooling facilities if water cutoff period is prolonged.

This water cutoff case ringed an alarm for other data centers that, although many would think that electricity is the one major factor for stable operation of a data center, water is another key element and business could be compromised due to water supply dependency on utility provider.

Water storage in a data center is becoming a bigger issue recently, as Korea is designated as one of the countries that face water scarcity. A lot of data centers are now establishing or enlarging their water storage tanks and preparing duplex water supply source. Furthermore, the government is working on securing water supply for industrial sites.

8.6 CONCLUSION

So far, we had a look into the organization of the Korean government related to ICT industry and the regulations and standards for data centers in Korea. The design factors and characteristics of data centers and the current status of the relevant market were mentioned, as well as a number of incident cases.

The data centers in Korea have a distinct building form wherein large data centers are mostly high-rise buildings due to high land prices. They also have office workspaces in the data center building to accommodate large number of staff and engineers.

Data centers were designed and built with reliability preference rather than energy efficiency until the 2000s, but in recent days, data centers are so called as energy beasts and gaining social interest with being a subject of the government's energy policy regulations. The interest in energy efficiency is rapidly increasing, and this phenomenon is expected to continue in the future in that much effort is going to be made in the market to increase the energy efficiency in various perspectives.

Additionally, there is an activity in progress to revise the law and define the data center industry as a unique specific category.

The authors would like to thank the NIA, NIPA, and TTA for their support in writing this chapter and especially to the ITSA for providing valuable statistical data on local data centers.

REFERENCES

[1] Korea Communication Commission, 2013. Available at http://eng.kcc.go.kr/user/ehpMain.do. Accessed on August 11, 2014.
[2] National IT Industry Promotion Agency, 2013. Available at http://www.nipa.kr/eng/main.it. Accessed on August 11, 2014.
[3] Telecommunication Technology Association, 2013.
[4] Guideline for Establishment of Green Data Center (KCS.KO-09.0065), RRA, 2012.
[5] Green Data Center Maturity Model (KCS.KO-09.0082), RRA, 2012.
[6] Building Certification System, Korea Energy Management Corporation, 2013.
[7] Green Data Center Certification Briefing session material, Korea Information Technology Service Industry Association (ITSA), 2012.
[8] ITSA, 2012.
[9] KMA (Korea Meteorological Administration), 2013.
[10] 2011 Regional Income Press Release, Statistics Korea, December 24, 2012.
[11] Press Release, Korea Exchange (KRX), October 12, 2012.

PART II

DATA CENTER DESIGN AND CONSTRUCTION

9

ARCHITECTURE DESIGN: DATA CENTER RACK FLOOR PLAN AND FACILITY LAYOUT DESIGN

PHIL ISAAK
Isaak Technologies Inc., Minneapolis, MN, USA

9.1 INTRODUCTION

The success of the data center floor plan design process is dependent on the participation of the facility design team and the information technology design team. The facility design team will consist of the building architect and the electrical, mechanical, and structural engineers. The information technology design team will consist of those responsible for the network topology, server and storage platform architecture, and network cabling infrastructure design.

The design of the data center layout must be developed with input from all the key stakeholders from the facility and IT groups within the organization. This integrated design approach will help to ensure that the data center will function and perform throughout the facility life cycle, providing operational efficiency to support the many technology life cycles that the data center facility will see.

9.2 OVERVIEW OF RACK AND CABINET DESIGN

9.2.1 Two- and Four-Post Racks

Two- and four-post racks are open frames with rail spacing that should meet the EIA/CEA-310-E manufacturing standards.

The mounting rail rack units (RU) (1 RU = 1.75 in.) should be clearly marked on all rails. The RU markings typically start at 1 on the bottom. However, there are some manufacturers with products that start RU number 1 at the top. The RU designations should be consistently applied throughout the data center on two- and four-post racks as well as cabinets so there is no confusion for the technicians or the ability to integrate the Information Technology Equipment (ITE) mounting positions with a Data Center Information Management (DCIM) application.

Two- and four-post racks should be provided with vertical cable management on either side of each rack, with sufficient capacity to accommodate the maximum number of patch cords or cabling infrastructure that is anticipated within the rack. The vertical cable management should have fingers at 1 RU increments to align with the ITE mounted within the rack. The depth of the vertical cable managers mounted on either side of the rack and the depth of the horizontal managers mounted within the rack should be coordinated so that they are in alignment, providing an even smooth consistent pathway throughout the cabling pathway provided for the rack.

One challenge with two- or four-post racks is the ability to mount zero-U power outlet units (POU), or Rack PDU, to the rack rails or frame. This often requires nonstandard mounting solutions as the POU are manufactured to install with button holes within cabinets that do not exist on two- or four-post racks. For this reason, horizontal POUs are often used for two- and four-post racks.

The position of a row of racks first needs to be coordinated with the floor grid (if on a raised floor), any adjacent cabinets, and any overhead pathways. (Refer to Section 9.4 for further guidance on coordination with pathways. Refer to Section 9.5 for further guidance on coordination with other systems.)

The racks should be provided with a bonding point for bonding the rack to the data center grounding system. The bonding point should provide metal to metal contact without any paint or powder coating inhibiting the effectiveness of the bond. The resistance to true earth shall be either 5, 3, or

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

1 ohm (max.) measured by the fall of potential method (ANSI/IEEE Std 81), depending on the Class of data center per ANSI/BICSI 002-2011.

The recommended methods of grounding racks and cabinets may exceed the minimum requirements to meet the building codes. While the grounding requirements within building codes are provided for life safety and power system protection, the grounding requirements in standards such as the ANSI/BICSI 002 Data Center Design and Implementation Best Practices standard and the IEEE 1100 Recommended Practice for Powering and Grounding Electronic Equipment provide guidance for safety, noise control, and protection of sensitive electronic equipment.

9.2.1.1 Two-Post Racks A two-post rack provides a single rail to which the ITE is mounted on. It is recommended that the ITE mounting brackets be set back from the front of the chassis so that the center of mass is positioned at the point the ITE brackets are installed. Large chassis ITE (either in RU or depth) may require a special shelf to adequately support the equipment.

Two-post racks have typically been used in network IDF/MDF closets where there is a single row of equipment. The two-post rack should only be used in the data center where space constraints limit the use of four-post racks or cabinets. ITE that are mounted in two-post racks are more susceptible to physical damage as the ITE is exposed beyond the rack frame.

9.2.1.2 Four-Post Racks A four-post rack provides a front and back rail to which the ITE is mounted on. Manufacturers provide four-post rack models that offer fixed-position front and back rails or a fixed-position front rail with a variable-position back rail if required. The variable-position back rail typically is adjustable as a single system from top to bottom, so all ITE mounted in the rack must accommodate the rail position selected.

Four-post racks are typically the preferred open-frame solution as they provide greater physical protection for the ITE mounted within them than what a two-post rack offers. For example, fiber enclosures mounted in a four-post rack will have the back of the enclosure within the footprint of the four-post frame. The fiber typically enters the fiber enclosure at the back. If the enclosure was installed in a two-post rack, the fiber entering the enclosure would be physically exposed, but in a four-post rack, it is within the four-post frame footprint.

9.2.2 Cabinets

Cabinet are closed frames with rail spacing that should meet the EIA/CEA-310-E manufacturing standards.

The mounting rail RU (1 RU = 1.75 in.) should be clearly marked on all rails.

The design and selection of cabinets often overlook the details that are required to ensure a suitable solution. The cabinet selection is not only based on the ability to mount ITE to the rails but on the ability to support the network cable entry from overhead or underfloor pathways, POU implementation, and airflow management through and within the cabinet becoming increasingly important. The cabinet selection requires the designer to be knowledgeable in hardware platforms, network cabling infrastructure, power distribution, and cooling airflow to be able to recommend the appropriate solution.

The cabinets should be provided with a bonding point for bonding the rails to the data center grounding system. The bonding point should provide metal to metal contact without any paint or powder coating inhibiting the effectiveness of the bond. The resistance to true earth shall be either 5, 3, or 1 ohm measured by the fall of potential method (ANSI/IEEE Std 81) depending on the Class of data center per ANSI/BICSI 002-2011.

9.2.3 Network

Open-frame two-post racks are typically not used in the data center for the core network equipment due to the physical size and weight of the chassis. Open-frame four-post racks may be used as they are able to support the size and weight and provide suitable cable management. However, they do not integrate into a controlled hot–cold aisle configuration since there is no airflow management.

Manufacturers provide cabinet models that are purpose built to support server platforms or purpose built to support network platforms. It is not recommended to implement core network platforms within cabinets that have been purpose built for server platforms. Purpose-built network cabinets have options to support core network equipment with front-to-back airflow or side-to-side airflow within the same cabinet.

The physical requirement nuances between manufacturers and models of network equipment places an increased burden on the designer to ensure that the appropriate mounting frame is incorporated into the data center. Core network equipment will often consist of equipment with front-to-back and side-to-side airflow within the same cabinet. It is important to identify the appropriate airflow management solutions for the specific network platforms prior to finalizing the equipment elevation design and cabinet selection. When coordinating network equipment requirements with cabinet manufacturer options, the airflow management may require additional RU space above and/or below specific chassis to provide adequate airflow (i.e., Cisco 7009 platform), while others require external side cars fastened to the cabinet, increasing the width of the total cabinet solution, to provide adequate intake and exhaust airflow capacity (i.e., Cisco 7018).

The specific airflow path through a chassis must be validated prior to finalizing equipment layouts and cabinet selection. The following are some examples of various airflow management solutions, each requiring different

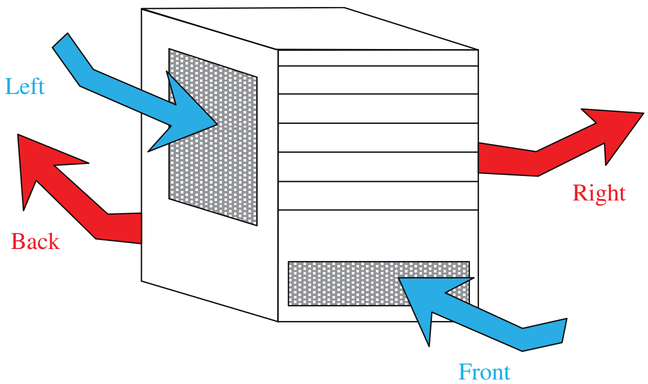


FIGURE 9.1 Network chassis with front-to-back and left-to-right airflow (e.g., HP12504 model). Courtesy of Isaak Technologies Inc.

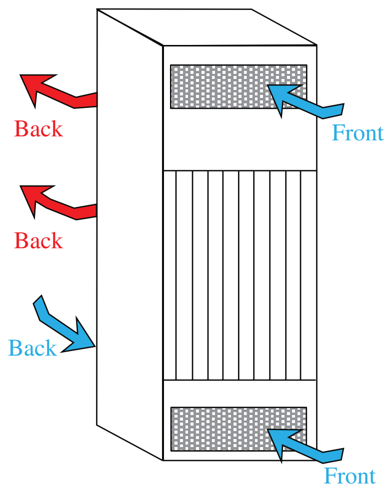


FIGURE 9.2 Network chassis with intake located on the back and front of chassis and exhaust out the back of chassis (e.g., HP12508 model). Courtesy of Isaak Technologies Inc.

approaches in cabinet design and equipment layouts (the port side of network switches are referred to as the front) (Figs. 9.1, 9.2, 9.3, and 9.4):

- Cisco 7004, 7009, 7018: side-to-side/right-to-left and front-to-back airflow
- Cisco 7010: front-to-back airflow
- HP 12504: side-to-side/left-to-right and front-to-back airflow
- HP 12508, 12518: lower chassis has intake in the front and back, and upper chassis has intake in front and exhaust in the back, resulting in front to back and back to back.

Cabinet manufacturers typically design their network cabinets to support side-to-side airflow in the right-to-left direction, with the port side of the network switch installed on the front rail (cold aisle). Network switches that have left-to-right airflow through the chassis may require that the network cabinets be provided by the network switch

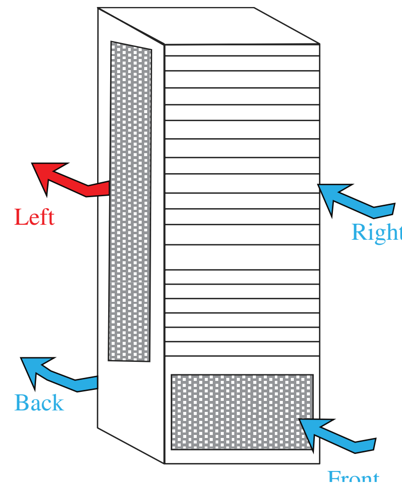


FIGURE 9.3 Network chassis with front-to-back and right-to-left airflow (e.g., Cisco 7018 model). Courtesy of Isaak Technologies Inc.

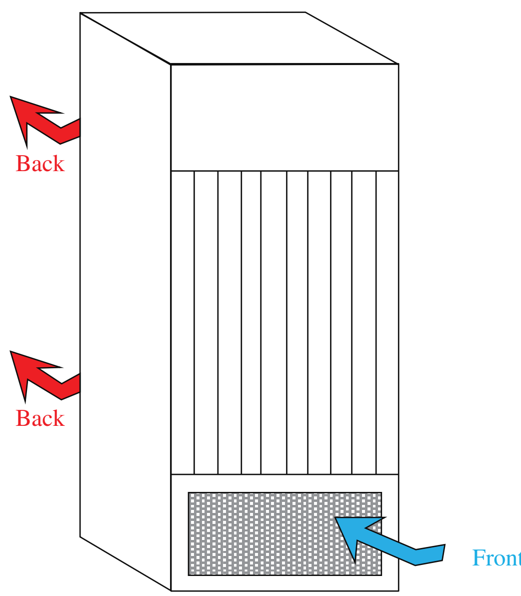


FIGURE 9.4 Network chassis with front-to-back airflow (e.g., Cisco 7010 model). Courtesy of Isaak Technologies Inc.

manufacturer to ensure proper airflow management as left-to-right airflow management accessories may not be readily available from the common industry cabinet manufacturers.

All associated fiber enclosures or copper patch panels providing interconnection to the network equipment should be installed so the port side of the panels aligns with the port side of the network equipment.

Network switches that are classified as top-of-rack or end-of-row switches often have the airflow through the switch from back to front. This is done to accommodate the port side of the switch (front) to face the same direction as the ports on the server, which are positioned on the back of the server. This enables airflow through the cabinet from the cold aisle to the hot aisle to support both the servers and the top-of-rack or end-of-row switches.

Network cabinets typically have more network connections for IT equipment than server or storage cabinets. Therefore, network cable management is one critical design criteria for the network cabinet design. Network connectivity may consist of copper twisted pair, fiber, and proprietary interconnections between redundant chassis or multiple chassis to make a single virtual switch stack. Adequate space with suitable bend radius should be provided within the cabinet for all network cabling, preferably with physical separation between the copper and fiber cables.

The management of power distribution cables is also a factor when selecting the appropriate cabinet manufacturer and accessories. It is recommended to have the power cables routed vertically in the cabinet in the opposite corner from the copper network cabling. Large network chassis typically will have multiple 20 A, or higher, power cords. The power cords should be provided with cable management solutions that are sized to support all of the cordage without exceeding the pathway capacity limits.

9.2.4 Server and Storage

Rack-mountable server and storage platforms are mounted within ITE cabinets. Some of the main considerations when selecting ITE cabinets for server or storage platforms include the following:

- Top or bottom entry of network cabling
- Top or bottom entry of power cables
- Width and depth of standard cabinet to be used
- Will a vertical exhaust duct be incorporated into the overall cooling solution
- Is physical security required for each or specific cabinets, and if so, are manual locks sufficient or is electronic locking to provide entry logs required

Cabinets that have bottom entry of either power or network cables will need to have floor tile cutouts positioned within the footprint of the cabinet. In order to provide flexibility with respect to where the floor tile cutouts are positioned, the cabinet frame should have minimal obstructions at the base of the cabinet. Some cabinet manufacturer solutions have flat plates that provide structural support between the inner rails and the outer cabinet panels. These flat plates may impede where the floor tile cutouts can be positioned.

The cabinet width and depth should provide sufficient space to mount at least two vertical power strips in one corner and route network cabling in the opposite corner.

Server and storage manufacturers with rack-mountable solutions may provide a swing arm accessory to manage power and network cable management to each device. The swing arms can significantly impede the airflow out the back of the server or storage platform, increasing the temperature within the cabinet. The benefit of the swing arm is that a device can be “racked out” without disconnecting the power or network connections. It is prudent to ask if standard operating procedures include powering down a chassis before any hardware upgrades or modifications are made. If this is a standard operating procedure, which is typical, then there is no need to have the swing arms on the back of the equipment.

ITE cabinets have historically been black in color. The designer may want to consider implementing white-colored ITE cabinets. White cabinets help to reduce the amount of energy required to provide the recommended lighting levels within the computer room.

9.2.5 Large-Frame Platforms

Large-frame platforms are systems that do not fit within a standard 2100 mm (7 ft)-high cabinet with standard EIA/CEA-310-E mounting rails. These systems often include large disk arrays for enterprise storage, mainframes, HPC systems, supercomputing systems, or tape libraries. The cabinets that are used to support these large-frame systems are often wider and deeper than typical server cabinets.

If large-frame platforms are used within a data center, the layout of these systems on the computer room floor must be planned early to ensure that the critical building systems are designed appropriately. Power distribution, cooling methodologies, lighting layouts, fire detection, and fire suppression system layouts must all be coordinated with the IT equipment layout.

It is also common for large-frame platforms to have the power and network connections enter from the bottom of the equipment. In a nonraised floor computer room, an appropriate method of routing the power and network cabling must be identified.

9.3 SPACE AND POWER DESIGN CRITERIA

Power demand density (W/sf) has often been used as the criteria to establish power and cooling capacity requirements. Power density is often used by facility engineers to define capacity requirements. However, inappropriate capacity projections may be defined if power density is used as the only metric to develop capacity planning.

Proper capacity planning does not simply identify the existing power density, say, 1000 W/m^2 (92 W/ft^2), and then apply some multiplier, say, ×2, and define the future power, and cooling requirements should provide 2000 W/m^2 (184 W/ft^2).

9.3.1 Platform Dependent Capacity Planning

The recommended approach to develop and define future space, power, and cooling capacity requirements is to analyze each hardware platform and the supported applications. This exercise is not a facility engineering analysis, but an enterprise architecture and IT analysis driven by types of applications used to support the business objectives.

Identify the hardware platforms and review historic growth by platform if available. Review the supported applications and identify impact of future requirements on hardware capacity planning. The hardware platforms are typically compartmentalized into the following categories: (i) network, (ii) server appliance (nonblade), (iii) blade server, (iv) large-frame processing (mainframe, HPC, etc.), (v) large-frame disk arrays, and (vi) rack-mounted disk arrays (Fig. 9.5).

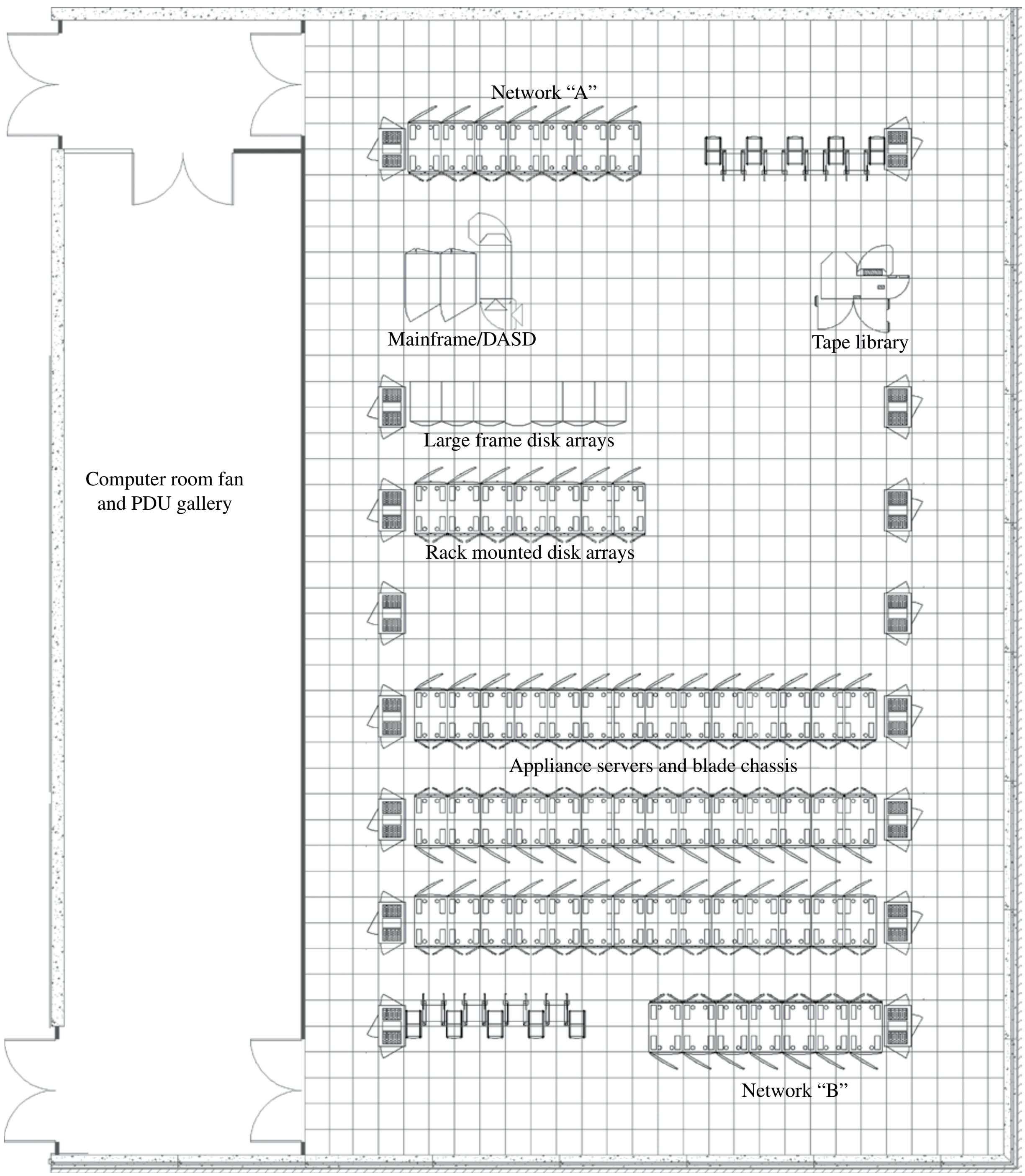


FIGURE 9.5 Example of multiplatform computer room layout. Courtesy of Isaak Technologies.

9.3.2 Refresh

Refresh capacity is required when applications or data are being migrated from legacy systems to new systems. Refresh also can have a significant impact on capacity planning. If an organization utilizes rack-mounted appliance servers and blade servers with very little storage requirements, refresh may not be that significant. These servers are typically refreshed individually, not as an entire platform. In this scenario, the refresh capacity required may be less than 5% (space, power, and cooling) of the total capacity supporting server platforms.

If an organization has implemented large-frame disk array platforms within their storage architecture and the disk arrays require a significant amount of the space, power, and cooling capacity in comparison to the processing

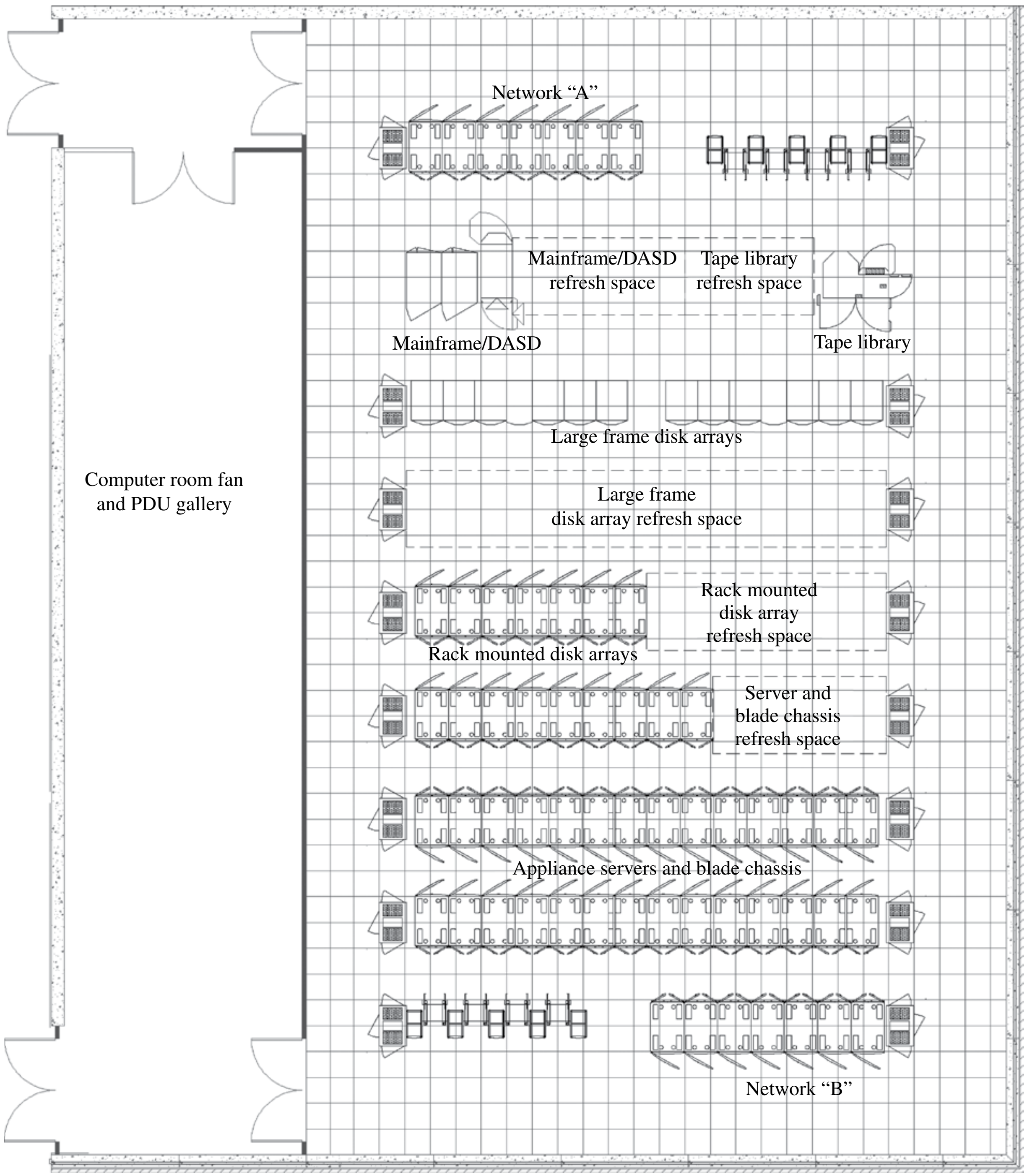


FIGURE 9.6 Example of multiplatform computer room with 18% space required for technology refresh. Courtesy of Isaak Technologies.

capacity requirements, the capacity planning results will differ significantly from the previous example. When disk arrays go through a technology refresh, they are typically refreshed as an entire system. If the entire disk array consists of two 9-frame systems, the refresh migration will require the space, power, and cooling capacity to stand up an entire new 9-frame system at a minimum. The legacy system and the new system will need to function alongside each other for a period of time (weeks or months). The new system will need to be powered up, configured, and tested before the data will be migrated from the legacy system. After the new system is in full production, the legacy system can be decommissioned and removed from the data center. This scenario results in the capacity plan requiring increased space, power, and cooling requirements for the anticipated disk array platforms to facilitate technology refresh.

If a data center's computer room, which is supporting multiple platforms, has more than 80% of the computer room space, power, and cooling capacity consumed, there may be little or no room for growth but only excess capacity available to support the refresh of the current platforms (Fig. 9.6).

9.3.3 Power Density

The power density of the computer room is an outcome of analyzing the space and power capacity planning exercise. It is not the recommended starting reference point for establishing power and cooling capacities. Once the growth and refresh requirements have been established as noted previously, the power density can be expressed.

9.4 PATHWAYS

9.4.1 Entrance Network Pathway

It is recommended that customer-owned maintenance holes be installed at the property line. Customer-owned conduits are to be installed between the customer-owned maintenance holes and the data center. This ensures that the customer manages and controls which network service providers have access to their data center and how they physically provision the fiber to the data center.

The elevation of each maintenance hole cover must be lower than the elevation of the entrance conduits terminated inside the data center. This will ensure that moisture does not migrate into the data center from a flooded maintenance hole.

The minimum recommended conduit size and quantity for entrance pathways are four 100 mm (4 in.) conduits to support up to three network access providers. The conduits will have either hard-walled or fabric inner duct placed inside the conduits to enable future fiber pulls without damaging the initial pull. When more than three network access providers are anticipated to serve the data center, one additional conduit is recommended for each additional provider. The conduits are recommended to be concrete encased from the maintenance hole to the facility, with a minimum of 1.2 m (4 ft) separation from any other utility.

The routing of the entrance pathway from the property line to the entrance room in the data center should meet the following requirements as documented in the ANSI/BICSI 002–2011 standard:

- BICSI Class 1: one route with at least four conduits to the entrance room
- BICSI Class 2: two diverse routes, each with at least four conduits to the entrance room
- BICSI Class 3 and 4: two diverse routes, each with at least four conduits to each entrance room

For BICSI Classes 3 and 4, where two entrance rooms are recommended, it is recommended to install 100 mm (4 in.) conduits between the entrance rooms as well. The quantity of the conduits between the entrance rooms should be the same as the quantity entering the building. These conduits are provided to give network access providers flexibility in how they route their ringed fiber topology to the data center, either in and out the same entrance room or in one entrance room and out the other. These conduits should not be used for any other function other than to support network access providers' fiber infrastructure.

9.4.2 Computer Room Pathway

Power or network pathways can be routed overhead or underfloor in a data center with a raised floor. There are viable solutions for either scenario. The decision as to which method to use is often designer or owner preference.

9.4.2.1 Overhead

Power Overhead power distribution methods consist of running power whips, or wire in conduit, from a distribution frame to receptacles above each cabinet or a wire busway with hot-swappable plug units.

The wire busway provides greater flexibility and reconfiguration as the ITE changes over the life of the data center.

Overhead power alongside overhead network requires greater design coordination to ensure that adequate physical separation exists between power circuits and any unshielded copper network links. Minimum separation between power circuits and unshielded network cabling is dependent on the quantity and ampacity of the power circuits (refer to ANSI/BICSI 002 standard).

Network Overhead network distribution is generally the preferred method used. Overhead copper network pathway options include top-of-cabinet trough systems, basket trays, and ladder rack. For overhead fiber-optic distribution, it is recommended to use fiber-optic troughs designed specifically for fiber-optic infrastructure. Cable trays, typically used to distribute power circuits, are not recommended for network distribution within the computer room of the data center.

The top-of-cabinet trough systems are often used in smaller data centers with one or two rows of ITE. The trough system requires less coordination, minimizes ceiling height requirements, and is a cost-efficient solution. The trough system impedes moving or replacing a single cabinet within the row in the future as the cabinet itself supports the pathway.

Basket trays are very common as they are a cost-effective solution. Basket trays ease the installation of the pathway in applications where there are numerous elevation changes.

Ladder racks are also very common as they provide the most options to assist in the transition of the copper cabling from the ladder rack down to the racks or cabinets. Top rung ladder racks allow the copper cables to transition off the side or through the rungs, with either method using water fall accessories to limit the bend radius. Ladder racks can also be used to directly support fiber-optic troughs if used.

Fiber-optic troughs designed specifically for fiber-optic infrastructure are the recommended method of distributing fiber throughout the data center's computer room. The fiber-optic trough ensures that the minimum bend radii are maintained throughout the pathway and at all transitions from the trough down to racks and cabinets. Split corrugated tubing should be used in the transition from the trough to the cabinet or rack to provide physical protection for the fiber-optic cables.

9.4.2.2 Underfloor

One simple and low-cost system that is often overlooked in new data center designs is to include lighting within the underfloor space. Including light fixtures below the raised floor will help provide a safe working space for any technician who needs to maintain systems within the underfloor space.

Power The routing of power cabling under the floor is often incorporated into current data center designs. The typical solution uses a liquid-tight flexible metal conduit or cable that lies on the floor below the raised floor system. Some jurisdictions require wire in conduit.

Network Routing network cabling under a raised floor requires coordination with all of the other building systems within the underfloor space. This includes power cables and conduits, chilled water piping, grounding systems, and fire detection and suppression systems.

The industry is trending away from routing network cabling in the underfloor space unless it is specifically for large-frame systems that require cable entry from below.

For copper twisted pair cabling, the most common method is to use a wire basket solution to distribute the network cabling. The wire baskets are either supported by stanchions independent from the raised floor pedestals, or they are supported by the raised floor pedestals themselves. When being supported by the raised floor pedestals, the designer must ensure that the pedestals will support the cable pathway system in addition to the raised floor and the equipment on the floor.

For fiber-optic cabling, there are purpose-built fiber troughs available to distribute the fiber cable. If the underfloor space is used as an air distribution plenum, the pathway must be rated to be used within a plenum space. The only products available to meet this requirement are metal troughs. Metal troughs are often not desired as they are typically made out of light gauge sheet metal that easily cut when being handled in a confined space.

Another method to ensure physical protection for fiber-optic cable that is distributed in an underfloor space is to use armored or crush-resistant fiber-optic cable. This method does require the designer to coordinate the quantity of fiber sheaths that will be terminated in any rack-mounted fiber shelves above the floor.

9.5 COORDINATION WITH OTHER SYSTEMS

There are many benefits to locating non-IT systems outside of the data center computer room. Non-IT systems such as Computer Room Air Conditioner/Computer Room Air Handler (CRAC/CRAH) units, power distribution units (PDUs), and uninterruptable power supply (UPS) systems are sometimes located in the computer room for smaller-sized data centers. Benefits of locating non-IT systems outside the computer room include minimizing the activity of non-IT personnel within the IT space, minimizing the activity of IT personnel within facility spaces, removing heat generated by the facility systems from the computer room, and simplifying the coordination of the IT systems placement within the computer room.

Various design solutions are available that result in non-IT systems located being outside of the computer room. One method is to configure galleries adjacent to the computer room where cooling and power distribution equipment is located. Another method is a multistory facility with the computer room above or below the level where the cooling and power distribution equipment is located (Figs. 9.7, 9.8, and 9.9).

In a raised floor application, it is important to understand the dimensions of the flooring system. The floor grid measurements of 600 mm (2 ft) are not a nominal equivalent typically used when expressing dimensions in metric, imperial, or U.S. units. In the United States, the floor grid is based on 2 ft square floor tiles that are 609.6 mm by 609.6 mm. In countries that have incorporated the metric system, the floor grid is based on 600 mm^2 floor tiles that are 23.622 in by 23.622 in.

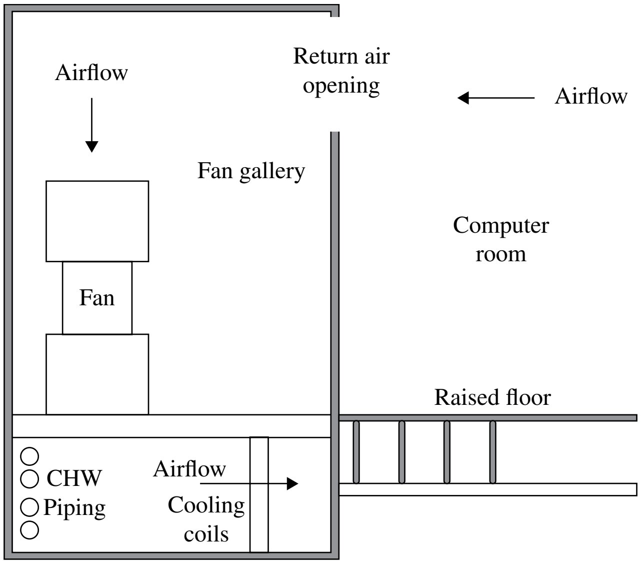


FIGURE 9.7 Example of Fan Gallery (section view). Courtesy of Isaak Technologies.

9.5.1 CRAC/CRAH

The CRAC/CRAH units are typically placed along the longest wall of the computer room, with the rows of ITE perpendicular to the longest wall. The CRAC/CRAH units are typically recommended to be aligned with the hot aisle to maximize the distance from each CRAC/CRAH unit to the closet perforated floor tile. This also simplifies the return airflow stream from the hot aisle directly to the CRAC/CRAH unit return air inlet, which is more critical for computer rooms with low ceilings.

A minimum of 1.2 m (4 ft) is recommended for cold aisle spacing. A minimum of 900 mm (3 ft) is recommended for hot aisle spacing, with 1.2 m (4 ft) preferred. The exact aisle spacing should be coordinated between the IT design and the cooling system design to ensure that adequate airflow can be delivered from the cold aisle and returned to the CRAC/CRAH unit from the hot aisle.

There should be a minimum 1.2 m (4 ft) aisle space between the ITE row and any wall or other non-IT equipment located around the perimeter of the computer room. It is also recommended to have one or two of the perimeter sides provided with 1.8 m (6 ft) aisle clearance to move CRAC/CRAH or large-frame ITE in or out of the computer room.

9.5.2 Power Distribution

PDUs are the electrical components that transform the voltage from the building distribution voltage levels down to the ITE voltage level (208 V, 240 V). It is recommended that the PDUs be located outside of the computer room.

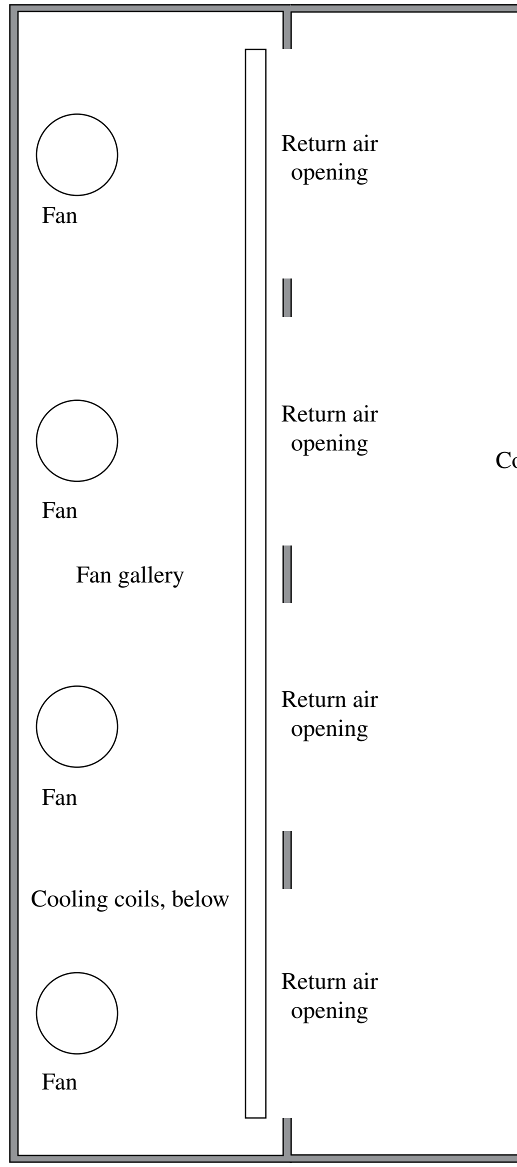


FIGURE 9.8 Example of Fan Gallery (plan view). Courtesy of Isaak Technologies.

Remote Power Panels (RPPs) are the electrical component that provides higher quantity of power circuit pole positions in a high-density frame compared to standard wall-mounted panels. The RPPs are downstream from the PDU and feed the POUs within the IT equipment racks and cabinets. POUs may also be referred to as power strips. The RPPs are recommended to be placed at one or both ends of the rows of IT equipment, depending on the level of redundancy required.

RPPs are typically made up of four 42 pole panels. The entire panel can be fed from one upstream breaker, an individual breaker feeding each panel, or any combination in

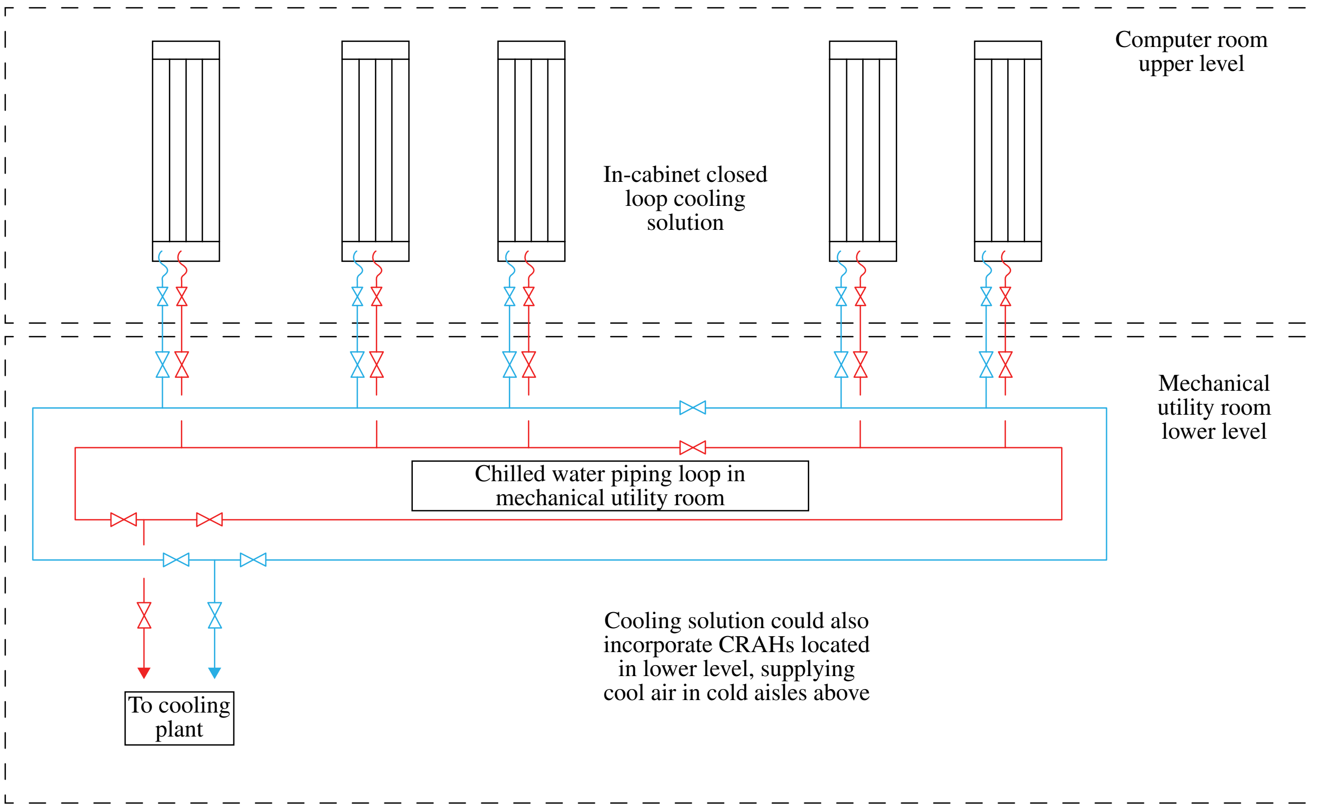


FIGURE 9.9 Example of multistory data center. Courtesy of Isaak Technologies.

between. It is possible to feed two of the panels from power source “A” and two panels from power source “B.”

For data centers designed to meet ANSI/BICSI Class F2 or lower, one RPP at one end of a row of IT equipment will meet the design standards. For Classes F3 and F4, in order to meet the minimum design standards, an RRP will be placed at both ends of each row of IT equipment.

Placing the RPPs at the ends of the rows reduces coordination with other systems compared to placing them against the outside wall. When they are placed at the ends of the row, the power whips feeding the racks or cabinets within the row are contained in the floor grid directly under the racks or cabinets.

When the RPPs are placed against the outside wall, the power whips have to transition across the perimeter aisle to the row of racks and cabinets. If this method is used, the power whip installation needs to be coordinated with, when used, other underfloor pathways running perpendicular to the ITE rows, chilled water, or refrigerant lines.

Overhead power busway may also be used to distribute power to the ITE racks and cabinets. Typically, RPPs are not used when overhead busway are implemented. The overhead busway design provides flexibility for the IT designer in that the exact power circuit feeding each rack or cabinet can be easily and quickly changed by inserting a plug-in unit that has the correct breaker and receptacle configuration required. Overhead busway need to be coordinated with all other overhead systems such as sprinkler head locations, network cabling pathways, and lighting. If there are two overhead busway providing an “A” and a “B” power source, the position in either the horizontal or vertical plane needs to provide sufficient separation to be able to insert the plug-in units without conflicting with the other busway or plug-in units.

Power cables entering cabinets, either overhead or through a floor tile below, should be sealed with an appropriate grommet. The grommets need to be sized to provide sufficient space to pass the diameter of one POU cord plus the diameter of one POU cord end cap through the grommet.

POUs are available in either horizontally or vertically mounted models. Vertical, referred to as zero-U, models are typically used in server cabinets. Vertically mounted POUs may not be the preferred model for network cabinets as the air dam kits required to support ITE with side-to-side airflow may restrict the placement of the POUs on the exhaust side of the cabinet. This is cabinet manufacturer and model dependent. For this reason, horizontal POUs are sometimes used for all network cabinets. Horizontal POUs are also often used for open two- or four-post racks as the racks typically do not have options for button holes required for the vertical POU mounting.

9.5.3 Sprinkler and Fire Protection Systems

The local authority having jurisdiction (AHJ) will define if sprinkler or fire protection systems are required below the raised floor or above a suspended ceiling in addition to within the computer room space.

Sprinkler and fire protection systems should be mounted the highest when required under a raised floor or above a suspended ceiling, with all other pathways and systems mounted below the fire protection systems. Fire detection devices can be mounted vertically if specifically designed and approved for vertical mounting applications.

Sprinkler head placement may be a critical coordination issue for computer rooms with less than 3 m (10 ft) ceiling height, especially when overhead power or network cabling is used. Sprinkler heads typically require 450 mm (18 in.) clearance below the sprinkler head; local AHJ may have greater restrictions.

The sprinkler and fire protection systems coordination challenges are often eliminated if ceiling heights of 4.2 m (14 ft) or higher are implemented within the computer room.

9.5.4 Lighting Fixtures

When a suspended ceiling is used, lighting fixtures are generally inserted in the ceiling grid system. This method requires close coordination with the ITE rows, overhead pathways, and sprinkler and fire protection devices.

Computer rooms with higher ceilings may implement indirect lighting by using suspended light fixtures with most of the light directed up and reflected off the ceiling (painted white) to provide sufficient light within the room. This method provides a more even distribution of light throughout the room, with less shadowing compared to light fixtures inserted in the ceiling grid. However, the amount of lumen output required to provide sufficient light may exceed local energy codes. As more options for LED lighting become available, the indirect lighting method using LED lamps may be a viable consideration. When using an indirect lighting system, it is recommended to have the suspended light fixtures installed above any other suspended systems such as power or network cabling pathways. This will minimize the risk of lamps breaking when technicians work on systems above the light fixtures.

9.5.5 Raised Floor versus Nonraised Floor

Building constraints in floor-to-deck heights may restrict the design from incorporating a raised floor. Incorporating a data center within an existing building may also restrict the ability to have a depressed slab for the computer room space. It is always desirable to have the computer room floor at the same elevation the entire route from the adjacent corridor space to the loading dock. Ramps to accommodate a change in floor elevation between the computer room and the adjacent corridor are not only a functional annoyance, but they also require additional footprint within the computer room. The recommended slope for the ramp is 4.8 degrees, a rise of 1:12. For a 600 mm (24 in.) raised floor, this would result in a 7.2 m (24 ft) long ramp. For a large computer room, this will not be significant, but for a small data center, this may significantly reduce the space available for ITE.

It is sometimes stated that if the data center cooling system does not use the underfloor space for air distribution, then a raised floor is not required. This is only considering one aspect of the raised floor when others exist. A raised floor environment provides flexibility to accommodate future ITE technology requirements that are not within the initial design criteria. This could include direct water-cooled ITE, where the water lines would be preferred to be routed below the ITE versus overhead. Another example could be fan-less servers, which require air passing through the cabinet from the underfloor space up through the vertical exhaust ducts to the return plenum. Both of these examples, although available today, are not commonly used throughout a data center.

If a raised floor is not incorporated into the design, coordination issues are introduced. The installation and placement of drains for CRAC/CRAH condensate lines must be provided, along with sufficient floor slope to keep any moisture accumulation away from ITE.

9.5.6 Aisle Containment

The design of aisle containment systems should always be reviewed with the local AHJ over fire protection systems. Containment systems either have fire detection and suppression systems integrated within the contained space or have the containment panels automatically removed, without impeding the egress pathway, upon detection of heat or smoke. The local AHJ may place certain constraints on how a containment system is integrated into the overall fire protection plan.

Lighting fixture type and placement will need to be coordinated with the aisle containment system to ensure that sufficient light levels are provided within the containment spaces.

Overhead or underfloor power and network cabling pathways can be easily incorporated into either a hot aisle or cold aisle containment system. The containment system itself does not introduce any new coordination challenges.

Vertical heat collars are technically not aisle containment, but rather a contained vertical exhaust duct. Vertical heat collars do introduce additional coordination challenges in that up to half of the space available on top of the cabinets is consumed by the vertical duct and therefore no longer available for routing the power or network cabling pathways. Additional coordination is required when incorporating overhead network cabling pathways and overhead power distribution busway together with vertical heat collars. All of these systems need to fit within the limited space over the cabinets.

9.6 COMPUTER ROOM DESIGN

9.6.1 By Size

For the purposes of our discussion, we will define small computer rooms as less than 280 m^2 (3,000 ft^2), medium computer rooms as less than 930 m^2 (10,000 ft^2), and computer room with more space as large. These parameters are certainly not a standard within the industry but are simply used to guide the discussion on the design nuances between different sizes.

9.6.1.1 Large Large data centers will require columns to support the roof structure or have the computer room space compartmentalized into multiple smaller rooms. The location of the columns should be coordinated between the structural engineer and the IT designer to minimize the interference of the column on the ITE layout.

Large data centers may also require additional network frames distributed throughout the computer room space to support distribution network switches. This will be dependent on the network architecture and topology that is deployed in the data center.

9.6.1.2 Medium Medium data center may require columns to support the roof structure. However, the designer should identify solutions that are available to avoid or minimize the quantity of columns. Columns are a coordination challenge with the initial ITE layout and all future technology refreshes.

Many options exist for medium data centers with respect to network architecture, topology, and the associated cabling infrastructure. This could include centralized, distributed, zone, top-of-rack configurations. Space to support additional network distribution frames throughout the computer room will likely not be required.

9.6.1.3 Small Small data center can be the most challenging to coordinate the ITE with all the other systems. They often push the design to high-density solutions as owners are trying to compress as much processing capability within a small defined footprint.

Columns within the computer room should be avoided.

Small data centers will typically have a centralized network core.

9.6.2 By Type

9.6.2.1 In-House Single-Platform Data Center Organizations that own and manage their own data centers with single- or minimal platform variations required to be supported can have a consistent repeatable ITE layout. Organizations that have all computer processing on rack-mountable appliance or blade servers and all storage on rack-mountable disk arrays have the capability to plan their ITE layout using consistent zones of cabinets.

The computer zones can be based on the standard cabinet width and depth and standard aisle spacing. As an example, if the standard cabinet is 800 mm (31 in.) wide and 1.2 m (48 in.) deep, the computer room layout will consist of repeatable rows with 1.2 m (4 ft) hot and cold aisles between the ITE cabinets. Since all platforms are rack mountable in standard ITE cabinets, it is not necessary to know exactly where each system, whether network, appliance server, blade server, or storage disk array, when developing the initial ITE floor plan layout. If the power distribution from the UPS downstream to the ITE cabinets is designed according to the ANSI/BICSI 002 standard, there should be sufficient flexibility in the capacity of the PDUs and RPPs to support each zone independent of the specific platform within each zone.

Since all the ITE systems are consistently installed in standard cabinets, the power and network cabling pathway design can also be consistently applied across the computer room space (Fig. 9.10).

9.6.2.2 In-House Multiplatform Data Center Organizations that own and manage their own data centers but have numerous platforms may require unique zones to support each platform type. Unique platform types may consist of rack-mountable servers (appliance or blade), large-frame computer processing (mainframes, HPC, and supercomputers), large-frame disk arrays, or carrier-class network platforms (580 mm [23 in.] rails).

The computer room ITE layout will need to identify each unique zone and the size of each zone. The placement of the power distribution (RPPs), power pathways, and network pathways will all need to be coordinated with the unique requirements of each zone (platform). Each zone will need to have sufficient spare capacity to accommodate the anticipated growth and the technology refresh requirements.

Large-frame systems can have ITE with depths up to 3.6 m (72 in.) and varying frame heights and widths. Systems such as tape libraries will have even larger footprints. Large-frame systems have various airflow patterns through the equipment for cooling such as front to back, side to side, and bottom to top that will be unique to the platform, manufacturer, and model of platform. Large-frame systems have various power or network cable entry points, often accessible from the bottom only. The cable entry points may have very little tolerance as to the placement of a floor tile grommet below the specific frame. These unique characteristics of large-frame systems require extra attention to the design details and coordination with the supporting systems.

Since large-frame systems typically have their power and network cable entry points at the bottom of the systems, the preferred power and network cabling pathway location supporting these systems may be under the raised floor. It is common to have overhead pathways for rack-mountable systems in standard cabinets and underfloor pathways for the large-frame systems within the same computer room (Fig. 9.11).

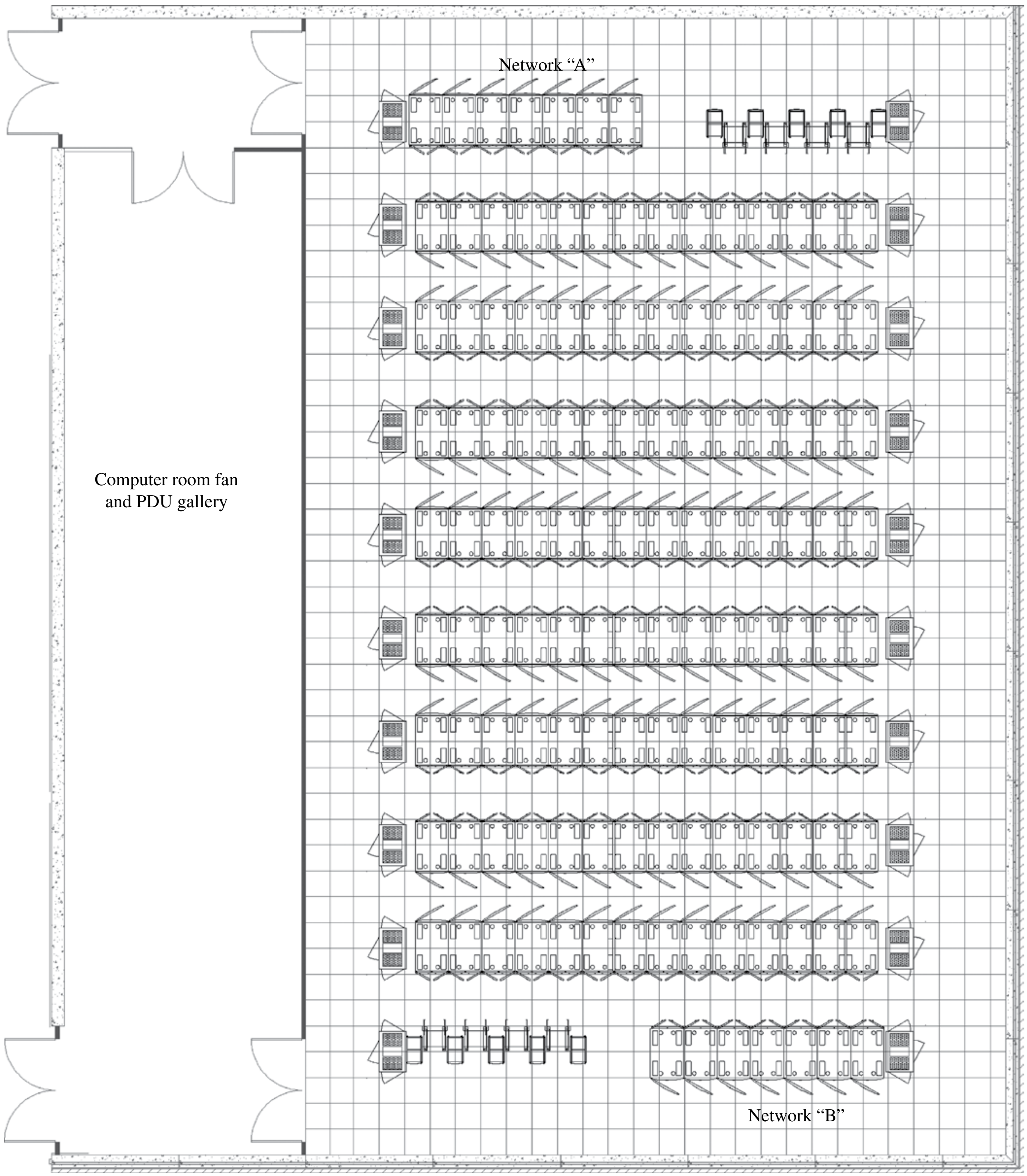


FIGURE 9.10 Example of computer room layout with all IT platforms mounted in standard cabinet—all equipment in 4 ft zones. Courtesy of Isaak Technologies.

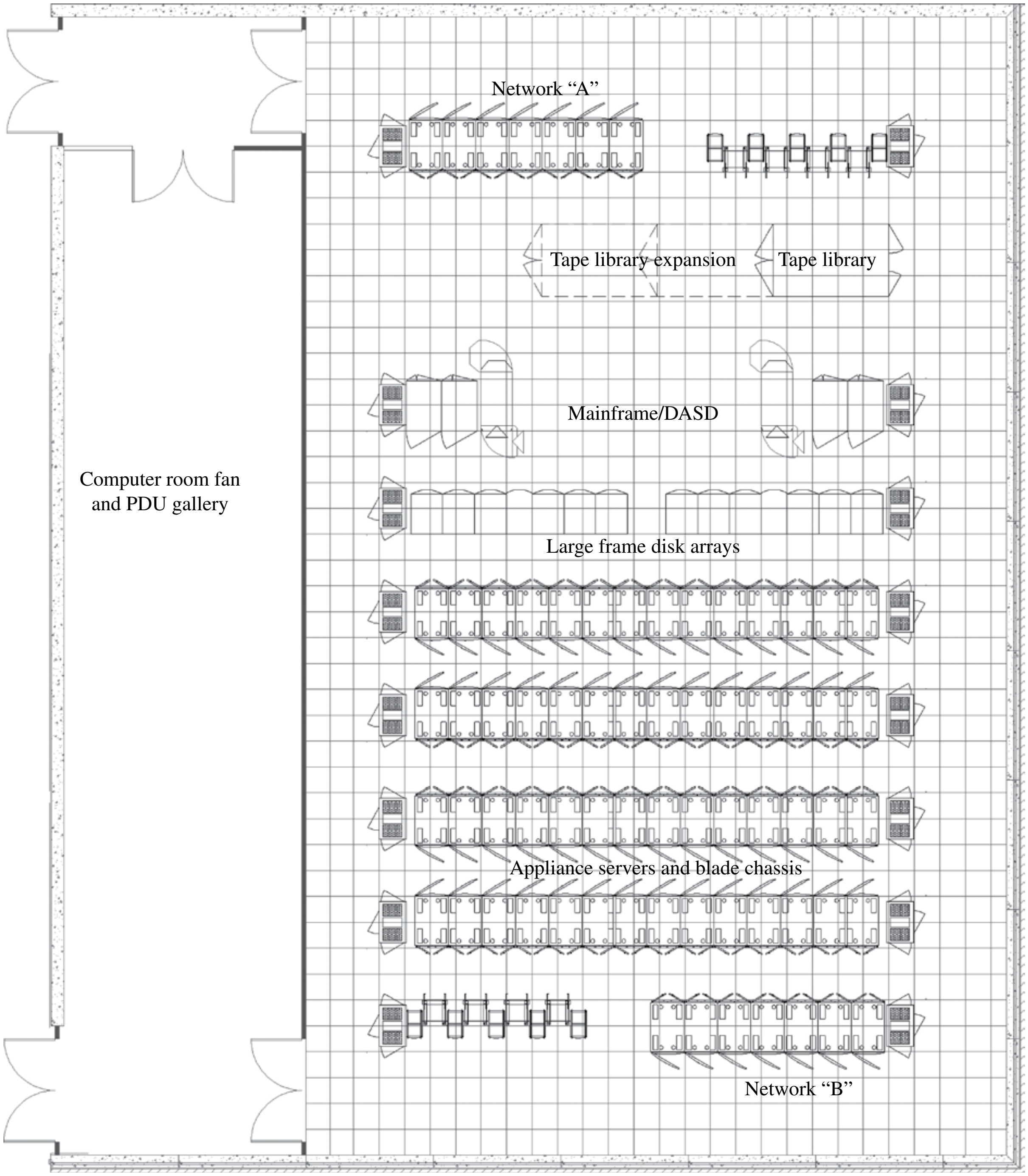


FIGURE 9.11 Example of multiplatform computer room layout. Courtesy of Isaak Technologies.

9.6.2.3 Outsourced Services Data Center Colocation data centers consist of organizations that own data centers and manage the space, power, and cooling infrastructure to support their customer's platforms placed in either caged spaces or cabinets.

This type of data center requires a different approach to defining the space, power, and cooling capacity requirements. The owner does not know exactly what the ITE layout will look like until customers have committed to their services and defined the systems they will be placing within the colocation data center. This information is not known at the time the colocation owner is planning and designing the data center. Therefore, colocation data center design drivers typically are cost control, flexibility, and scalability.

Cost control is required to ensure that the level of reliability, redundancy, and the services provided by the data

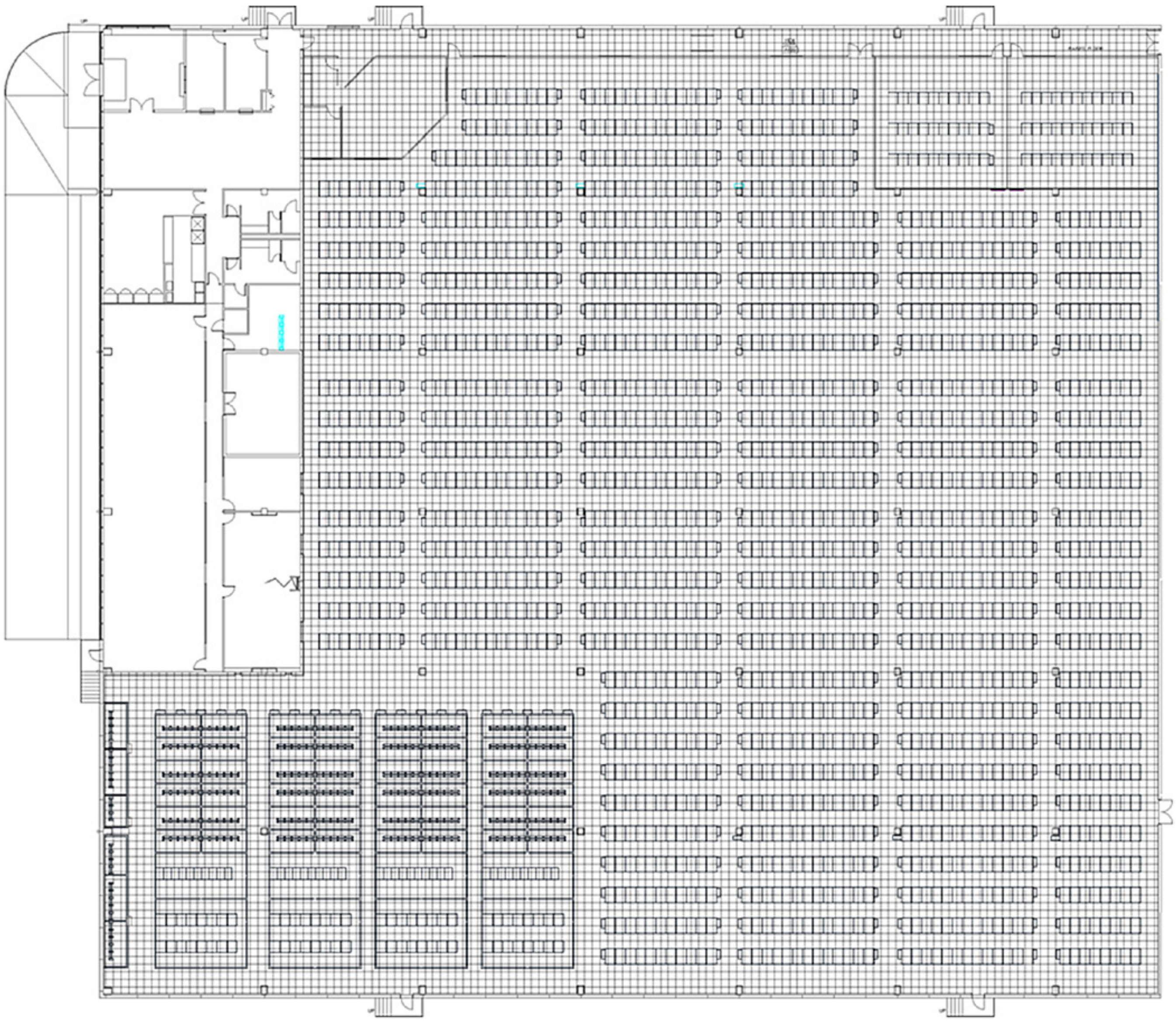


FIGURE 9.12 Example of colocation data center with customer caged space and customer cabinet layout—lease by cabinet or caged space. Courtesy of Isaak Technologies.

center are in alignment with the potential customer's requirements and price point. Flexibility is required as the capacity requirements of each individual caged space or cabinet will vary significantly over the life of the data center, as customer's technology changes, or as customers move in and out of the data center. Scalability is required to enable the colocation data center capacity (space, power, cooling) to be built out as customer demand requires.

Colocation data centers typically provide space, power, cooling, and connectivity to network access providers. The colocation owner typically does not manage the network, but rather simply provides a handoff of the service providers circuit at the entrance room to the customer's equipment located in a caged space or cabinet. There are distance limitations on various circuits provided by the network service providers. For large colocation data centers, there may be a requirement to have multiple entrance rooms so that T-1/E-1 or T-3/E-3 circuits can be extended to the customer's space without exceeding distance limitations. The multiple entrance rooms do not provide any redundant capabilities in this scenario (Fig. 9.12).

9.7 MODULAR DESIGN

Incorporating a modular approach to a data center design is generally always applied. The exception to this is small data centers where the ultimate power and cooling capacity is not less than 40% of a single module or component. Modular design needs to address space, power, cooling, and network capacity.

A critical consideration when incorporating a modular approach is that the design must be able to support future expansion without reducing the level of redundancy of the critical systems. Implementing future expansion must also be accomplished without disrupting the normal IT operations within the data center.

9.7.1 Computer Room Space

Of all the facility-related aspects that are affected by modular designs, space is the one that often impacts cost the least. The total cost of data center facilities is generally comprised of 30% of the building shell and interior build-out and 70% in the electrical and mechanical systems (land and IT systems not included). This ratio will fluctuate based on the level of redundancy required and the size of the computer room space. Since the building represents the smaller portion of the total facility costs, it is a common approach to build out two or three times the initial required floor space to accommodate future growth.

It is also common to plan for the expansion of space to accommodate future growth. This can be accomplished by constructing additional computer rooms adjacent to the initial building or incorporating knockout panels in the computer room perimeter wall, which can be removed in the future.

9.7.2 Power and Cooling Infrastructure

Incorporating a modular design in the power and cooling systems is a standard approach to use. It is very common for new data centers to have the initial power and cooling capacity less than 50% of the ultimate capacity design.

The initial build-out of power capacity must have the power distribution coordinated with the ITE layout. It is more practical to build out the computer room from one end and systematically expand across the computer room space. This allows the initial electrical distribution (PDUs and RPPs) and cooling equipment to provide capacity to the initial zone of ITE. Future PDUs, RPPs, and CRAC/CRAH units will be added in the adjacent ITE zones as additional capacity is required.

It is critical that the future PDUs, RPPs, and CRAC/CRAH units can be added without disrupting the systems initially installed. It is recommended that the installation of future PDUs not require a shutdown of any upstream distribution that is feeding “hot” PDUs and the installation of future RPPs not require a shutdown of any upstream PDUs that are feeding “hot” RPPs.

9.7.3 Network Capacity

Incorporating a modular design for the network not only addresses capacity but also the physical location of the entrance room within the data center.

9.7.4 Scalability versus Reliability

Data center operators often desire to have reliable and scalable solutions; however, these are fundamentally opposing criteria. Scalability requires smaller capacity components, in greater quantity to make up the ultimate design capacity (i.e., seven 500 kVA UPS modules vs. five 750 kVA UPS modules):

- The 500 kVA UPS module example can scale in five 500 kVA increments from 500 to 3000 kVA (assuming redundancy is required for the UPS modules). If each module had a reliability value of 80% over a defined period, in an $N+1$ configuration, the seven 500 kVA UPS module example would have a system reliability of 85.2%.
- The 750 kVA UPS module example can scale in four 750 kVA increments from 750 to 3000 kVA (assuming redundancy is required for the UPS modules). If each module had a reliability value of 80% over a defined period, in an $N+1$ configuration, the five 750 kVA UPS module example would have a system reliability of 88.2%.

Increasing scalability inherently decreases reliability. The designer of any system, whether it is for power distribution, cooling distribution, or network architecture, must balance scalability and reliability to ensure that the appropriately sized building blocks are selected for the initial design and for the future incremental increases in capacity.

9.8 CFD MODELING

Computational Fluid Dynamics (CFD) is a method of modeling the effectiveness of the cooling system and its ability to meet the demand of the ITE being supported. In order to conduct a CFD analysis, the computer room space must be modeled, including the room dimensions, the placement of heat producing equipment within the room, the placement and type of the supply cooling (CRAC/CRAH), the placement and type of perforated floor tiles, all openings within the floor tile system, and any obstructions to the airflow (pipes, cable trays, etc.).

The output of a CFD analysis will model the temperature and pressure variations throughout the computer room space (three-dimensional). This has proven valuable in data center design as the designer can validate the cooling system design prior to installing the system. It is also beneficial to data center operators as they can:

- Model how the placement of future ITE will impact the cooling systems ability to meet the computer room demand
- Simulate various failure scenarios by "turning off" components within the CFD model and analyzing if the remaining cooling system can support the ITE load

There are a few vendors that have developed the CFD software tools, with varying degrees of accuracy, level of modeling complexity, and cost.

9.9 DATA CENTER SPACE PLANNING

9.9.1 Circulation

The data center must support the replacement of all ITE and power and cooling system components by providing adequate clearances from the loading dock to the computer room, and electrical and mechanical rooms. Corridors should be at least 2.7 m (9 ft) high. Doors should be a minimum of 2.4 m (8 ft) high and 1.1 m (3.67 ft) wide for single doors or 1.8 m (6 ft) wide for a pair of doors. Consideration for corridors with higher ceilings and 2.7 m (9 ft) high doors should be made since a packaged standard 42 RU cabinet on a pallet jack typically does not fit under a 2.4 m (8 ft) high door.

The data center layout should be defined into various access types such as noncritical, critical facilities, and critical IT. It is recommended to minimize personnel traffic between these zones as much as possible, keeping facility personnel out of IT spaces and IT personnel out of facility spaces.

9.9.2 Support Spaces

Any function that is required in supporting the IT systems within the computer room is considered part of the data center. Functions that are not directly required to support the IT systems within the computer room are considered to be outside the data center.

The following critical spaces are required to support the IT systems within the computer room.

9.9.2.1 Entrance Room The functions of the entrance room are to provide a secure point where entering network outside cable plant from access providers can be transitioned from outdoor cable to indoor cable and to house the access provider-owned equipment such as their demarcation, termination, and provisioning equipment.

The entrance room should be located adjacent to, or in close proximity to, the computer room. The pathways from the entrance room to the computer room should not transition through any nonsecure spaces. The entrance room should also be located in close proximity to the electrical room where the main building ground busbar is located in order to minimize the length of the bonding conductor for telecommunications.

For data centers with redundancy requirements, a second entrance room is recommended to provide physical separation between redundant access provider services. These entrance rooms should be located at opposite ends of the computer room from each other.

The entrance room often houses multiple network service providers. The configuration of the entrance room should be coordinated with each network service provider to ensure that their requirements are met and that all clearance requirements and special security concerns are understood.

9.9.2.2 Network Operations Room The network operations room or network operations center (NOC) supports IT operations. The NOC has technicians within this room monitoring the network and IT system operations, typically on a 24/7 basis.

The NOC is typically located adjacent to the computer room with an entry door into the computer room. This can act as another level of security in that everyone that enters the computer room would gain entry through the NOC, enabling the NOC personal to physically see each individual accessing the computer room.

Since the NOC provides 24/7 operations, personal comfort is a driving design criteria to ensure that technicians are alert and can easily access the critical information. This influences the type of furniture selected, the multiunit display systems, and possibly some level of natural lighting provided.

Even though the roles of the technicians within the NOC are primarily IT related, it is recommended that the building management systems (BMS) have monitoring capability within the NOC as well. This will enable the technicians to have an understanding of the building systems status in real time. The BMS should not have control functionality within the NOC.

9.9.2.3 Entry Way The entrance into the data center should have a physical security station to monitor and control all access to the facility. Visitors and outside vendors should have to sign in and verify the need for them to gain access to the computer room. No access to critical spaces should be allowed without proper authorization past the main entrance into the data center.

9.9.2.4 Support Staff Support staff that directly manages the daily operations of the data center will have their offices or work space within the data center space. Data Center support staff may consist of the following:

- Data center manager
- Data center facility manager
- Data center facility engineers and technicians

- Data center shipping/receiving clerk
- Data center security
- NOC personnel

IT network or system engineers and administrators are not necessarily located within the data center. The IT personnel may be located off-site from the data center with remote access capability.

9.9.2.5 Electrical Room The electrical rooms should be located adjacent or in close proximity to the computer room to minimize the lengths of copper feeders from the electrical distribution to the ITE within the computer room. There are significant quantities of power circuits feeding the ITE, and minimizing the feeder lengths helps to reduce installation costs.

The size of the electrical room is directly related to the ultimate design capacity and the level or redundancy of the electrical distribution. When redundant electrical distribution is required, it is recommended that these rooms be positioned within the data center with as much physical separation as possible to reduce common modes of failure.

9.9.2.6 Battery Room Data centers that utilize battery-based UPS systems are recommended to have dedicated battery rooms. Wet cell batteries require dedicated battery rooms with special ventilation requirements to meet building codes. Other battery technologies may also require dedicated battery rooms and/or special ventilation depending on the total quantity of battery acid within the battery system or local building codes.

9.9.2.7 Mechanical Room The mechanical equipment room requirements vary depending on the type of cooling technology used. The water-based cooling system that incorporates a chiller system requires sufficient space for the chillers, pumps, and piping. The mechanical equipment room should be in close proximity to the computer room to minimize the routing of piping through nonmechanical spaces between the mechanical room and the computer room.

9.9.2.8 Storage Room Data Centers have a need for storage rooms to support two different functions. A storage room is required for facility-related spare parts. The spare parts that should be on hand include belts, filters, and other general maintenance-related items. A storage room is also required for IT systems, which include temporary placement of high-value equipment prior to deployment in the computer room, spare network line cards, interface cards, network modules, optical interfaces, power supplies, and critical components with higher failure rates.

Secure storage may be required for vendor storage. Vendors that are supporting IT platforms within the computer room with defined SLAs may need to store critical components on-site in order to meet the terms of the SLAs. Even though these spare parts are stored on-site, they are still in the vendor's inventory until such time as they are required to be installed in the owner's IT systems. Therefore, the vendor may require a secure storage space to ensure that their high-value components are securely stored. This vendor storage may not need to be a dedicated room, but simply a secured space, closet, or shelving within a larger storage area.

9.9.2.9 Loading Dock/Receiving The loading dock should provide protection from the elements so that the delivery of high-value equipment is not exposed to rain or snow during the receiving of a shipment. It is recommended that the loading dock have a secured entry between the loading dock and the rest of the data center space to ensure that only authorized personnel can gain access from the loading dock to the rest of the data center. The loading dock should be sized so that there is sufficient space to temporarily house all the equipment from the largest anticipated delivery at one time. Once the high-value equipment is received and the loading dock overhead door is closed and secure, the equipment should be moved into an adjacent secure staging space.

The staging space is where the packaging will be removed from the equipment. All packaging material should be placed in waste containers, helping to ensure that cardboard dust does not enter the rest of the data center facility.

It is recommended that the route from the loading dock to the staging space, burn-in room, equipment repair room, and the computer room have the floor at the same elevation. There will be several technology refresh occurrences throughout the life of the data center facility. Each refresh requires the delivery of new equipment and legacy equipment being shipped out; therefore, it is preferred that no ramps or changes in elevation be required as this introduces risk and increases the difficulty when delivering the high-value equipment.

9.9.2.10 Burn-In/Equipment Repair A burn-in or equipment repair room is recommended so that the IT equipment can be initially powered on and tested prior to being placed inside the computer room. This ensures that the equipment is not defective or will cause a short circuit within the critical computer room space. A separate dedicated UPS system should be considered for the burn-in and equipment repair room to ensure that a burn-in process is not disrupted due to a power utility outage. The UPS circuits for the burn-in or equipment repair room should not be fed from the main computer room UPS.

The burn-in or equipment repair room may be incorporated together with the storage room depending on internal operating procedures. The combined storage, burn-in, and equipment repair room would need to provide sufficient space to support all these functions.

9.9.2.11 Security The security space requirements include space for the security personnel to monitor and control building access and a space to support the security systems.

The security personnel space should be at the main entrance into the data center to control access into the building.

The security system space can be a dedicated secure room with ITE racks or cabinets housing the security systems such as access control and CCTV monitoring. The security systems are critical to the data center operations, and as such, the power circuits should be fed from a UPS source.

Other critical building systems that require rack-mounted systems that are not managed by the IT department may also be placed within the security systems room. Other building systems may include servers supporting the HVAC control systems or the building management systems.

9.10 CONCLUSION

The data center is a complex combination of facility systems and IT systems working together to support the critical business applications. These systems do not function in isolation from each other and should be designed with a methodical coordinated approach. A design or operational change in one system can have a cascading effect on numerous other systems.

A data center project begins with understanding the IT applications and supporting IT platforms. The process continues with coordinating the facility requirements, the IT network architecture and topology, and the computer room layout of IT and non-IT equipment and is completed with the business applications migrating to the new platforms supported by all critical data center infrastructure.

FURTHER READING

ANSI/BICSI 002-2011. Data Center Design and Implementation Best Practices standard.

ANSI/NECA/BICSI 607. Telecommunications Bonding and Grounding Planning and Installation Methods for Commercial Buildings; 2010.

ANSI/TIA-942-A. Telecommunications Infrastructure Standard for Data Centers; 2010.

IEEE 1100-2005. *The IEEE Emerald Book*, Recommended Practice for Powering and Grounding Electronic Equipment.

NFPA 75. Standard for the Protection of Information Technology Equipment; 2009.

NFPA 1600. Standard on Disaster/Emergency Management Business Continuity Programs; 2007.

UL 60950-1 2003. Information Technology Equipment—Safety—Part 1: General Requirements.

10

MECHANICAL DESIGN IN DATA CENTERS

John Weale

The Integral Group, Oakland, CA, USA

10.1 INTRODUCTION

Data center mechanical design is not inherently complex, but the requirement for high reliability combined with very obvious (and expensive) failure if it is not met adds a degree of challenge not seen in common mechanical design. Against this high-stakes design background, traditional design has leaned heavily on repeating proven legacy designs—often at the expense of innovation that can improve reliability, flexibility, cost, operating efficiency, and other aspects of design quality. The objective of this chapter is to acquaint a mechanical designer with data center design and give them the technical grounding required to move beyond replication of proven, yet often obsolete, designs and into creating optimized solutions that meet the unique requirements of their clients.

The best mechanical designs for data centers show not just skill in system design but also a clear understanding of the fundamental purpose of a data center: to make money. A careful investigation of the design criteria and consideration of their impact on the design help to best serve the actual needs of the client. But, surprisingly, this is often not done. The reliance on reusing old, "proven" designs is often used to justify doing only a cursory investigation of the needs of the current client. Some level of assumption is required to maintain the flexibility to accommodate future, unknown IT equipment requirements, but the needs of the initial IT equipment set and operations should be evaluated for each individual project.

The system configurations and equipment used in data center design should be familiar to experienced mechanical engineers, but there are a number of specializations made to adapt them to the needs of data centers. Equipment is configured to provide the high reliability required by the data center, in addition to serving the sensible-only nature of the dominant internal loads. System configurations are designed to accommodate the point source loads of IT equipment, with various approaches used to provide cool air to the intakes while reducing recirculation of the hot exhaust.

How the design process for data centers fits into the traditional design stages and milestones is discussed at length. Consistent communication with the design team and owner is important in any project, but the high costs and critical nature of mechanical systems for data centers increase the need for clear and direct communication.

With a strong grounding in the equipment, system configuration, and design process used for data centers, a discussion of current best practices offers a springboard into the final subject—future trends— which is not a conclusion to the chapter but rather where the dialogue is handed off to the dynamic world of practice.

10.2 KEY DESIGN CRITERIA

There is no single type of data center. While it is feasible to design a generic "standard" data center based on common assumptions alone, the best balance of flexibility and first cost requires careful collection and evaluation of specific client requirements including the following.

10.2.1 Reliability

High reliability is a, and often the, critical criterion of data center design. As it has a wide impact on the mechanical design, the level of reliability is defined early in the process.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

It is common that a redundancy level such as $N+1$ (no single equipment failure will result in any loss of capacity) will be an explicit part of the design criteria. Sometimes, a standard or guideline will be referenced to define the reliability and redundancy requirements, or there will be an insurance company requirement document or internal client standard that must be met.

It is important for the mechanical engineer to fully understand the reliability requirement and explicitly state how it will be met—particularly if cost is driving an aggressive interpretation. For example, a common area of interpretation and compromise is on how an $N + 1$ reliability requirement impacts chilled water piping design. Some clients will require two independent chilled water loops for the absolute highest redundancy (a burst pipe does not bring down the cooling system), while others accept a single piping system with parallel paths and valves to allow any segment to be bypassed (planned piping work for repair of slow leaks and planned maintenance can be performed without interrupting the system operation). These two approaches offer different operational capabilities and first costs.

This question of whether chilled water piping requires redundancy or not illustrates where the standard data center approach to reliability—requiring a constant redundancy for all components—is a gross simplification: the probability of a pipe failing is far, far less than the probability of a complex device such as a chiller failing, yet it may be given the exact same redundancy design requirement ($N+1$, $2N$, etc.). Yet the alternative, a detailed analysis of the actual probability of failure based upon the component probability of failures, is simply not done as a normal part of mechanical design.

The specific reliability requirements have a large impact on system cost and determine if the final product meets the client's needs or is a catastrophic failure. Fully understanding all the details of the redundancy requirement and communicating all of the implications of it to the client in a clear documented manner is an important task that spans all design phases. It is usually not a difficult-to-define issue, but due to its wide-reaching impact, it is not an aspect that should be rushed through with any assumptions.

The project location can have a significant impact on design for reliability. Designing for tornado or hurricane resistance can have a large impact on mechanical design, often leading to hardening approaches ranging from bunkered dry coolers to cooling towers protected behind (very low free area) armored louvers. Client concerns over local forest fires or even objectionable odors from industrial neighbors can limit the value of air-side economization. Local water consumption restrictions during drought years or code inspectors' proclivity to shut down equipment feared to harbor excessive legionella (particularly in Europe) may present nonnegligible failure modes that influence design.

The importance of the *appearance* of reliability can be missed by mechanical engineers. Data centers that serve external clients (colocation, or colo, facilities) place a high value on a marketable design. Due to the high demand for reliable and abundant cooling, the mechanical system often is a key part of the sales pitch. Even for owner-occupied data centers, a failure can prove very costly for a company, and nontechnical executives are often quite interested in the reliability. These critical clients are not trained engineers; the system design needs to not only be highly reliable in fact but also easily assure all the nonengineer stakeholders—from the marketing manager to the chief financial officer—that it is highly reliable. This can be a powerful driver toward using legacy design approaches, even when such legacy designs are not the technically most reliable option. The best designs appreciate the importance of appearances without compromising on providing on optimized design. Close coordination with the architect and client is needed to ensure this common soft, but critical, design requirement is defined and met.

10.2.2 Security

Another common characteristic of data centers is security. A visible aspect of reliability, security measures can become closely intertwined with marketing. Security requirements are usually relatively simple design parameters to accommodate as long as they are identified during the appropriate phase of schematic design. Adding the equivalent of security bars to an exterior air economizer's louver bank during construction can be an expensive and embarrassing consequence of neglecting security concerns. Space pressurization control, particularly when air-side economizer is implemented, can also be a major security issue if overpressurization results in doors not closing properly.

It is not unusual that large data centers desire an anonymous curb presence that does not advertise the nature of the facility's use. Architecturally, this means an exterior treatment without signage. Typically, this only impacts the architect and signing scopes, but in some cases, it may dictate the placement and screening of exterior equipment.

10.2.3 Safety

Unlike some critical facility designs, such as chemical laboratories, there are few opportunities for the mechanical engineer to kill people by the particulars of their data center design. Fire control systems, including dry extinguishing systems that use a gas to smother fires, are one area with serious life safety implications. Exiting requirements can also trip up some air management schemes that may drop curtains or have other obstructions in exit paths during a fire event. Close attention should be paid to the fire code requirements,[1] which are continuing to evolve to catch up to the current state of data center design.

Worker productivity can suffer as high-density data center facilities come to resemble industrial facilities. Very effective air management designs can result in some portions of the data center operating at hot exhaust air temperatures in excess of 95°F (35°C). Often, the high-temperature heat exhaust paths correspond to spaces that are occasionally occupied by workers installing cabling or performing other tasks. Design consideration needs to be given to accommodating these workers (and meeting applicable OSHA codes). Noise limits can also be a concern, although the operational accommodation for a high noise space is simply ear plugs, while an excessively hot space may require operationally intrusive frequent mandatory worker breaks.

10.2.4 Aesthetics

As mechanical design approaches evolve, the interior appearance of data centers can vary dramatically from the traditional. This is usually a problem for clients. Data centers that need to attract rent-paying occupants want tours of the facility to immediately project an image of a traditional—read highly reliable—data center. Internal data centers, being a large investment with banks of high-tech equipment and intriguingly blinking lights, are also popular tour material for executives wishing to show off their organization's technical prowess. While these groups are rarely at the design table, their desires are ignored at the designer's peril.

The client's expectation for the space appearance can be difficult to define for several reasons. A highly placed executive rarely sits at the table during schematic design but can force an 11th-hour design change by a brusque condemnation of the appearance of hanging curtains for containment after seeing a design rendering. Or a concerned question about how it can be a data center if it does not have a raised floor from a trusted facility manager can exert pressure to completely redesign the airflow system. Sometimes, clients delay raising concerns due to some embarrassment about bringing up appearances during the early-stage design discussions filled with concerns about important technical issues like kW capacity, tons of cooling, redundancy, and tiers—but that early design stage is exactly where concerns about appearance should be raised. Good visual communication, ranging from renderings to rough sketches, dealing specifically with the system's appearance during initial selection is important to keep aesthetics from upsetting the design process.

10.2.5 Flexibility

The primary data center load, that is, the IT equipment the system is supporting, typically passes into obsolescence and is replaced in whole or piecemeal on a 3–5-year cycle. Beyond the standard need for the mechanical system to be able to support changes in load size and physical location in the space, changes in infrastructure requirements ranging from the need for liquid cooling to the equipment airflow requirements may need to be considered. These future changes typically need to be accommodated while the balance of the data center is in full operation, increasing the need for the design to provide appropriate access to installed components and consideration of system expansion in future.

The need for flexibility can be a challenge to distribution design. In an air-based system, having excess ducting capacity provides for future flexibility and can yield day-one efficiency benefits through lower pressure drop if the fans are designed to turn down. Likewise, oversized piping is also prudent. It is not unusual for air-cooled spaces to provide for the future installation of chilled water as a hedge against future cooling requirements. Flexibility also frequently justifies the addition of valved stubbed-out connection points to allow for future expansion to be possible without any system downtime or costly work on operating hydronic loops (hot taps, freeze plugs, etc.).

10.2.6 Waste Heat Reuse

An unusual aspect of data centers is that they are very reliable, constant sources of large quantities of low-quality heat. Capturing and reusing the waste heat stream may be a profitable design goal, providing free heating and often good publicity. The waste heat is low quality (relatively low temperature) but can be a tremendous asset if there are adjacent spaces that require heat. There is a particular synergy between laboratory facilities with constant outdoor air requirements and data centers; projects incorporating both are a treat for the designer who treasures elegant design. Heat recovery chillers or heat pump-based systems can boost the heat quality, at some electrical consumption cost, to even feed a campus loop.

Computer chips themselves commonly have safe operating temperatures over 150°F (66°C), but maintaining that chip temperature requires much lower air temperatures with traditional air-cooled heat sink design. The low quality of heat available for recovery from IT equipment is often not a function of the chip requirement itself, but of the heat sink and casing design—changes in either area could make waste heat harvesting more practical.

[1] Any curtain system that interacts with the distribution of fire suppression systems, in particular by dropping upon alarm to remove barriers to agent distribution, should comply with the current NFPA 75 requirements regarding not blocking exit paths in addition to other applicable NFPA standards and local codes.

10.2.7 Profitability

Every member of the design team knows but rarely bothers to say that the primary reason for a data center to exist is to make money. This impacts the mechanical design in countless ways. There are the obvious construction budgeting exercises the mechanical engineer often supports, ranging from construction cost estimating to life cycle maintenance and energy cost. There are also less explicit aspects, such as providing a high enough level of reliability and adequate flexibility to allow for economical future expansion or providing a system that is attractive to potential tenants. A tight focus on the technical and first-cost challenges of the design is natural, but stepping back and regularly considering this larger picture, which can be more heavily influenced by long-term maintainability, flexibility, and efficiency aspects of the design, can help ensure the best ultimate design for the client.

10.2.8 Efficiency

While often overshadowed by reliability and schedule demands, efficiency is relevant in the long term since data centers make money by running computing equipment, not cooling equipment. Increasing the efficiency of the supporting mechanical system allows for more power to be dedicated to supporting the profitable equipment.

As data center design matured past the building boom of the Internet bubble, attention has turned to the electrical costs of these facilities. Dedicated data center operators may view higher efficiency as being a key competitive advantage, while a data center supporting a corporate office may look to data center efficiency improvements to meet carbon emission reduction goals. Large data centers may find their growth limited by the amount of power available from the utility and look to mechanical efficiency to free up capacity for IT equipment expansion.

A common metric used to help evaluate the efficiency of a data center is the Power Usage Effectiveness, typically referred to as the PUE. The PUE is roughly defined as:

Total electricity used by data center/electricity consumed by computing equipment.

By definition, the theoretical best possible PUE is 1.0. There are occasional debates as to whether a PUE of less than 1.0 is possible by taking credit for recovered heat, but credit for that type of approach is captured through use of a different metric. The PUE is often the total facility electricity usage as measured by the utility meter divided by the uninterruptible power supply (UPS) system delivered electricity, but the precise application of the PUE calculation still varies in practice.

While standards currently under development are expected to remove ambiguity and tightly define performance metrics,[2] PUE is currently often used quite loosely, occasionally quite incorrectly, and remains an ambiguous term in design practice. There can be question as to whether the annual average PUE or peak day PUE is being discussed. For comparisons of existing facilities, the local climate conditions must be considered; the exact same data center design in Oregon will have a better PUE than if it were in Florida. There are also difficult questions in where to draw the line between the IT equipment and the supporting mechanical equipment, particularly in innovative system approaches. For example, if the UPS is incorporated directly into the IT power supply in the IT server box itself, does that move the UPS loss from the numerator to the denominator? What about the opposite, a system that breaks the power supply out of its traditional place in the IT server box to consolidate in a large, potentially more efficient, centralized power supply? Small fans within the IT are considered part of the IT load, yet a design that measurably increases their "computing equipment" power consumption results in an "improvement" in PUE but a net increase in facility power usage.

Despite the limitations, PUE is currently the best metric available for data center owners to require and evaluate their facility efficiency. It is the responsibility of the mechanical designer to ensure their client understands the ramifications of design decisions on not just the PUE but the underlying efficiency of the facility.

10.2.9 Design Standards and Guidelines

Clients often have guidelines or standards that they wish to meet, such as conditioning to meet an ASHRAE TC9.9 Class rating, meeting an Uptime Institute Tier rating, or a legacy internal design criteria document. It is important that the design engineer understand both, the standard and *why* it is being sought. In some cases, meeting a standard may be a matter of precedent—"that's how we did it last time"—while in other cases it is a hard insurance requirement that will be fully audited prior to profitable occupancy. Clearly defining the driving requirement with the client at the start of project can help focus design effort and resources toward meeting the underlying objectives of the client.

10.3 MECHANICAL DESIGN PROCESS

Three common design styles can be summarized as implementation, optimization, and revolution. They all follow the same process but influence it at every step.

An implementation design relies heavily upon using off-the-shelf data center systems and very mature configurations. For example, locating a handful of CRAC units around the perimeter of a data center with an underfloor supply

[2] Work is currently under way in ISO technical committee JTC 1, subcommittee SC39, to provide robust performance metric definitions.

plenum and through-space return air path is an implementation design. This approach allows for quick design of a high-reliability space but often falls prey to the colloquialism that you can have it fast, cheap, or efficient—pick two. For a small or temporary data center, the small investment in design and integrated controls can make this an attractive option. In some cases, the entirety of the system design can be adequately, or at least at low cost (design cost, not equipment cost) and quickly, provided by the equipment vendors.

An optimization approach evaluates several different options and requires significant engineering calculation to implement. The system type and equipment will ultimately copy an existing design but be tweaked and optimized to best fit the current client requirements. Chilled water or glycol water systems often fall into this category, as do systems with air distribution more complex than underfloor with through-space return. Central control systems and evaluation of several different system types during the schematic design phase would be expected. This is what most mechanical engineers assume is desired.

The final design style of revolution seeks an optimal design but allows it to differ from precedent. Due to the top priority of reliability, this approach is challenging but can be critical to meet an impossibly low construction budget or an impossibly high efficiency, density, or other program requirements. Common hallmarks include using systems not marketed "off the shelf" to data centers, unusual distribution systems, integration of typically independent design aspects (e.g., HVAC components designed into the racks, heat recovery for other uses on-site, custom requirements on the IT equipment operating envelope, etc.), and a sophisticated client. A revolutionary design requires significantly more design work, a closely coordinated design team, and a technically proficient client. It is not appropriate for every project, but as it is open to embracing the best solutions, it is the theoretical ideal.

The standard data center design path is now well worn, and while it has an unforgiving cliff next to it (there is no margin for error), its challenges lie in smoothly and economically executing the design more than creating it. There are always different components to consider, but short of the inevitable handful of unique project requirements (a tape drive room! a glass tour hallway! no roof penetrations!), the design is an implementation and optimization exercise. Efficient design follows the same path but with a more deliberate effort to question assumptions and quantify the efficiency of design options. This requires more engineering time and/or a design team with a wide variety of experience. It is the questioning of assumptions—whether driven by a desire for a more efficient facility, first-cost constraints, unique site opportunities, etc.—that can combine an acceptance of risk, by design effort an acceptably minute additional risk, which can lead to a revolutionary design approach.

The system type is generally selected early and is often the single most critical choice driving the ultimate data center efficiency. Loads are based primarily on the program, how many kilowatts or megawatts of computing equipment desired, plus the overhead of the mechanical equipment supplying cooling. Typically, program loads are known at the start of the project, although in some cases they will be refined and adjusted as further information about available electrical capacity and mechanical system loads is developed.

After system-type selection and determination of the internal load, the design process continues with equipment selection and layout. As with any project, it is common to have some iterations of design revisions driven by the need to reduce the construction budget. More specific to data centers, there is a tendency to have design revisions to increase the cooling capacity of the mechanical system as the design team identifies opportunities to add more kW of usable power capacity for IT equipment (internal load to the mechanical designer) into the budget.

Drawing production and construction administration phases can vary somewhat based upon the construction model in use, ranging from design–bid–build to design–build. Beyond the delivery model, the scope and schedule of the project can also impact the phases of design, in some cases condensing phases or on the opposite end of the spectrum requiring a phase to have multiple iterations. While there is no one universal design process, the most standard is some variation of the following design phase progression: Predesign, Schematic Design, Detailed Design, Construction Documents, Construction Administration, and Postdesign Support.

10.3.1 Predesign

There are several different ways a data center design project is initiated that can impact the mechanical design challenge. While not required, input from the mechanical engineer at the earliest stages can ensure the most efficient facility. The type of project will inform the type of mechanical system insight needed.

10.3.1.1 Greenfield Development The least-defined designs begin with only a desired facility size and capacity goal (stated in kW or MW of IT equipment capacity). If the mechanical system's power requirement significantly influences site selection or other key activities common in the predesign phase, the mechanical engineer should provide the design team with estimated design condition system efficiency as a design parameter. System efficiency varies widely, so the estimate will be only approximate but can be adopted as a driving parameter that must be met as the design progresses.

The key mechanical efficiency needed in this case is not an annual average efficiency, but rather the highest mechanical power requirement in the worst-case, peak cooling load, extreme design condition. It is this highest demand that will dictate the amount of the electrical feed that will have to be diverted from supplying IT equipment—also known as the reason the entire facility is being built—to supporting the mechanical system.

Data centers produce an enormous quantity of low-quality (low temperature, ranging from 72 to 100°F (22 to 38°C), dependent upon the design) waste heat. In cases where there is flexibility in the location of the facility, such as on a corporate campus, demand for the waste heat can play a role in the site selection.

Sometimes, potential data center sites can differ by tens or hundreds of miles, bringing an evaluation of the climate zone into the siting question. The most extreme case is where the data center can be sited in a climate zone that requires little cooling. With less geographical span, the data center may still have varying access to a large body of water that can be used for cooling, such as a river, large lake, or even a large wastewater treatment facility. A small investment of mechanical designer input can catch when such rare, but highly valuable, opportunities appear during site selection.

10.3.1.2 Converting an Existing Building into a Data Center Data centers are surprisingly flexible in the kinds of buildings they occupy. A renovation situation can offer mechanical options, and limitations, that would never occur in a purpose-built facility. For example, an extraordinary amount of floor-to-floor height may be available for a data center sited in an unused warehouse, offering great opportunities for efficient airflow—or a drop ceiling with a wasted 18' high space above it. Or the data center may be crammed in an existing office building's windowless basement with no exterior walls and only an 8' floor-to-floor height. Either of these example situations can successfully house a data center, but the design team should be informed of the impact they will have on the type of mechanical system options that would be available.

10.3.1.3 Expansion It is quite common to expand an existing data center. The mechanical engineer should assist in evaluating the existing systems to determine if there is usable capacity to dedicate to the expansion. If the operator has had a negative experience with the existing system, that can drive a change in system type for the expansion. One common opportunity with expansions, particularly with chilled water systems, is to reduce the cost of redundancy by intertying the expansion system with the existing mechanical system. There may also be an opportunity to improve the system efficiency of the existing system by integrating it with a newer, more efficient expansion system. An example of this would be a chilled water plant expansion to support new data center space that installed new, high-efficiency chillers to serve the new load and retired an existing, low-efficiency chiller system to standby backup operation.

10.3.1.4 Remodel of Existing Data Center Remodeling of an existing facility without an expansion of footprint or cooling capacity is not very common. When reliability is the top priority, it is rare to modify anything that is not broken. The mechanical engineer should use any predesign phase to identify explicitly the driving motivation for the remodeling: solving hot spot issues, reducing energy use, capitalizing on energy reduction incentives, achieving a corporate carbon reduction goal, meeting a potential client's redundancy requirements, etc. Often, the motivations for a remodel project are typically well known, and these predesign tasks are condensed into a single effort that generates a request for proposal document or a preliminary meeting with contractors to request a quote for services.

10.3.2 Schematic Design

There are several common objectives in the SD phase, with the priority placed on each, varying by client. The typical SD process begins by identifying the key design requirements. Creating an estimate of the load and selecting the system type can proceed in parallel to some extent, with development of initial equipment lists and/or schematic drawings of the ultimate deliverable, which will vary depending on the project schedule and objectives. When design time or budget is very short, schematic design may be combined with design development.

10.3.2.1 Objective Identifying the objective may entail no more than a careful reading of the detailed scope document or request for proposal provided when the design team was selected. Bear in mind that most design requirements appear obvious—all clients want low cost, effective space control, redundancy, etc. *But* the relative priority and underlying operational requirements need effort to clearly define and understand. In most cases, trade-offs between design priorities must be made. A good understanding of what motivates each design requirement allows the best trade-offs to be evaluated and offered.

The objectives will drive the required deliverables. Some projects may only require an initial equipment list and material estimate to aid in a schematic cost estimate required to assess the business plan. In this case, while schematic diagrams and block equipment layouts are likely to be performed to some level to ensure an accurate equipment list, the traditional schematic drawings might be omitted entirely from the final Schematic Design deliverable in lieu of a text narrative description. Or data center expansion may place speed over early cost estimation and prioritize advanced drawings over equipment sizing and listing. Understanding the specific needs of the project allows the mechanical engineer to most efficiently allocate their time.

10.3.2.2 Define Space Requirements The temperature and humidity that must be maintained in the data center are critical design parameters that are too often assumed based on tradition. Significant money can be saved in both the construction and operations budgets by properly assessing the true requirements for these parameters.

By tradition, data centers are kept quite cool, at 72°F (22.2°C) or even less—far cooler than most IT equipment requires. Many data centers operate with low temperature set points simply because designers and clients are copying prior data centers, all the way back to the dawn of data centers when low temperatures and humidity were required to keep the punch card feeders operating smoothly. While there can be risk in changing any aspect of a critical design from what "worked last time," in the case of modern data centers, it is worth research to evaluate the actual current equipment that will be conditioned to suggest the best design conditions. Standards bodies have begun to offer concrete recommendations for IT rooms design set points. For example, the well-established international building technology society ASHRAE technical committee 9.9 has design professionals as well as IT manufacturer representation to develop temperature requirements [1]. Based upon their work, a maximum normal operating space temperature of 80°F (27°C) is recommended for data center space, with excursions to significantly higher temperature allowable in some cases.

A common practical reason for lower space temperature set points in existing data centers is to compensate for a failure of airflow management. With poor (or no) airflow management, a common situation in older data centers, at some point in the room, the hot air exhaust from an IT rack is recirculated into the intake of other, or even the same, IT equipment. This recirculation results in a localized hot spot that could eventually cause equipment damage. Lowering the temperature set point of a data center is a common reaction to hot spots and it does help.

Another reason given for maintaining a low space set point is to provide a reservoir of cooling to the space in case of equipment failure—but the hope for cooling buffer provided is far less than one expects. When the amount of "stored" cooling is calculated, it is found to offer a negligible safety buffer to all but the most lightly loaded facilities.[3]

There are other rationales for designing and operating a data center at a low-temperature set point, including client expectations, but a proper assessment of the actual set point needed can often yield significant first-cost savings, higher space capacity, and lower operating costs for an educated client.

The required humidity set point is another area where the requirement is often set based upon custom and an assumption that "tighter is better," yet too aggressive a humidity control band can actually harm data center reliability. Humidifiers are a potential source of catastrophic failure due to their water supply. They should be minimized or even eliminated if possible. They also carry a significant operational cost, including maintenance and power consumption, which in extreme overdesign conditions can even impact emergency generator sizing. The true need for humidity control should be carefully evaluated, and modern guidance on static control considered (i.e., that humidification is not necessarily an accepted means of protecting components from electrostatic discharge, particularly in the rare situations that modern IT equipment requires static protection beyond that included in standard chassis design).

10.3.2.3 Cooling Loads One characteristic aspect of data centers is that their cooling load is almost entirely internal space load, the heat generated by the IT equipment inside. The space load is defined by the amount of IT equipment that the client wishes to house. Typically, the load is discussed in terms of watts per square foot by the mechanical designer.

Unlike an office building or other common commercial designs, data centers have an industrial load profile—almost flat, 24 h a days. "No windows" is a common design requirement for data center spaces, removing the largest source of shell cooling, which is usually already a negligible fraction of the load. Likewise, very little outdoor air is supplied for these transient occupancy spaces, with just enough for pressurization if code justification can be made to deem the space normally unoccupied. Typically, the heat generated inside by the IT equipment is an order of magnitude or higher than even the peak heat gain through the shell, so there is negligible impacts from the shell. All this results in a very flat load profile.

Shell loads can be calculated using traditional load analysis methods but are really only explicitly determined out of the abundance of caution that underlies all critical facility design; in a typical data center, the peak shell load is negligible (far less than the sizing safety factor) and could be assumed to be zero with little impact on the design. For designers less familiar with data centers, understanding the nature of this load can have a surprisingly wide impact on the design. Not only are load calculations radically different, but the stable and efficient part-load performance of the system takes on a higher priority. Unlike an office building,

[3] For a space with a 9 ft high ceiling and 20 W/sf IT equipment load, based on the thermal mass of air, the temperature would rise 5–10°F/min in the absence of cooling; consideration of thermal mass such as the floor gains little due to a low rate of heat exchange. In practice, unless loads are very low, the benefit of overcooling a data center to gain some margin for error in a total failure situation is illusionary.

most data centers have a minimal increase in cooling load on even the hottest summer day, and they are also typically designed to have redundant capacity at all times; these two characteristics combined result in them very rarely (never by typical design) operating cooling equipment at full load.

The design load assumptions should be regularly checked. It is not uncommon for them to change significantly as business plans change or additional information about utility power availability is discovered. It is often convenient to coordinate with the electrical designer who often is the first informed of internal load changes, since they have a direct impact on the sizing of the facility's extensive infrastructure.

In some large data center projects, there is a limited electrical capacity economically available to the site, and this limited feed capacity makes the assumed efficiency of the mechanical cooling system a critical factor in determining the power available to run IT equipment—every watt in cooling equipment is a watt less to run the profit-generating IT equipment. In this case, the mechanical system efficiency can become a critical design parameter that must be met to maintain the integrity of the business plan, and the designer needs to regularly calculate it and defend it (typically from cost-cutting exercises) accordingly.

A final aspect of the design is the exterior design conditions. While they typically have little impact on the cooling load that has to be delivered (which is overwhelmed by the large IT equipment internal cooling load), external conditions do significantly impact the capacity of the heat rejection plant. As a critical facility with 8760 h operation, extreme outdoor climate design conditions are often used rather than the more typical 1% or even 0.5% conditions. These can be significantly higher than the standard conditions used for mechanical design and will impact the sizing (and cost) of heat rejection systems. The outdoor design condition needs to be appropriate for the project and clearly documented for communication to the client.

10.3.2.4 System-Type Evaluation The mechanical system type may not be completely set during the schematic, but a preferred approach is often selected. The key parameters of the system type that should be evaluated include the cooling medium, delivery path, heat rejection method, and airflow management. The objective to selecting a design basis is primarily to assist in cost estimation, define the footprint requirement, and evaluate the efficiency. The very high-level selection of these system parameters can set the efficiency of the final product and have a major impact on operating energy costs. The system-type selection can also impact architectural parameters including ceiling height, external equipment space, and interior layout.

The selection of system type has an enormous impact on the mechanical design and the capabilities of the final product. During schematic design, different system types should be assessed for the ability to meet the design objectives. Beyond the main requirements, details like lead time requirements in a fast-track project and cost impacts on other disciplines such as the requirement for a raised floor or a larger emergency generator system should be noted and considered. A thorough high-level coordination need not take a lot of time, but can often be skipped if not made an explicit part of the process.

Some sophisticated clients may require that the data center meet a very specific efficiency metric or require a formal value analysis of multiple system options. The choice of system type and options will heavily influence the ultimate efficiency, so the relative efficiency of different options can be a key deciding parameter in scoring what system is the best option for the site. Even when the client does not require it, in light of the magnitude of energy consumption over the life of a data center mechanical system, the relative efficiency of differing system types should be considered.

10.3.2.5 Footprint Evaluation Data center mechanical systems have a significant impact on the architectural program, layout, and costs. At this stage of the design, a full layout is impractical due to the fluidity of the design, but rough estimates of the footprint of the major pieces of equipment, general paths for piping and ducting, distribution concepts, and machine room spaces are needed to coordinate the program requirements with the architect. The largest pieces of the system (including air supply mains if applicable) can be represented as rough rectangular blocks to quickly generate layout estimates. Any equipment located on the data center floor is of particular concern as it subtracts from the program space available to house IT equipment. The space required for airflow management and ducting is another large element of data center mechanical systems.

Significant cost and efficiency benefits can be realized by closely coordinating the architectural design with the mechanical system. The method of coordination varies greatly, from three-dimensional (3D) computer models to hand sketches on tracing paper, but regardless of the method, they all serve to allow the mechanical engineer to communicate to the architect the size of the system, the ideal layout, and the compromises that are implicit in the proposed actual layouts. All designs have compromises, and it is important to consciously identify them and use the design team's combined expertise to quantify them as much as schedule and budget allows.

Savings from architectural integration tend to be most significant in large, dedicated data center spaces. Beyond the traditional use of a raised floor, there can be opportunities to optimize and distribute airflow using architectural elements such as a ceiling plenum, or partitioning walls. Cost savings may be realized by placing the mechanical plant on a subfloor below the data center or by using exterior rooftop-mounted air handlers to reduce the conditioned space that must be built (in temperate climates where maintenance would not be hindered). Most designs can benefit from centrally locating utilities to shorten the lengths of the largest mains and offer opportunities to reduce the power requirements from the fans and pumps. Some system solutions, such as air-side economization, are heavily dependent on the architectural configuration to produce a functional system. Some products allow for integration of air handlers into the building structure, for example, by replacing an exterior wall with built-up air handlers with easy access to exterior air for economization.

Smaller data centers can also benefit from close integration with the architecture. A common potential benefit is the harvest of low-quality heat from a data center housed in an office building to warm adjacent office space during winter. There can also be low-cost-efficiency opportunities that can be realized by utilizing an adjacent office mechanical system during nonbusiness hours to provide air-side economization to a data center. Or cost savings from using the office HVAC system as a redundant cooling source (with the clear communication that the office space will sacrifice cooling to support the data center when necessary). Opportunities in these cases are typically limited by the design cost and a match of the humidification requirements between the office and the data center, with significant custom engineering to realize a workable interplay between the spaces.

10.3.2.6 Code Evaluation As with any project, an overlooked code requirement can become a late-in-the-design land mine. Code review should be part of every phase of design. Different localities will face different code challenges and inspector expertise in the area of data centers. An open evaporative cooling tower may be a good standard design solution in California but wrought with code implications in the United Kingdom where a legionnaire's disease scare can result in shutdown orders for all cooling towers in miles. Major code impacts like this are rare and should be familiar to the design team based on past experience; explicitly identifying and documenting code concerns is an important part of schematic design.

Specialized fire control systems that use a gaseous fire suppression agent or dry pipe preaction systems are common. While the fire suppression system is typically designed by a fire protection engineer, purge fan requirements, isolation dampers, and distribution piping often require coordination and assistance from the mechanical designer. Management of airflow is a critical task for high-density data centers, and associated partitions may impact the fire system design. As a longer-term concern, the future flexibility of a design should be evaluated in light of the fire control code requirements. For example, the use of flexible curtains to control the airflow of hot spent air is currently a common air management approach to allow ease of reconfiguration, but the curtains can interfere with the dispersal of fire extinguishing agents and require integration with the fire control system.

In some areas of the country, typically those with stringent energy efficiency written into the local codes, utilities offer incentive money to encourage more efficient design. This opportunity is only available in a limited number of areas, but it is worth checking with the local utility as early in the schematic as possible to identify any incentive money that may be available to invest in more efficient systems and protect them from deletion during the inevitable efforts to make budget later in the design process.

10.3.2.7 Prepare Deliverables Deliverables for the schematic design phase will vary depending upon the client and design team but at a minimum should serve to document the design assumptions, compromises, and recommendations developed during the schematic design phase. The most common deliverables are the same as any other design project: a design narrative and a set of schematic drawings. But deviations from these common deliverables can be called for in some cases.

In a cost estimate-driven project, drawings may be omitted entirely in favor of a more detailed narrative with an equipment list. The reasoning behind this is that when the primary objective of the schematic design is to develop a construction cost estimate, traditional schematic design deliverables like system-level single line drawings or block layouts of the main mechanical spaces are of little value; the design budget can be better spent developing a more detailed list of the basis of design equipment, area required, feet of piping, and pounds of duct. For the most efficient design process, the cost estimator will be accessible throughout the schematic design to make clear what they need for the estimation exercise and to highlight key design areas with a high cost sensitivity.

A more speculative developer-driven project may focus on generating a marketing piece out of the schematic. They may require a deliverable with attractive cartoon-style sketches of the proposed design, a layperson-level narrative of its advantages, and little focus on equipment lists and sizes. While a properly sized pump with the perfect break-water distance balancing efficiency with longevity is a beautiful thing, few nonmechanical engineers care; a nice 3D rendering of the space in full color is more important if the objective is to attract client deposits or sell a building owner on a project.

Because the expected SD deliverables can vary, it is important that the mechanical engineer communicate with the design team and (often indirectly via the architect) the client to ensure the correct materials are developed. Regardless of the primary materials required for delivery, a document that clearly states the design assumptions and limitations must be generated. While usually part of the design narrative, it could be a separate memo to the design team lead outlining parameters including the design load, space temperature requirements, and system requirements, such as the need for water piping on the data center floor or the exterior space required by dozens of independent air-cooled computer room air conditioners' condensers.

10.3.3 Design Development

In this phase, it is expected that the system-type selection is finalized and equipment sizing begins. Layouts of the mechanical spaces and distribution are made and coordinated with the architect. The ducting and piping mains are defined and documented in drawings to allow for clear coordination. Controls should be considered, although it is common (if often unwise) to do so only to a cursory level. Code authorities may be directly contacted to test any code interpretations and, in jurisdictions unfamiliar with data centers, begin an education process. Cost estimating, and the closely associated efforts to reduce construction costs to make the construction budget, often starts in earnest during design development.

10.3.3.1 Finalize System-Type Selection The system type, which can vary from air-based cooling of the entire room all the way to cool water piped directly to the computing equipment, has wide impacts on the mechanical design. Making a firm system-type selection early is good for controlling budget by keeping to a tight schedule, but there can be tension to keep the system type flexible to accommodate changes in the layout, incoming cost information, client preferences, and other concerns. The design budget and schedule will dictate how critical it is to end system-type comparisons. The mechanical engineer should be sensitive to the needs of the client and architect, but be clear on when aspects of the system-type to be used need to be decided to maintain schedule versus aspects that can change later to accommodate additional information. And regardless of when the system-type selection is finalized, be aware that a high cost estimate will almost always lead to a reopening of the discussion, so some amount of design rework should be assumed if the cost estimation or "value engineering" exercise is planned late in the phase.

Once made, the system-type selection should be explicitly documented by an email a memo or incorporated in a progress drawing set sent to the entire design team to help in coordination. There is little that hurts a designer's budget as much as a late change in system type, for example, from air-cooled computer room air-conditioning units distributing via a raised floor to water-cooled built-up air handlers using overhead ducting and plenum space. When a base system selection is made, declare quite explicitly to the team that it is a foundation assumption and changing it could result in additional cost and delay. The mechanical system also impacts most aspects of the data center design. Clear coordination of the selected type is important enough to warrant the redundancy of documenting the final decision even if all design fields were directly involved in it.

10.3.3.2 Value Engineering As with any project, there is a need for the final design to be constructible with the budget available. Commonly referred to as value engineering, this exercise of cutting construction budget from the design is becoming more common in data center projects as they become more of a common commodity space. The large size and expense of the systems within the scope of the mechanical engineer typically requires their significant participation in value engineering.

When investigating lower-cost design options, it is important for the mechanical engineer to coordinate with the electrical engineer to ensure the client understands that an extra kW used on HVAC equipment, perhaps due to the use of lower-cost mechanical equipment, is a kilowatt of generator and utility capacity not available to make money. The assessment of an alternative mechanical system or equipment option needs to take into account not just a potential reduction in the installation cost of that mechanical component but also any *increased* costs that may be incurred on the electrical system by the alternative. Impacts on redundancy, space flexibility, and expandability must be clearly defined and communicated to the client to ensure that an accurate assessment of cost-saving measures is made. Good value engineering can reduce the cost of the whole project without harming performance, but a design team myopically focused on only their own discipline's line item costs can reduce the final space's utility and actually increase the whole project cost.

10.3.3.3 Revise Load Estimate The key component of the load estimate is the power of computing equipment that will be supported. As the design process progresses, this critical design parameter can abruptly shift. Regular communication with the design team should ensure that the mechanical designer is aware of any relevant revisions. The mechanical engineer should also keep the electrical engineer updated of any changes in the need for power to support the mechanical system, with a keen awareness that decreases in the mechanical system's efficiency can cascade into a nonnegligible need for more generator and transformer capacity.

10.3.3.4 Preliminary Layouts Floor plans and data center layouts take shape during design development. Targeted and succinct input from the mechanical designer can ensure that mechanical concerns are met and issues such as minimizing distribution length (a cost and energy efficiency driver), providing enough space for appropriate maintenance access, airflow management, and planning for future capacity expansion are well handled.

The mechanical room layout has a significant impact on the system efficiency and operational requirements. There are often numerous trade-offs, such as desiring a very compact footprint but needing space to allow for maintainability or minimizing first cost by downsizing mains sizing at the cost of hurting future flexibility and efficiency. Mechanical layouts should be generated as early as possible. It is easy enough to make a high-velocity air system with a very small footprint, but the future flexibility, expandability, and operational energy cost implications of such an approach are grim. Optimization of the mechanical system layout is critical.

Where high efficiency is the primary goal of the system, mechanical equipment should be laid out accordingly. Airflows and water flows inherently waste energy when they make sharp right-angle turns. Recognition of this can often result in a mechanical room where equipment is located at an angle to the walls, piping is kept near the floor rather than routed over a rigid grid of aisle ways, and long radius turns and 45° laterals are common. One pipe fitter compared a particularly efficient plant layout to a sanitary sewer system—an apt comparison, since gravity-driven sanitary sewer systems are forced to adhere to low pressure drop layouts. While piping layouts to this level of detail are not appropriate in design development, the modest extra effort and (sometimes) floor space required for efficient layouts should be acknowledged and planned for if efficiency is a high priority. Air handler sizes should be optimized for 8,760 hours of power-consuming operation per year, rather than by office-based rules of thumb such as a 500 fpm (2.5 meters per second) coil face velocity.

Rejecting heat from the data center to the outside is the primary task of a mechanical system. The size, type, and location of the exterior heat rejection components, whether they are a cooling tower or a louvered wall, should be identified during the design development phase and any limitations identified. For example, a ban on rooftop equipment for security or leak concerns, hardening against tornado and hurricanes, or other uncommon but critical specific project requirements need to be determined and accommodated in the system selection and layout. Aesthetic and acoustical concerns can also be factors if the facility is located in a residential area or within the line of sight of a residential area; expensive houses on the hill with a direct view of the best place for a noisy cooling tower yard anecdotally tend to house local politicians and code inspectors with sensitive hearing.

Future expansion also plays a role in determining how much space is required, both on the interior and exterior. If a future expansion path is desired, it should be an explicit project goal and be directly incorporated in design development by considering and documenting where future equipment and distribution would go to support additional load. It often is cost-effective to provide some infrastructure to support future equipment, such as extending a tower structural support platform to fit more cells in the future, oversizing piping to provide future capacity, and adding empty electrical conduits when casting foundations.

10.3.3.5 Equipment Selection The selection of equipment is an important step in ensuring that equipment exists that can provide the desired performance within the rapidly solidifying space, cost, and energy budget available.

After the system type is finalized, and in parallel with developing layouts, preliminary basis of design equipment selection should begin by calculating equipment capacities and sizes. At the beginning of design development, a detailed equipment list should be started and the equipment schedule drawings begun. The most expensive equipment should be sized first, followed by the physically largest equipment and finally the auxiliary equipment, with the overall goal being ensuring that equipment is available that can provide the desired performance within the rapidly solidifying space, cost, and energy budget available. Items like pumps and fans can usually be approximated by calculation based on estimates of pressure drop requirements, while larger equipment such as chillers, CRACs, air handlers, cooling towers, and other similar items should have preliminary selections made to better define size, cost, and efficiencies.

The nature of the cooling load presented by a data center differs in several ways from a typical commercial office building load. The selected system equipment must be able to stably carry the design load even during design heating (lowest outdoor air temperature) conditions, at a time that office cooling plants are often shut off. The cooling system also must switch seamlessly and stably between any economization mode and mechanical cooling. In air-based systems, airflows are sized to accommodate the sensible-only load. Reheat of IT space is unnecessary.

Projects with an energy efficiency requirement to meet use the preliminary equipment selections to calculate the predicted system efficiency to ensure contract or design requirement compliance. While there are many energy modeling programs available for buildings, due to the simple nature of data center load (approximately flat, 8760 h a year), a spreadsheet calculation that uses hourly typical meteorological year data available from a number of sources or bin weather data can be successfully used to streamline this task. System interactions should be considered throughout the design. For example, a successful airflow management system that collects heat exhaust from the IT equipment can increase the air-side delta T and allow for smaller air handlers, paying for some of the first cost of the airflow management elements. Using low pressure drop plenums for air movement instead of ducting and allowing a higher temperature and humidity range in a portion and all of the data center are other system design decisions that can have far-reaching impacts on the mechanical system.

10.3.3.6 Size and Locate Distribution The data center mechanical system exists to move heat out of the data center. Regardless of the medium it uses to do this (air, water, glycol, refrigerant), there will be a significant distribution system (ducts, pipes, or both) to move the heat around.

An air-based system will require large ducts or plenums to allow for the volume of airflow required. Within the data center footprint, plenums formed by a raised floor and/or a false ceiling are typically the most efficient and flexible method of air distribution. The space itself is often used as a plenum to move the large volumes of air needed to cool the equipment. Ducting can offer a more controlled distribution system that can avoid some code requirements regarding wiring through space used as an air path plenum, but it is often less efficient. The choice of air system can significantly impact the cost of the fire suppression system by increasing the active volume.

Raised floors are often used to create a supply air plenum. This is a simple design approach but can run into limitations at high load densities as floor heights become economically unattractive (particularly in zones with extensive seismic requirements). If the underfloor space is shared with any other utilities, such as electrical distribution systems or data cabling, it can become surprisingly congested, resulting in inadequate airflow to portions of the spaces—close and consistent coordination with other trades is required, starting from when the floor height is initially estimated and continuing throughout design.

Raised floors are rarely used as a return air path; while having a floor plenum that serves as a return path is theoretically feasible (the buoyance effect of hot air is negligible at the air velocities seen in all but the most lightly loaded or specially designed data centers), current design practice and commercial products available only support use of raised floors for supply air.

Overhead plenums are often used for return air. In the common legacy design of CRAC located on the IT equipment floor using a raised floor for supply air distribution and through-space return, converting dead space above the ceiling into a return plenum is a common method of reducing mixing of supply and hot exhaust to eliminate hot spot problems, improve capacity,[4] and increase system efficiency. Code requirements on the type of electrical supply wiring and equipment allowed in the plenum space need to be considered, particularly when considering retrofit of an existing facility, along with any impacts on the fire suppression system from the added active volume.

Overhead plenums are rarely used for air supply, with ducting preferred for overhead supply. A return plenum can be combined with supply ducting to offer a hybrid plenum/ducted air management solution that does not require a raised floor.

[4]The cooling of most CRAC is a function of the temperature difference between supply and return air. Improved airflow management can increase this temperature differential and increase the usable capacity of currently installed equipment.

10.3.3.7 Investigate Airflow Management Airflow management is a critical aspect of avoiding potentially damaging hot spots in high load density data center design that relies on air for cooling (as opposed to cooling water to a rack-level system). The airflow management approach needs to be considered early in the design phase as it has extensive impacts on most areas of the mechanical design, including cost, effectiveness, efficiency, and system sizing. Architecture may also be significantly impacted.

The IT equipment housed in most data centers draws cooling air in one side and then ejects a high-temperature exhaust out the opposite side, ideally drawing air from the front and exhausting hot air out the back. Airflow management can take many forms, but the objective of all of them is the same: capture the hot exhaust and cool it before it is pulled into the cooling airstream of another (or the same) piece of equipment. Discussed in greater length elsewhere in this chapter, airflow management can take the form of anything from hung plastic curtains partitioning off the intake side of racks from the exhaust side of racks to distributed floor tiles with integrated and independent variable speed supply fans that vary the volume of cool air supplied from an underfloor plenum on a per-rack basis. In highly customized cases, the airflow management will likely dictate the architecture by dictating the space height or layout of spaces relative to the exterior wall.

In design development, the main priority is to determine the kind of airflow management system design that best fits the program and communicate to the other trades the impact it has on their design work.

10.3.3.8 Drawings While drawings may be skipped in favor of a costing-targeted narrative in the schematic phase, it is rare that the design development phase does not produce drawings. For larger projects, there are often one or two progress sets compiled during the design development phase to assist with interteam coordination.

Drawings are developed to support costing exercises, document design progress, and aid coordination in this phase. Any coordination agreements between the disciplines, ranging from the location of mechanical rooms to the electrical capacity (or, more crudely, the motor horsepower) required for mechanical, should be clearly documented as a valuable product of this phase that could be lost if left in notebooks or buried in an email chain. Common drawings required in this phase include an equipment schedule, air-side diagram, and water-side diagram that serve to record the current state of load calculations and system selection. Layouts of plant rooms and equipment yards are also typically provided, although they are subject to adjustment during the Construction Document phase.

Preliminary layouts of mechanical equipment and distribution are an important coordination tool between the architect, mechanical engineer, and electrical engineer. They also serve to inform more experienced clients of the scope and type of systems they will need to support with operations staff.

Detailed drawings are primarily a task for the next design phase, Construction Documents, but when significant detailed design work was done in the Design Development phase, it is appropriate to document it. This most often occurs when an unusual system or design approach is being considered, and it must be designed to a significant level simply to verify it is a feasible option. Such design aspects tend to be defined by their unpredictability, but they could include features ranging from the suspension hardware configuration of a hung curtain air management partition to the construction details of a built-up air handler with direct evaporative cooling/humidification incorporated into a structural wall. Beyond the case of subsystems that are developed to unusual detail to prove feasibility, a significant number of generic details will be included in this phase as they are available from the designers' standard work to "cut and paste" into the project; while not necessary, this can help coordination for projects with short schedules, little communication between the design disciplines, or design team members who are unfamiliar with the proposed data center systems.

10.3.3.9 Code Investigation Any outstanding code questions that impact the design should be settled in this phase. They may be settled in a variety of ways, ranging from verbal or email confirmation from the authority having jurisdiction to an agreement with the client representative about the interpretation being used and the worst-case cost of the interpretation being rejected. At this stage, the risk is typically design rework and the associated costs and possible delays. The implications of worker safety codes on operation should also be determined and communicated to the client. For example, as data centers move to creating hot aisles that operate at high temperatures, operators may be legally obligated to limit worker time in those areas—which can be a problem if extensive rack wiring and hookup needs to be regularly performed from the hot aisle side of the IT rack.

10.3.3.10 Cost Estimating and "Value Engineering" Supporting cost estimating efforts and investigating opportunities to reduce system first cost are often a high priority throughout the design process. Any deviations from very traditional standard design should be clearly documented for the cost estimator and reviewed closely by the engineer. The mechanical engineer should review the cost estimate for equipment type, sizing, pounds of ductwork, and other key cost elements. Cost estimating is often done by a contractor, who in the process of a cost estimate can often offer a useful viewpoint on the design's apparent constructability and the clarity of documents.

If a design change being considered to reduce cost will impact the cost of other trades, the mechanical engineer should inform the cost estimator and review the ultimate cost estimates to ensure it was accurately captured. For example, utilizing smaller, high-velocity air handlers may reduce air handler cost but need more fan power and increase the cost of electrical support systems ranging from panels to the building transformers. Whole-building impacts of this type are often overlooked in the early design cost estimates, which can lead to poor value engineering decisions being made. Some savvy clients may also request a net present value analysis to capture the operating cost impact of changes.

10.3.3.11 Controls Control design is often left to the Construction Document stage. This is a reasonable strategy to avoid rework, but research should be completed and documented by the end of design development to identify the type of control system desired to ensure cost estimates are accurate and assume an adequate level of control investment to support the proposed system. Common types of control include central direct digital, independent integrated computer room air-conditioner unit controls, or some combination. It is not unusual for smaller data centers to have the control system consist of the onboard controls of computer room air-conditioning units—which have a very different capability and cost profile than a central direct digital control (DDC) system. The intended type of control should be clearly defined and communicated to ensure that it is captured in the cost estimate, electrical and architectural coordination issues are identified, and owner expectations are appropriate.

Any unique or complex control approaches should be described and detailed as far as necessary to verify feasibility. Water-side or air-side economization features in data centers, which offer huge operating cost savings, often require control approaches that differ significantly from the standard controls used when these common systems are applied to office space.

10.3.3.12 Prepare Deliverables Deliverables for design development phase will vary depending on the client and design team. Again, early coordination with the client and/or architect that clearly defines what the mechanical designer *will* deliver as opposed to what they *can* deliver is critical to providing a high-quality and complete deliverable. Most, if not all, design development deliverables represent a preliminary version of a Construction Document set deliverable. Typical deliverables include a design narrative summarizing the system design, initial specifications, and drawings that illustrate the location and size of major pieces of equipment, air distribution paths, required active plenum spaces, main piping distribution paths, and preliminary sizing and power requirements (for electrical coordination) of major pieces of equipment. Significantly more detailed information may be required in some cases, for example, if the project delivery model includes some form of bid and award at the end of design development to bring in a contractor, developer, or another external entity to take the project through construction.

The design narrative will typically be an update of the schematic narrative deliverable. While the schematic deliverable will often discuss and compare different options, the design development deliverable focuses on the single selected system approach. Space load assumptions in terms of the computer equipment power consumption in the data center are clearly defined, ideally in both a watts per square foot capability for each program space and a total system kilowatt for the entire building. Where basis of design equipment selection have been made, it is appropriate to include preliminary submittal data as an appendix.

Specifications should be focused on defining the equipment requirements and any expensive execution requirements, such as requiring all welded piping or high-efficiency axial vane fans. While ideally a full set of draft specifications are collected, they may be very preliminary with minimal placeholders used for typical areas. Not all projects will require preliminary specifications in the design development phase, but even if not required for the submittal, it is often a design efficiency to begin tailoring them as the equipment selection tasks of design development are completed.

Drawings allow for more detailed coordination between the disciplines and should provide enough data for peer review, be it external or internal to the design team. Drawings should contain as much information as available on calculated loads, equipment sizing, distribution duct and piping sizes, system configuration, and layout. Avoid adding "filler" information hastily cut and pasted in merely to make the drawings look more complete to avoid problems arising from the unpredictable use of the design development drawings. A commissioning plan may be developed from this design deliverable, or an energy model created, or additional cost estimation, or other tasks that require information on the mechanical system configuration. It is better that incomplete areas are left undefined rather than a hastily added filler misleading other works and ultimately resulting in wasted time.

10.3.4 Construction Documents

The construction document phase is the completion of all design tasks required to allow for the permitting, bid, and construction of the facility. It is often the most costly design phase, but at the same time, the majority of big design decisions impacting system capacity, flexibility, and efficiency have been completed at the outset of this phase.

10.3.4.1 Finalize Equipment Selections Load calculations are finalized and the final equipment selections are made during this phase. Depending on the construction schedule anticipated, lead time of major pieces of equipment may be a factor in the final equipment selections. It is good standard practice to ensure that there are multiple providers of equipment that can meet the specifications—often a requirement with large or government clients. Beyond the typical savings advantage of a competitive bid to supply equipment to the project, verifying multiple suppliers of equal equipment ensures that the project will not be disrupted by a single supplier withdrawing the basis of design equipment from the market. Or, at a minimum, clearly highlight where equipment substitution may require design changes. Such postbid design changes tend to be costly, be it merely a forced increase in mechanical room size because an alternate air handler has a larger footprint or a full redesign to an entirely different air management solution.

If initially performed by the design team using software, website, or catalog procedures, key basis of design equipment selections should be verified with a manufacturer representative to ensure accuracy. All details of the equipment selection need to be defined, verified, and recorded in the design documents. The number of details that need to be verified are as varied as the types of equipment that may be applied to a data center. Care must be taken to properly specify the right options, particularly in the area of controls and low outdoor temperature operation (unlike office buildings, data centers will need to generate cooling even on the coldest days).

The redundancy strategy used, such as $2N$ or $N + 1$, should be included in equipment schedule notes to record the basis of design and aid commissioning efforts. Equipment should be selected with consideration of the reliability and maintainability required for data center operation.

10.3.4.2 Clearance and Interference Issues The general equipment layout and distribution paths should be defined by this design phase. The final coordination of equipment layout with all other trades should ensure that there will be no interference or conflicts between trades. It's a risky game to count on contractors in the field to solve interference issues during construction, even if the job utilizes a design–build delivery model. Pipe sizes need to be laid out with allowance for insulation thickness, ducts fitted with consideration for the size of flanges, and equipment placed with the required code and desired maintenance clearances around them.

When coordination is done primarily by two-dimensional (2D) plan layouts and sections, piping and ducting need to be shown with thickness (double line) on the drawings. In congested areas, sections need to be provided to verify that the systems fit. Sometimes, elevation levels are assigned for different equipment, for example, defining the ceiling and lights as being in the band 9 ft 0 in. to 9 ft 10 in. above finished floor (AFF), mechanical piping and hangers at 9 ft 11 in. to 12 ft 11 in. AFF, and fire and electrical distribution at 13–15 ft AFF. This method of assigning elevations can be effective but may require more height than is absolutely required and additional coordination to ensure that vertical elements, typically hangers and seismic bracing to the aforementioned, are accommodated. Equipment should show clearly the clearance and service space required around it, including code clearance requirements in front of electrical panels.

3D modeling is becoming more common and can be a valuable tool to solve interference problems before they cause trouble on the construction site. 3D modeling significantly impacts the Construction Document process. Designer time and budget is shifted out of the construction administration phase, where the final coordination was often in practice completed, and into the construction document phase. Budget and staffing hours need to be shifted accordingly. The objective of this design investment is a better coordinated design that minimizes construction delays and change orders—ideally saving time and change order costs that more than offset the additional Construction Document time.

An often-overlooked coordination issue is the location and airflow around exterior heat rejection equipment. Data centers are designed for continuous control of the space, including during hours of extreme high temperatures. This will highlight any problems such as cooling towers that suffer from recirculation due to the placement of a screening wall or dry coolers bunched together in the middle of a black roof heat island with local air temperatures a dozen degrees higher than ambient. The common presence of redundant capacity that can be used during extreme heat periods provides some leeway but only a small amount since failures often occur on the extreme hottest days (not due just to bad luck, but rather the highest cooling loads correspond with the worst operating conditions for bearings and windings). Extreme hot exterior conditions will expose poor heat rejection airflow design on a fully loaded data center. Lawn sprinklers wetting overtaxed dry coolers on the roof of a data center are a depressingly common dunce cap placed on inadequate heat rejection designs.

10.3.4.3 Controls The building controls are a critical element of a successful system yet are often left until late in the design process to be designed. To some extent, they are delayed simply because there is not a pressing coordination need for them to be defined earlier. Beyond defining a few locations where electrical power is required or wall space is needed to hang the control boxes, control coordination occurs entirely within the mechanical design.

Coordination of the control design with the equipment selections and specifications is critical. While a small data center facility may require no more than onboard controls that are integrated into the air-conditioner units installed on the data center floor, many system types used for larger facilities will require a networked system with external sensors or the flexibility of a more customized central DDC system. A number of control aspects require definition. Each piece of equipment must have the proper interface type defined and control input capabilities. Commissioning, an important testing aspect for a critical facility, may also require control features such as trending or remote Internet access (a very helpful monitoring and diagnostic tool, albeit one that carries a security requirement).

The control sequence is the logic that defines the system operation. The best control approaches will be simple enough to ensure reliability but complex enough to provide flexible and efficient control. As energy costs increase, the demand for controls to minimize system power consumption also increases. While reliability and robustness are the primary design concerns, a good control design will implement common efficiency best practices, such as varying airflow, controlling the temperature of the air supplied into IT equipment rather than returned, and efficiently adjusting system operation to most efficiently match part-load conditions.

The most traditional control strategies tend to be reliable but very inefficient. For example, maintaining a return air temperature set point equal to the desired space temperature is simple (and simple is reliable), but since the return air temperature should be higher than the temperature of air supplied into the IT equipment intakes (the point where temperature control is required), this approach will chronically overcool the space. It will not directly control the parameter of concern, that is, the air temperature supplied into the IT equipment intakes. As overcooling is not typically viewed as a failure in data center control—the expectation of a computer room as being almost refrigerator cold is common—traditional control sequences are often biased toward inefficient and even uncontrolled overcooling. Many CRAC manufacturers have begun to offer more efficient control options that utilize supply air temperature sensors, remote sensors located near the IT equipment intakes in the space, and variable speed air supply fans to offer improved efficiency. DDC systems have offered the flexibility to provide this kind of control for many years but at the cost of increased complexity and design effort.

Every effort should be made to minimize control complexity to the extent that doing so does not harm control capability and efficiency. Complexity tends to introduce delay in system start-up as problems are identified and corrected, as well as introduce more points of failure that can reduce the reliability of the system. Some complexity is a requirement to provide the best space control—with the exception of lightly loaded and expensively overdesigned data centers, simply turning the mechanical system to full on is not an acceptable modern design. A good control system has the ability to match cooling output to the actual load, prevents overcooling of the space upon sensor or actuator failure, provides efficient space control, and can be fully understood—and therefore maintained—by the future system operator, not just the design engineers.

Humidity control in particular can be a problem in data centers. The humidity control set points should be relaxed to properly match the actual needs of the housed IT equipment to ease the control problem. The control approach needs to acknowledge and accommodate expected sensor drift over time, since humidity sensors are significantly less reliable than temperature sensors. The design should also take pains to avoid a situation where sensor error over time can result in independent systems serving the same space fighting, a situation commonly seen with CRACs using independent humidity sensors where due to sensor error one is humidifying, while another serving the same space is dehumidifying.

10.3.4.4 Coordinate with Electrical All disciplines must coordinate and integrate their designs during this phase. Coordination with electrical is sometimes relegated to "throwing drawings over the wall," but significant system savings and optimization may be achieved through more frequent coordination. The design capacity of the UPS system typically dictates the IT load that the mechanical system must be sized to support, so this design capacity should be verified regularly to catch any last minute changes that could greatly impact mechanical sizing. Impacts run from mechanical to electrical too; for example, the size of the emergency generator is significantly driven by the mechanical system efficiency. If the generator is near a size break point where a minor reduction in load could allow the use of a smaller unit, the cost–benefit assessment of the value of buying more efficient equipment to reduce the peak mechanical system kilowatt can change radically, perhaps to the point that a first-cost reduction and operating cost reductions can be achieved from taking the whole-building assessment approach. Similar effects may occur all the way back to the building transformer level. Capturing the extensive interaction between mechanical efficiency and electrical first cost during life cycle cost estimation and evaluation should have occurred in design development, and it should continue with the greater design resolution available during the finalization of the design in Construction Documents.

10.3.4.5 Coordinate with Fire Protection Fire protection systems are highly specialized and jurisdiction dependent, so their final design drawings are typically produced by a fire control specialist—introducing another discipline that requires coordination. Airflow management design approaches often interact heavily with the fire protection scheme by introducing partitions in the space. The fire protection design must accommodate the partition scheme and any active plenums to ensure code compliance and proper protection of the space. The mechanical engineer should also capture the fire behavior required of the mechanical system during an alarm condition. While office space air handlers are commonly shut off during a fire alarm event, the critical nature of data center cooling often calls for a fire control scheme that keeps the cooling system, including air handlers, operating during a fire alarm. Another coordination issue is meeting any space exhaust requirements associated with the fire protection system if a dry gas-based system is utilized, including exhaust fans, relief dampers, and the associated control integration.

10.3.4.6 Coordinate Equipment Layout with Architectural and Electrical A minimum level of coordination with electrical is achieved by accurately showing mechanical equipment locations and requirements on the coordination drawing sets that are generated regularly through this phase. It is also important to ensure that the electrical parameters for equipment shown on the schedule—the phases, voltage, and design amperage—are accurate. If control panels require UPS power to avoid unacceptable reboot delays upon a loss of power, then that should be clearly communicated along with the locations of all control panels that will be powered. The equipment that requires emergency generator backup also needs to be clearly defined, with any equipment that does not need backup clearly identified.

10.3.4.7 Coordinate IT Layout The layout of IT equipment is often defined to some extent by the mechanical system airflow design. High equipment loads require an airflow management design that prevents hot exhaust streams from overheating an adjacent piece of IT equipment. Most airflow management designs enforce some limitation on where IT equipment will intake cool air and exhaust hot air. A common requirement is that IT equipment will be placed into standard-size racks and arranged in rows with cool air pulled in from the front "cold aisle" and hot air be ejected out the back into the "hot aisle." Most (but not all) IT equipment follows this airflow arrangement; if it is required for proper space control, the mechanical designer should clearly state that limitation of the system and coordinate with the client to ensure that the design will meet their needs. And if it does not, the first-cost and operating cost penalties of incorporating more flexibility should be summarized and communicated before the design is modified. Designing to allow for random rack layouts, as may be required for some applications where space is rented out to multiple different clients (often referred to as colocation facilities), is more expensive and cannot handle high-density loads unless rack-based cooling (which solves the air management problem by placing the cooling coil or water-cooled heat sinks literally within inches of the heat load) is used.

10.3.4.8 Complete Distribution Design and Calculations As with any design, pumps and fan sizing is customized for the project's specific distribution layout. The final sizing is done with consideration of the data center operation profile. Data centers operate 8760 h a year without downtime available for reconfiguration work; flexibility must be designed in with features such as "oversized" distribution sizing to allow future expansions or rezoning of loads. Such oversizing can also reap significant energy savings if the system is designed to capitalize on it by turning down efficiently.

10.3.4.9 Complete Specifications Design specifications fully define the required equipment, components, and installation methods for the mechanical system. While the outline of the specifications is produced in Design Development, significant work occurs in Construction Documents to complete the specification book. For data center designs, particular attention should be paid to the allowable equipment substitutions. Commercial air-conditioning equipment may be significantly cheaper than data center-specific equipment but wholly unsuitable to serve the primarily sensible load and 24 h reliability needs of a data center. The specifications need to tightly define all aspects of the equipment, in particular the redundancy, reliability, part-load efficiency, and control components that tend to be significantly different from and more expensive than in commercial equipment.

Specifications are developed from a number of sources. The starting point is often a library of standard specifications either produced over time by the designer or licensed from a specialist source. Equipment manufacturers often provide guideline specifications, which are useful once unimportant aspects ranging from trademarked coil treatments to the color of primer coat are trimmed out to allow reasonable substitutions. Regardless of their initial source, specifications must be fully reviewed and revised to meet the reliability and critical facility nature of the data center. Using specifications produced for a prior successful data center is an acceptable starting point, but full review by the designer is a (tedious and time-intensive) must. Submitting a specifications set that carefully defines by name the equipment that does not exist on this job is embarrassing but is only an inkling of the expensive grief that can occur. Erroneous specifications combined with contract language on allowable substitutions can make denying the substitution of unacceptable (but relatively cheap) commercial air handlers in lieu of purpose-built CRACs an expensive change order.

The basis of design space conditions, loads, and design weather conditions should be included in the specifications if they are not stated in the drawings. These critical parameters are defined and approved in the prior design phases' narratives. Including this information clearly in the construction documents, which will usually become part of the building operator's documentation while design narratives do not, is of significant value as the data center is loaded up and potentially remodeled in future.

10.3.4.10 Generate Coordination Drawing Sets The final construction drawing set is developed in this phase. Multiple drawing sets are submitted to aid in design team coordination, usually including a 30%, 60%, 90%, Permit, and Final CD set. A small data center may combine some of the coordination sets into a single review, and a larger data center may rely on bimonthly meetings to review a common 3D model. The number of drawing sets should be clearly defined in the contract scope and verified during normal design coordination communication between the mechanical designer and design team lead (architect, client, design–build general contractor, etc.). The exact scope of each coordination set varies with the project; the following discussion is a general guide.

Care should be taken to ensure that estimated data placed on the drawing to allow for early coordination is clearly tracked and replaced with the correct, calculated data as it becomes available—round numbers for design parameters such as 100.0 ft water gauge for all pump heads or 10 horsepower for every fan motor are common telltales of placeholder data that has escaped proper update with the final, calculated sizing. Small oversights such as not updating the estimated pump size to match the final calculated pipe loop pressure drop (plus safety factor) can prove costly.

The 30% drawing set provides information on the proposed equipment layout and types, with a focus on aspects that require coordination with the other design disciplines. An early priority to support the electrical design is to set the location of major pieces of equipment and define their electrical demands. To support architectural integration, the location of all outdoor air intakes and exhausts, major ducting, external equipment, and piping are early priorities. This data should be clearly presented in the 30% drawing set and is therefore often the subject of the earliest coordination meetings. In a tight layout situation, problems with architects are traditionally solved with drawings and sketches on overlaid tracing paper. 3D modeling software is a rising alternative for coordination.

All coordination concerns raised by the 30% set should be resolved by the issuance of the 60% set, with additional information added. Additional coordination problems may arise as detail is added to distribution routing, and duct and pipe sizes are fully defined. The 60% set has most or all equipment selection finalized and distribution pathways clearly shown. Final pump and fan sizing calculations, which are based on the final routing and sizes of ductwork and piping, have at a minimum calculated estimates completed and in the equipment schedule.

Controls are defined for all equipment with preliminary sequences shown to illustrate the intended operation in the 60% set. If the controls are integrated into the specified equipment, the required options and intended settings are defined by clear notes on the equipment schedules (defining options only in the specifications or in control details is technically acceptable but in practice is more prone to being missed by contractors, causing trouble during construction). Integrated control capabilities can vary significantly between equipment suppliers, so it is important to define the control requirements fully, assess their impact on availability of alternate equipment, and ensure any bid or construction alternates offered provide equal control capability.

All drawing sheets that will be present in the final design package should be represented in the 60% set. The 60% set should include drafts of all drawings, including controls, permit sheets, plumbing, and fire suppression. Requiring that the responsible team members provide drafts of these drawings for this set ensures design team members are fully aware of their scope. While there should be no scope confusion this late in the design, if the mechanical design is assuming that an external consultant will provide a fire protection design and permit documents while the client is expecting fire protection to be integrated into and provided with the mechanical set, the 60% coordination set can be a painful but not catastrophically late point to recognize and correct the scope confusion. The mechanical engineer should check and verify that all expected sheets are in the coordination set and follow up to verify if any are missing. It is also important to verify that electrical is supporting all the mechanical equipment, including any accommodations for future equipment.

The 60% level is often the point where all disciplines have provided drawings (and/or electronic 3D models) with the level of detail and accuracy suitable for identifying interferences and other conflicts. Regular meetings, either in person or by voice conferencing with Internet screen sharing, are often begun to resolve interference issues as the final design package is completed.

With all drawings represented in the 60% drawing set, the 90% drawing set is simply a completed version of the 60% set. While rarely attained, the objective of the mechanical designer is for the 90% set to be the final design and require only cosmetic title block updates prior to release for bid and construction. Equipment sizing is completed and based upon final design calculations of load, pressure drop, and layout-specific parameters. Equipment layout is completed, including maintenance access and verification of installation/removal corridors and door heights. All distribution requirements, including the minor but critical plumbing associated with humidifiers and condensate control, are fully defined, sized, and shown. Plumbing is captured, and airflow management is shown in the set and integrated with the fire suppression system and equipment layout as shown on the architectural backgrounds.

Controls are defined for all equipment in the 90% set, including completed points lists and sequences. Coordination with electrical should be completed for all aspects of the design, be it circuiting support for high-amp electric humidifiers located in nonmechanical spaces, control panels that require a UPS power circuit, or control of a restroom fan by a wall switch that will be installed under the electrical contractor's scope. When custom control sequences are defined, care should be taken to carefully check them and ensure they are complete, correct, and simple enough to be properly implemented. In a critical environment, the limiting factor on controls logic should not be what the specified system can do, but rather what is the minimum it must do to provide the required reliability, control, and efficiency. As the last portion of the construction completed, flaws and errors in the control sequence can lead to delays and costs late in the critical final stages of the construction calendar when both time and contingency funds are often exhausted.

The production of details is a major task to complete the 90% set. Details show the exact construction and installation designs for equipment and mechanical components. Where details are pulled from the designer's predeveloped library, care should be taken to ensure they are applicable to the design. It can be confusing to include details for how to hang ducting from a concrete slab when the building is a single-story structure with steel roof trusses and inappropriate to include steam trap details for a building with no heating requirements at all. If distribution or mechanical room layouts are tight with significant coordination and interaction with other trades, sections and room layouts are appropriate. While piping layouts can be fully defined in plan view by noting bottom of pipe elevations or heights AFF, carefully selected section details tend to reduce field confusion and catch more interference problems in the drawings rather than in the field. Section details can also be valuable to ensure that air distribution plenums are not being clogged by mechanical, fire, electrical, and architectural elements.

10.3.4.11 Permit Drawings Drawings submitted for building permit should be as complete as possible. Depending on the jurisdiction, changes between the permit drawings and final construction drawings may need to be noted by revision bubbles on the set—cumbersome bookkeeping if there are extensive changes. Depending on the schedule, the 90% drawing set may be used as the permit drawing set. If a separate permit set is produced, it usually differs from the 90% set by including permit-specific forms (sometimes inserted into drawing sheets). Time requirements often also result in it being less complete, with control sheets often neglected since they tend to have few code requirements. Details such as seismic bracing of ductwork, fire smoke dampers, fire alarms for air handlers (or the justification for omission of them), outdoor air ventilation rates (as low as possibly allowable), and other code-related design aspects need to be included and complete. Sections and large-scale room layouts dimensioned for construction layout are less important in the permit set.

The permit set is the completion of the code compliance research and design that occurred throughout the entire design process. Notes and narratives from the schematic and detailed design phases should be referenced, and any code concerns that were raised should have their resolution clearly shown, noted, and specified in the permit drawing set. Code concerns that were not raised in communication with officials may benefit from not being highlighted in the interest of keeping the set clear and concise for review.

10.3.4.12 Bid Package: Construction Drawings and Specifications The bid drawings include the final design drawings and specifications. There should be few changes from the 90% set, and any signification changes should be explicitly coordinated by phone, email, and/or meeting with all affected disciplines. A final review of the specifications to ensure they are complete and applicable may result in additional changes. To ensure completeness, at a minimum, all equipment noted on the schedule will be represented in the specifications, all distribution systems will have installation and accessory information in the specifications, and every control point type will be fully described in the specifications. Changes that occur after the bid set is released can be costly; while it is inevitably difficult to find time and budget in these late stages, a final quality control review at least 3 weeks prior to the release of the bid set is a must for all but the smallest projects or the largest contingency budgets (and most understanding—often incredibly rushed—client).

10.3.4.13 Bid Support During the bid period, requests for clarifications from contractors may be submitted. Following the protocol set by the client, the mechanical designer should be prepared to promptly offer written responses as required. Bidders will be looking to assess the lowest cost options to satisfy the design—which is in the interest of the owner, as long as the cost savings do not reduce the reliability, redundancy, and operational capabilities desired for the data center.

In some cases, due to time constraints, the bid package is released incomplete with additional addendum packages released prior to the bid due date to complete the design documentation. If any mechanical scope needs to be included in the addendum, it is critical that the design team leader (typically the architect) knows what to expect and incorporates the mechanical materials in the addendum. Addendums may also be used to respond to bidder questions that reveal ambiguity in the design documentation or last minute cost reduction opportunities.

10.3.5 Construction Administration

The design job does not end until the data center is properly operating. Construction administration is a significant time demand and critical to a successful project. The mechanical designer provides submittal review of equipment selections, site inspections to determine if installation requirements are being met, interpretation of the design documents when questions arise, quick correction of design ambiguities (or outright errors), solutions to interference problems, support for commissioning, and final inspection to ensure correct installation. None of these tasks differs significantly from any other mechanical design, other than the high reliability demand of the data center that increases the importance of delivering a fully debugged system on day one.

10.3.5.1 Submittal Review Submittal review ensures that all equipment meets the design requirements. A methodical approach should be taken to check the submitted equipment against the drawing schedule information and the specifications. When the submitted equipment matches the basis of design, the submittal review is primarily limited to verifying that the correct configuration and options are specified. Substitutions require more in-depth investigation to ensure they meet the letter of the design documents—as well as any design requirements that were not explicitly included in the design documents but assumed as a standard equipment feature that was included in the basis of design equipment selection.

10.3.5.2 Site Inspections Regular site visits should focus on verifying that equipment and distribution are being properly installed. Contractors sometimes will install equipment as they have done in the past rather than as dictated on the design drawings. This can be an advantage in some cases, where an experienced contractor can compensate for design documents with weak specifications or drawing detail. But it can cause significant trouble if the contractor is not familiar with data centers and begins to make incorrect design interpretations, perhaps removing redundancy to save money or placing dry coolers closer together to reduce the size of the mechanical yard but at the cost of harming extreme day cooling performance. And as with any project, there is always the need to ensure that the quality of installation meets the design requirements. The sooner an incorrect installation technique is caught and corrected, the lower the potential for adverse schedule impacts.

10.3.5.3 Design Interpretation The design documents will fully describe the mechanical system and how to install it. But for more complex mechanical systems or ones that have unusual systems, there can be value in the mechanical designer discussing the design intent directly with the installation contractors. Caution is required to ensure that all parties understand that nothing in the discussion represents any approval for deviation from or addition to the contract documents scope. Casual discussions during site visits risk misinterpretation as being changes made (with cost implications) to the contract documents. Set meetings with documented meeting notes that clearly state no design changes were implied or approved at the meeting can be useful. For example, if an unusual piping design is used to provide piping redundancy, a 30 min meeting at the job site with the pipe fitters to describe the intent can ensure it is installed as shown. Control sequences are another area where direct discussion between the contractor and design engineer typically saves more time than it consumes; a half day spent reading through the control sequence and ensuring the actual programmer understands the intent can save considerable time versus correcting erroneous assumptions when they show up as failures during commissioning.

10.3.5.4 Design Modification The installation offers the final word on whether all interference and coordination issues were properly resolved during the design process. There is no ignoring a column that is directly in the way of a pipe-run or a gravity-driven condensate drain line that hits the drop ceiling as it "slopes down to drain." Interference issues are often corrected by on-site meetings, but more significant problems (that often carry cost impacts) require written responses to Requests for Information. Promptly correcting any problems found is critical to minimizing the disruption and potential delay. Lead time issues may also come up during construction, which may require the mechanical designer to provide substitute equipment options. When a tight construction schedule is expected, lead time concerns are another key reason to try to ensure that multiple vendors are available for all critical pieces of equipment.

10.3.5.5 Commissioning Commissioning is a methodical testing of the installed systems. Theoretically, commissioning should be unnecessary: if the design is installed exactly per design intent in all respects and all equipment functions perfectly, it is not needed. In practice, commissioning is a very important process to ensure that the design is properly installed, that the design meets the requirements, and that the system will operate without failure in all anticipated conditions. Support for commissioning is often integrated deeply into the fundamental design of the system. Test ports provided into piping, duct access doors at key points to allow inspection of dampers or turning vanes, and the requirement for extensive trending capability in the central control system are all common accommodations made for both commissioning and ongoing system maintenance.

The mechanical designer will need to ensure that the commissioning agent is fully informed of the design intent, in particular the control sequences, interior design conditions, and outdoor design conditions. If no commissioning is planned or budgeted for by the overall project, the prudent mechanical designer (and experienced data center contractor team) allows time and budget to perform a targeted commissioning of their own. Without the active testing of the system provided by commissioning, there is a risk of system failure well after system start-up as internal loads change and external weather varies.

Commissioning is ideally performed only after all systems are completed and operating. Trouble can arise if it is performed prior to system completion; for example, use of load banks to test the capacity of a UPS system before the UPS room cooling system is operational can lead to such extreme room overheating that it can pop sprinkler heads.

10.3.5.6 Final Approval The final inspection generates a punch list or list of problems that need to be corrected before the system installation is deemed completed per design documents. It is critical to note that the common punch list is no substitute for commissioning. Commissioning applies in-depth active testing to ensure that all systems meet design intent under all expected operating conditions. A punch list is generally based on a passive inspection of the installation and balance reports to verify that everything looks like it is installed per requirements.

Any changes to the design are collected and applied to the construction drawing set to provide a final and accurate as-built set of design documentation to the owner. The as-built documentation is critical to support the ongoing operation of the system as well as any future modifications or build-out. As-built documentation collects information from the field into the design documentation and needs to be completed before the construction team dissolves.

After all items on the punch list have been corrected and the as-built has been delivered, the mechanical designer signs off that the contractor has met their contractual obligations and their role in the construction project is completed.

10.3.5.7 Postconstruction Support The traditional designer's scope ends after project completion, but there are often continuing services that would benefit the owner. The designer is in the best position to offer recommendations for how to build out and operate the system to the highest possible efficiency. Questions regarding changes in the intended loading, slowed or accelerated build-out approaches, and operational optimization can all benefit from direct designer input. There is also often opportunity to improve the system efficiency by tailoring operation to the actual IT load and installation; once the system is constructed and operating, it can be optimized to the actual conditions rather than design assumptions.

A well-trained site staff can operate a properly designed system well, but the ultimate expert in the system will be the designer of record. Keeping them involved at a low level of review and comment can significantly improve operation. Enabling the remote monitoring of the system can help make this involvement an economical option.

10.4 DATA CENTER CONSIDERATIONS IN SELECTING KEY COMPONENTS

Most data center cooling systems rely on a number of common cooling components. While data center cooling applications have strict reliability requirements and a unique load profile, these can often be met by the proper application of common commodity mechanical equipment. When selecting components for data center use, all the standard selection concerns and approaches apply with additional considerations such as the following.

10.4.1 CRAC (Also Known as Computer Room Air Handler)

CRAC, as indicated by their name, are specifically designed to provide cooling to data centers. Integrated controls are typically provided, the system is designed to carry the sensible load typical of a data center (i.e., a far higher airflow per ton of cooling provided than typical), and reliability is a primary design concern.

Many CRAC offer the option of a reheat stage. Reheat is typically used to prevent overcooling a space when a system is in dehumidification. As most data centers are not concerned about overcooling, reheat is usually an option best eliminated; it can be a surprisingly expensive energy consumer and offers little benefit. It is exceedingly common to find in operating data centers that reheat has been disabled by data center operators due to its expense and very justified operator confusion at the purpose of electric heaters operating in their high cooling demand facility in the midst of hot, humid summer weather.

Humidifiers incorporated in CRACs are another area of operational concern. While often necessary to meet client humidity control requirements, they tend to be high-maintenance components. Humidity sensor drift also commonly results in facilities with multiple units "fighting," that is, one unit may be humidifying, while a literally adjacent unit is dehumidifying—a significant adder to maintenance requirements and energy waste. A shared control system for all humidifiers in a space is often appropriate to prevent this situation and reduce the number of sensors that inevitably require frequent service (calibration or replacement).

System efficiency at part load, which is not always publicized by manufacturers, is the critical parameter in determining the operating cost of a piece of equipment in a data center system design. A data center with all units running at full load is very rare due to the designing in of redundancy and surplus capacity for the future. Since the systems will operate at part-load capacity in the majority (if not all) of the time, if the selection of CRAC is to consider the operating cost, carbon footprint, or other common efficiency metrics, then the part-load-system efficiency must be defined for an accurate analysis.

10.4.1.1 Chiller Plant The chiller plant efficiency can be significantly improved by recognizing and designing specifically to meet the sensible nature of the data center load. It can be a major operating cost benefit to recognize that the vast majority of the cooling load from a data center is sensible only—no latent load requiring dehumidification is generated by the IT equipment and ventilation rates are negligible—and to serve that load with a plant optimized specifically for that regime, a medium temperature chilled water plant operating at 55–60°F (13–16°C) or even higher. Mechanical cooling equipment operates significantly more efficiently when the temperature difference between the evaporator temperature and condenser temperature (dry coolers, condensing coil, or cooling towers) is reduced, typically reaping far more energy savings from reduced compressor load than the fan or pumping cost associated with using a higher-temperature and lower-temperature delta cooling medium loop. A small conventional air-cooled system can be dedicated and optimized to provide the small amount of dehumidified outdoor air required to maintain space pressurization.

10.4.1.2 Air-side Economizer Use of an air-side economization cycle, which is bringing in outdoor air directly when the outdoor air is cooler than the space requiring cooling, offers tremendous energy savings for data centers, has a few well-known but specialized design requirements, and when properly designed improves reliability. Often initially considered for energy savings, the challenge of properly controlling the system and fear of potential contaminates from outside air are common hurdles that successful implementations overcome. Economizer systems tend to increase complexity simply by their existence adding additional controls that can fail, but they also offer an excellent source of cooling redundancy during much of the year and can be designed in a fail-safe manner.

Data centers are a perfect candidate for economization since they have significant cooling loads 24 h a day even when cool outside. In many climates, an air-side economizer can cut annual mechanical system power usage in half. There are also benefits from reduced maintenance of mechanical cooling equipment; for example, compressors and dry cooler fans will see significantly reduced run hours and have long periods of time when they are not required to operate.

Economization savings are maximized when combined with an effective airflow management regime and modern cooling set points. Current standards for maintaining a supply air temperature of up to 80°F to IT equipment inlets result in an exhaust airstream around 100°F, theoretically allowing savings from air-side economization whenever outdoor air temperatures are lower than 100°F. In practice, control sensor accuracy and humidity control concerns do reduce this opportunity but only slightly in most climates.

As with all aspects of data center design, air-side economization must be carefully engineered to provide improved reliability and control of the space. Contamination from the outdoor air is a common concern, although there is little research to support this concern. Proper filtration has been found to provide appropriate protection. Consideration of unique local conditions, such as a location directly adjacent to saltwater, intakes near diesel exhaust or other hazardous-to health fumes, odor problems that raise comfort issues, or unusual IT equipment requirements, must be understood and accommodated in the design. The large loads of data centers require large volumes of outdoor air to remove; the significant size of filtration and associated maintenance cost and accessibility should be considered during the evaluation of this design option. Humidity control is the other design concern with air-side economization. Redundant control sensors and an adiabatic humidifier system that utilizes the data center waste heat are the common approaches to ensure that an air-side economizer does not create a wasteful false humidification and dehumidification load.

Reliability of the data center as a whole can be improved by the second source of cooling provided by economization. The majority of IT equipment can operate for a period at temperatures significantly above the design temperature of a data center—temperatures that can often be maintained through just the operation of an economizer system if the primary source of cooling failed. This additional level of redundancy offered is often overlooked since any temperature excursion above the design temperature is unacceptable. However, it should be noted that an economizer system may be able to prevent downtime even if it cannot maintain the design temperature—a safety factor no data center design should ever depend on, but it does add value to the owner. The benefits of an economizer system are organically understood by most operators, with "open the doors" a common last-ditch maneuver in response to cooling equipment failure in a computer room.

The economizer is fully backed up by mechanical cooling and therefore does not require redundancy to protect reliability, but reliability can be hurt if the economizer controls are not designed to be fail-safe. The system should be designed such that no single failure, such as a temperature sensor giving an erroneously low reading, can result in *hot* outdoor air being brought in and overwhelming the mechanical cooling system during periods when the economizer should be inactive. Damper actuators controlling the economizer outdoor air intake should fail to be closed and be monitored by end switches with appropriate alarms. Redundant sensors should be used to sense outdoor air temperature and humidity and a regular sensor maintenance regime (replacement or calibration) followed.

The relative benefits and disadvantages of air-side economization versus water-side economization are discussed in the following section.

10.4.1.3 Water-side Economization or Free Cooling Use of a water-side economization system, that is, bypassing the energy-intensive mechanical compressor equipment to create chilled water through evaporative cooling alone, offers tremendous energy savings for data centers, has specialized design requirements, and has varied impacts on reliability. The greatest savings are seen in climates with a significant wet-bulb depression, but most climates offer good opportunity. Design concerns focus primarily upon ensuring reliable chiller plant staging. When properly implemented, water-side economization offers an additional source of cooling in case of chiller failure during cool weather.

Data centers are a perfect candidate for economization since they have significant cooling loads 24 h a day even when cool outside. In many climates, a water-side economizer can cut mechanical power usage in half. There are also benefits from reduced maintenance of mechanical cooling equipment; for example, chillers will see significantly reduced run hours and have long periods of time when they are not required to operate (or be rushed back into service to provide redundant standby capacity) and can have preventive maintenance performed on them.

The primary design challenge is to ensure that the operation of the water-side economizer will not result in loss of cooling from the plant when the plant is switching from chiller operation to water-side economization operation, and vice versa. The system must be designed to allow a seamless transition since, unlike office buildings, loss of cooling for even 10 min is not acceptable.

To ensure stability, the chillers must be able to start while the water-side economizer system is in operation—a challenge since in free cooling operation, the towers may be full of water at 45°F (7°C), while many chillers cannot operate stably until towers provide them with water at 60°F (16°C) or even higher. There are several possible methods to ensure chillers can start even when the towers are cool and operating at free cooling temperatures. One is the use of some form of head pressure control within the chiller to allow it to start up using the same cold condenser water as the free cooling system is using. Another common approach is to create an independent water-side economizer loop (often by temporarily isolating the redundant cooling tower capacity and providing a dedicated water-side economizer condenser water supply line) to ensure that the main chiller condenser water loop temperature can be quickly raised high enough for stable operation. Alternatively, some form of mixing loop can be configured to ensure the chillers can be provided with an acceptably high condenser water supply temperature for start-up even when the cooling tower sumps are at a free cooling temperature. Regardless of the design method selected, the designer must allow for reliable start-up of the chillers simultaneously with operation of the water-side economizer system.

Retrofitting a water-side economizer system often offers attractive paybacks but carries its own design challenges. The retrofit typically must be done with no system downtime, requiring luck in the location of access valves or, more often, use of techniques ranging from hot taps to freeze plugs to do the work while the system operates. A careful evaluation of the system should also be made to identify any existing problems that may assert themselves when the free cooling system is in operation, such as chiller problems that would prevent low load-operation or exterior piping with failed (or missing) heat trace—problems that could be unmasked by the new operational profile introduced by water-side economization.

If properly designed for a data center application, water-side economization offers a backup source of cooling for much of the cooling season. It is often compared to air-side economization. Compared to the air-side approach, water-side isolates the interior environment more since outdoor air is not brought in. But when evaluating reliability, an air-side economizer often offers backup to more failure modes; for example, a burst condenser water pipe during extreme cold weather could shut down the entire chiller plant—including water-side economizer—but a data center with an air-side economizer could remain fully operational while its entire chiller plant was down and likely remain up until repairs could be made. Which approach offers better energy savings depends on the local environment, specifically if the average wet-bulb depression is enough to overcome the lower temperature required for water-side economization due to the approaches of the cooling tower, flat plate heat exchanger, and air-side coils. Ultimately, the choice of system type requires a broad system evaluation by the mechanical designer and discussion with the client.

10.4.1.4 Humidification Humidifiers introduce a significant source of catastrophic failure to a data center. The relatively high-pressure domestic water lines are a source of flooding concern. A leak detection system and supply shutoff is recommended, along with an appropriate maintenance and testing schedule to ensure the proper operation of the system. Domestic water piping can supply an almost infinite volume of water, making it a higher risk than a chilled water piping system that typically has a very limited volume of water.

The need for humidification in data centers is an evolving debate. Based upon a building body of research, if humidification were not a traditional standard started in the age of punch card feeders, it is unlikely it could be justified for use in most critical facilities today—precedence is its primary remaining justification. The vast majority of IT equipment is protected from static discharge by chassis design. And if the static discharge protection designed into the equipment casing system is bypassed for internal work, then humidification alone does not provide an acceptable level of protection. However, precedence carries significant weight in data center design, and many clients still require humidifiers in their data centers. Despite precedence, from an engineering viewpoint, it is peculiar how wedded to humidification operators tend to be since it is far more common to hear a verifiable data center horror story involving knee-deep water under a raised floor caused by humidifier system piping or valve failure than a horror story involving static discharge.

The humidification requirement for a data center is theoretically low since there is little outside air brought in; however, operation and control problems regularly result in excessive humidification. Uncontrolled dehumidification is an expensive energy waste that is also common, particularly in direct expansion (DX) refrigerant cooling coil-based systems that often have simultaneous dehumidification and humidification due to an unnecessarily high space relative humidity set point, low dry-bulb temperature set point, and a tendency for portions of DX cooling coils to run cold. In systems with multiple independently controlled CRAC serving the same space, sensor drift over time will often result in adjacent systems "fighting," that is, one in humidification, while the other is in dehumidification; this problem is also exacerbated by tight set points, with the humidity deadband frequently far smaller than required.

Standard humidification systems are a significant operating cost in both energy and maintenance, with the partial exception of adiabatic systems. The most common humidifiers use electricity to vaporize water, a very energy-intensive task. If a low-humidity set point is used and there is minimal outdoor air (and no air-side economizer), the inefficiency of an electric humidifier may be of little net annual cost due to infrequent use.

An adiabatic humidifier that uses the waste heat of the data center itself to vaporize water offers an energy benefit by providing free direct evaporative cooling and reducing electrical demand on the generator. Adiabatic humidifiers raise maintenance and operational cost concerns in the larger sizes, where many atomizing types (ultrasonic, high-pressure water nozzles, compressed air nozzles) require significant water purification. While atomizing nozzles are a technology of choice in many critical environments that need to condition large quantities of outdoor air—such as a data center taking advantage of air-side economization—the associated water treatment plant can rapidly grow to expensive proportions and lead to a reconsideration of the simpler but more fouling-prone wetted media approaches.

10.4.1.5 Dehumidification Uncontrolled dehumidification in a data center can be a significant overlooked design challenge. A data center run with outdated but common set points, such as 70°F (21°C) dry bulb and a minimum humidity of 45%, has a dew point of 48°F (9°C). If the space cooling coil is operating below that temperature, it is possible for condensation, that is, uncontrolled dehumidification, to occur on some portion of the coil—a situation that reduces the capacity available to cool the space and can significantly increase operating costs and energy consumption. The most effective protection against uncontrolled dehumidification is to separate the sensible cooling system from the dehumidification system entirely and design the sensible cooling system such that the working fluid temperature never drops below the desired space dew point. An example would be using a 52°F (11°C) chilled water supply temperature to supply air handlers cooling the data center space and a small dedicated outdoor air system[5] with a stand-alone DX coil to provide dry air for pressurization.

10.4.1.6 Fans The use of variable speed drives for fans in data centers is becoming more popular and offers the most significant energy savings opportunities to air-based systems if properly controlled. Variable speed fan systems take advantage of redundant capacity to allow all fans to operate at lower speed even when the data center is at design capacity. The savings accrue quickly due to the cube law nature of fan power—turning fan speed down by only 15% reduces fan power consumption by almost 40%. The fan speed needs to be controlled in series with the cooling coil to ensure the fan speed is reduced, with common algorithms including controlling the fan speed to temperature sensors in the space while the coil is controlled to maintain a constant exit air temperature or sequencing the fan speed and coil output in series (turning down fan speed and then reducing coil output). Control approaches are still developing, but the key is simply to ensure that the fan speed turns down and that the coil and fan speed control loops do not interact in an unstable manner.

Traditional fan optimization techniques can also yield significant savings. For example, the common raised floor data center configuration can improve fan efficiency with features such as turning fans beneath downflow units or even the use of plenum fans lowered under the floor to directly pressurize the underfloor plenum. If a built-up air handler is used, there are greater opportunities ranging from large and high-efficiency vane axial fan systems to the use of many smaller fans in parallel configured to create a wall.

Minimizing fan power installed helps reduce generator and electrical sizing, but it also has a major impact on operating costs. Data center fans operate 8760 h/year at a near constant load, which justifies a significantly higher first-cost investment in larger ducts and air handlers to reduce fan power operating costs than would be considered for a 2600 h a year variable air volume office system; applying standard rules of thumb that have evolved from office system design to duct sizing or selecting coil face velocity in data centers will result in a working system but miss many opportunities to optimize operating costs and energy usage over the life of the facility.

[5] In some cases where first cost is a driving factor, humidity control and positive building pressurization are not deemed a critical system that requires full $N+1$ design, and only a single outdoor air unit is provided.

10.4.1.7 Cogeneration Cogeneration is typically in the realm of the electrical designer, but it crosses into the mechanical design when the waste heat is used to generate cooling (sometimes referred to as trigeneration, even if there is no creation of heating from the plant). The use of an on-site cogeneration plant to power the data center and drive a cooling plant can offer a compelling story, but the business case can be difficult. Local incentives, the marketing aspect of a "power plant on-site," and the specifics of how it aids redundancy and reliability are all key factors that the mechanical engineer can assist in defining.

10.5 PRIMARY DESIGN OPTIONS

We will discuss the most common system approaches to four key areas of data center design: the cooling medium, heat rejection method, air delivery path, and air management. The selection of cooling medium defines whether the system will be moving heat primarily using airflows or water flows, which has profound design implications all the way down to the client's IT equipment selection in some cases. The heat rejection approach influences the equipment options for the mechanical system. The last two areas apply primarily to air-based cooling system: the delivery path used for supply and return and the air management system used to avoid hot spots and efficiently collect waste heat.

These are only the most common current approaches; there are many other configuration options for an analysis-driven design to assess and potentially pursue. Designs that represent evolutionary changes, the development of fundamentally new system approaches, are strongly incentivized by the magnitude of operating cost savings possible from more energy-efficient design approaches (and, at the moment, a few Internet behemoths with the savvy and design budgets to vigorously pursue those cost savings).

10.5.1 Cooling Medium

10.5.1.1 Air from CRAC CRAC are specialized air handler units available from a number of manufacturers. Placed directly on the data center floor, they are an imposition on the finished data center floor space. They often include integrated controls and options for humidification and dehumidification that are directly tailored to the data center environment. With capacities typically ranging from 5 to 50 tons each, they offer the simplest design option—essentially requiring little more than following the catalog selection process. While the use of computer room air conditioners (often referred to as CRAC units or sometimes CRAH units—Computer Room Air Handlers—when chilled water based) offers a predesigned system in many ways, there are additional mechanical design details required including pipe routing for the cooling fluid (be it refrigerant, a glycol/water mix to and from external dry coolers, or chilled water from a central plant), humidification piping or system, condensate removal drain system, any associated ducting, service of support areas, layout of interior and exterior (heat rejection) units, correct specification of control options, and fire system integration (including any required purge ventilation). A traditional design approach, they are favored for the relative simplicity of design and the redundancy of having multiple units. For large facilities, maintenance can become costly due to having many small distributed fans, condensers, compressors, and other components.

A recent evolution of CRAC is the in-row unit. An in-row unit is designed in the same form factor as a standard IT equipment rack and is made to be installed directly adjacent to IT racks. This can significantly simplify the air management design and often offers variable speed fan capabilities that further improve the efficiency. Currently popular for small data centers and as a retrofit to address hot spots in larger facilities, it is an interesting design evolution in the area of CRAC.

10.5.1.2 Air from Central Air Handler A central air handler system can offer the opportunity for a more customized and lower-cost system but requires more design effort. The air handler itself must be carefully specified; in particular, it must be sized for the almost all-sensible load expected from a data center. Extensive control system design is needed to provide appropriate data center control and robustness (no single point of failure, be it a single controller or a single control point such as a common supply air temperature sensor). Layout is typically more complex, but the data center floor is kept clear of most if not all mechanical equipment. For larger facilities, the flexibility, efficiency, and cost-saving potential of using central air handlers rather than the prepackaged CRAC options often outweigh the significantly greater design effort required.

Systems integrated into the building, where entire plenums may be combined with wall-sized coils and fan systems to essentially create massive built-up air handlers, have resulted in some elegant and highly efficient central air handler designs.

10.5.1.3 Liquid to Racks The most efficient method of moving heat is using a liquid like water, which has a volumetric heat capacity over 3000 times greater than air. Some data center equipment takes advantage of this to cool equipment directly with water, either through a plumbed heat sink system or (more commonly at this time) a cooling coil integrated directly into the equipment rack that cools hot exhaust air directly as it is exhausted. This approach requires piping in the data center footprint, raising concerns for some clients who are very concerned about liquid leaks onto IT equipment, although leak problems from chilled water supply systems designed for this kind of industrial environment are very rare (liquid process cooling loops are the norm in many critical facilities, such as semiconductor cleanrooms or pharmaceutical laboratories). Locating the distribution piping in a manner that allows free rolling of equipment around the data center floor, provides for future addition and repositioning of equipment, and meets redundancy requirements is a key consideration when selecting equipment. The use of liquid cooling offers tremendous potential for efficiency improvements and can allow high-power densities easily, even in spaces with very little height and space to move air. The greatest efficiencies are achieved by closely integrating a water-side economizer system and minimizing heat exchanger steps between the heat rejection and the coil itself to minimize the resistance to heat rejection (the combined approach of all the components between the heat sink/outdoor temperature and the IT equipment hot airstream).

With most liquid cooling applications, there is still a need for a small air-side system to carry the minor heat load picked up from convective and radiative heat transfer through the sides and tops of the racks, as well as lighting and any shell loads. A humidity control system must also be provided since liquid-cooled racks cannot provide dehumidification at the rack level. Depending on the liquid cooling system cost per rack, it may not make sense to provide low-power racks, usually under 2 or 3 kW, with water cooling—their load is another that usually is handled by an (much smaller than typical) air-based system.

10.5.1.4 Others Some manufacturers offer systems that use refrigerant loops to coils in the rack. With most of the same design concerns (piping layout to racks, flexibility, ambient system for humidity control), these systems are essentially liquid to racks. They do address concerns about liquid piping on the data center floor since any leak would flash to a gas without harming any data center operation but often at a significant cost and efficiency penalty.

Research is always continuing into innovative methods of transferring heat out of racks, ranging from bathing the servers directly in a cooled dielectric fluid to using very long heat pipes or solid heat sinks to produce entirely passively cooled systems. These methods are currently not feasible and/or available, but the mechanical engineer should remain vigilant for unexpected changes in both the form and temperature of the load itself and mechanical equipment options to move it.

10.5.2 Heat Rejection

10.5.2.1 Dry Cooler A common method of heat rejection to the outdoors is a dry cooler, which consists of a coil and fan unit placed outside (similar to a car's misleadingly named radiator, which works almost entirely through forced convection). Dry coolers can be used to cool a liquid condenser loop or to condense refrigerant directly. As dry heat rejection systems, they are theoretically limited to cooling to a temperature above the outdoor dry-bulb temperature and in practice tend to be the lowest efficiency heat rejection option. They do offer a low-profile solution and do not require any significant operation maintenance. Design can also be very simple, with everything but choosing options completed by the provider of an associated CRAC or, in the simplest possible solution, the dry cooler is integrated into a traditional packaged rooftop air handler system (properly selected to deal with the sensible nature of the load).

10.5.2.2 Open Cooling Towers The most effective form of heat rejection is typically an open cooling tower. Open cooling towers have the advantage of using evaporation to increase their capacity in all but the most humid climates, providing a more compact footprint than dry cooling-based systems. They are dependent on water, so for reliability reasons, backup makeup water is stored on-site, sometimes in the form of a sump that holds several days of water; at a minimum, enough makeup water to account for evaporation during design load for the same length of time as on-site diesel storage will run the generators is required. Freeze protection is a significant concern during build-out and operation. During system start-up, the level of waste heat from the facility may be vastly lower than at design, creating a period where the towers freeze due to the lack of waste heat from the data center load. During operation, any piping that is allowed to sit without flow, such as redundant piping or piping bypassed during free cooling operation, may become a potential single point of failure for the whole system and should be freeze protected and possibly isolated as necessary to ensure system reliability design goals are met.

10.5.2.3 Air-side Economizer An air-side economizer system rejects heat by exhausting hot air and bringing in cooler outdoor air. Air-side economizer is a tremendous opportunity to improve many data center designs as long as it is carefully designed for data center use. Having the ability to condition the space with outdoor air provides a degree of redundancy: if the compressor-based cooling system fails, the data center temperature can at least be maintained at around the same temperature as the outdoors. In most cases, this will allow the equipment to remain operational during emergency repairs even if temperatures may exceed the recommended operational temperature envelope.

Concerns of contamination by outdoor air are addressed by proper filtration, designed for ease of regular filter replacement. Recent studies have indicated that moderate filtration is more than adequate to eliminate particulate concerns. Some sites may have unique pollutant concerns, such as ammonia gases near agricultural land or forest fire smoke, but there is little hard evidence suggesting outdoor air contamination beyond that dealt with through filtration is a concern. However, reality can take second place to client perception, so any implementation of an air-side economizer should be carefully vetted and approved early in the process to avoid wasted design effort on a system the client just does not want.

Humidity control is an important part of economizer design. Dry, cold winter air will usually require humidification to meet data center requirements. An adiabatic humidification system is strongly recommended to minimize operational costs. Adiabatic systems can be configured to use the heat energy in the hot return airstream from the data center itself to evaporate the water needed for humidification. There are many adiabatic technologies available, including wetted evaporative media pads, atomizing sprayers, and ultrasonic. The ultimate system selection should balance operational cost (including any water treatment requirements and auxiliary pump or compressor power), maintenance requirements, and reliability. Electric steam humidification systems will have comparatively high energy costs at the volumes of air that economization introduces into the space and may negate the energy savings of economization entirely.

Another aspect of air-side economization is ensuring that pressurization control is maintained. The large volumes of outdoor air introduced into the data center must have a proper exit path provided. Poor pressurization control can lead to doors that do not close properly—an inconvenience in commercial construction but an unacceptable security risk to most data center customers.

As an efficiency feature, economization systems typically do not require redundant design. However, they do require proper fail-safe design. Outdoor air economizer dampers should fail to a closed position. At a minimum, including redundant outdoor air sensors is needed to ensure a sensor failure does not result in hot outdoor air being allowed in during the summer. Also advisable are end switch feedback from damper actuators to alarm stuck dampers and control logic that is biased toward locking out economizer operation when its benefit is in doubt.

10.5.2.4 Others There are other, less common methods of heat rejection. A geothermal system that rejects heat to a large ground closed-loop heat exchanger is theoretically an option; however, it is usually quite expensive in the scale required by a data center and poorly suited to the year-round, cooling-only operation of a data center. A geothermal system that rejects heat to surface water can be a very good option but requires excellent siting adjacent to an appropriate body of water. Integrating the data center into a larger facility that requires heating so the data center's waste heat can be rejected to a building that requires heat or even a district heating loop can be a very successful and efficient approach. Note that the temperature of heat coming out of a data center tends to be quite low and air or water from the data center floor in the range of 75–85°F (24–29°C) is common, so a heat pump or other compressor-based system is often used to boost it to make it more useful for reuse. There are also options that combine evaporative cooling with a dry cooler, such as closed-loop cooling towers that can be appropriate in some situations.

10.5.3 Air Delivery Path

10.5.3.1 Underfloor The most common traditional data center air delivery path is through a raised floor. Several manufacturers provide raised floor systems consisting of 2ft × 2ft square tiles placed on pedestals. The underfloor plenum is pressurized with conditioned air, and perforated tiles are placed where air supply is required. The tiles can be easily picked up and moved, providing a great deal of layout flexibility.

The height of the raised floor is usually specified based upon the amount of airflow it must accommodate, which defines the free area needed under the floor (free area height is usually a smaller number than the nominal floor height due to the height consumed by the tile itself). It is critical that the mechanical engineer carefully consider the amount of underfloor free area height that will be blocked by electrical conduit, data cabling, and any other infrastructure intended to be routed underfloor. It is quite common to find data centers suffering from poor air distribution (manifesting as localized overheated "hot spots") due to the underfloor airflow being blocked by bundles of wiring, electrical raceways, and mechanical system piping. Underfloor plenums are not magic; adequate free area must be provided and maintained to allow for design airflow.

While the common approach to provide underfloor air delivery is to use a pedestal and tile system, it is not unheard of to take a more industrial approach and provide air supply up from a mechanical subfloor (sometimes shared with electrical infrastructure) through the structural floor of the data center. This approach can provide very low pressure drop and controllable air distribution and be a good fit if a retrofit space or clever architecture (and/or high pedestal and tile system costs) provides an economical way of constructing the subfloor. It also can offer a great deal of operational flexibility, with the ability to add water supply piping and ducting, add cooling equipment, or make other significant modifications on the mechanical subfloor level while the data center is in operation.

10.5.3.2 Overhead Ducted Delivering air overhead can be a quite effective approach. With the volumes of air motion required by most data center loads, the impact of natural convection flows is almost irrelevant, so it matters little if air is blown down from overhead versus up from a raised floor. Overhead ducting can integrate well with central air handling systems. Note again that data centers typically have high loads and large volumes of air in motion, which can result in the need for very large ducting, ducting that must also coordinate with cabling, lighting, and fire suppression. Care should be taken to optimize the ducting side for operational costs and future flexibility; undersizing ducting trades first-cost savings (which can be small if you must upsize fans and electrical infrastructure to support higher duct pressure drop) for a lifetime of higher operating costs. As data center loads rise, ducting can become prohibitively large and blur the line between ducting and plenums.

A ducted approach can carry the penalty of limiting data center flexibility. The locations of air supply need to be fixed during the initial construction and cannot be moved easily during future operation if the locations or footprints of the racks change. The initial ducting design can provide some allowance for future flexibility by sizing and laying out ducting in a manner that allows large increases in airflow with minimal pressure drop cost as required to meet future movement of load in the space. There are also creative approaches that can be used to stretch the flexibility of a ducted approach, including easily reconfigured fabric ducting or a methodical inclusion of taps and distribution paths to add future duct runs.

10.5.3.3 Overhead Plenum An overhead plenum is another option for distributing air to the space. The most common approach is to use a standard drop ceiling to create a return air plenum. This can be a very effective and low-cost approach that integrates well with many air management design approaches. Return grills are located over hot aisles and the air handlers are configured to draw return air from the plenum space. Exhaust fans and adiabatic humidification systems can be added to the plenum as part of an air-side economizer system (and serve double duty as fire suppression gas exhaust fans if necessary).

Without the structural need to carry the weight of the racks and provide a floor surface, overhead plenums can usually be much more generously sized than underfloor plenums. The overhead plenum can grow to become a de facto mechanical interstitial floor that houses mechanical equipment and provides access for maintenance without intruding into the data center floor.

10.5.3.4 Through the Space It is difficult to beat the simplicity of supplying cooling air by simply blowing it through the space toward the racks. Oddly, this design approach can be a very poor approach with hot spot problems and low efficiency or very well-performing with excellent control and efficiency—it all depends on the airflow management separating supply and return.

With fully mixed airflow management (also known as no airflow management), air delivery through the space typically reduces usable cooling capacity by haphazardly mixing hot exhaust air with the supply air. But it is extremely low cost and easy to implement in nearly any space, requiring little more than placing a CRAC unit on the floor and hooking it up. If the expected equipment loads are low, then the simplicity and low cost of this approach can be compelling.

With good airflow management, a through-space approach can be extraordinarily elegant. If all hot exhaust air is collected as part of a heat exhaust, the access walkways provided for moving equipment and operators through the space can serve as very low pressure drop supply air ducting. While a simple concept, this optimized airflow management integration can sometimes require significant analysis to ensure that it will provide reliable distribution during real operational conditions such as when racks are moved in or out. At the volumes of air movement required, computational fluid dynamics analysis may be recommended. While good airflow management systems usually pair through-space supply with a hot air return plenum, through-space air delivery can be paired with creative air management containment systems to provide a comprehensive single-floor air supply and return solution by partitioning the space into strips of hot through-space return aisles and cool through-space supply aisles.

10.5.4 Airflow Management

10.5.4.1 Fully Mixed Low-load data centers may have airflow similar to the standard design used for office spaces. A standard conditioned office space throws conditioned air supply throughout the space, mixing the cool air throughout to dilute the heat being generated in the space. In data centers, it is typically done through the use of CRAC units that throw air out from the top of the unit and draw it back in from the lower face of the unit—with little to no ducting required.

This approach typically has poor to acceptable performance but is very easy to implement. However, while simple, a fully mixed approach is limited in the capacity it can serve due to its inability to address localized hot spots. A prohibitive amount of airflow (equipment size and power

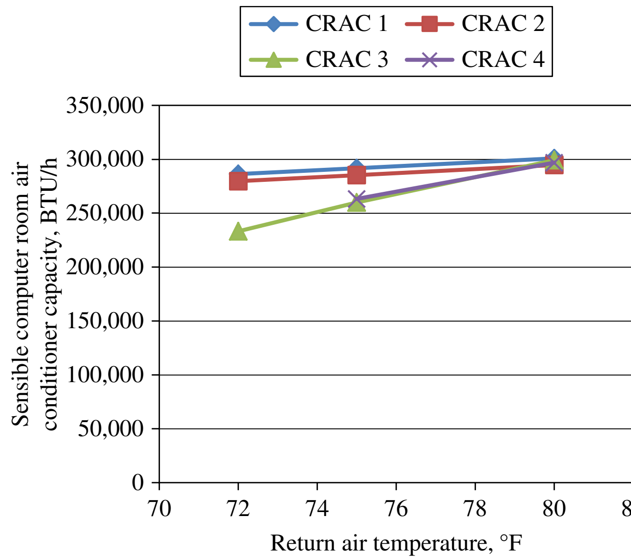


FIGURE 10.1 Impact of return air temperature on Computer Room Air Conditioner cooling capacity varies by specific model but can be significant.

usage) is required as loads increase. It is simply less efficient and effective to dilute hot air exhaust from IT equipment than to remove it directly. The dilution approach also impacts the effective capacity of the cooling units since the return air temperature is usually only as high as the desired space temperature set point, limiting the temperature differential possible across the coil (Fig. 10.1). Data centers with high loads rarely use this approach because it does not work with high loads.

10.5.4.2 Balanced Distribution Recognizing how IT equipment uses air for cooling, the distribution system can be designed to supply cooling air directly to the IT equipment intakes in approximately the quantity required. The return can then be configured to collect the hot exhaust air. For example, a design that uses a raised floor as a supply distribution plenum places the perforated supply tiles at the front of IT equipment, where the cooling air is drawn into the rack, and return is drawn through the space hot aisles or a ceiling plenum to the air handler. Another approach would be to use in-row air handler units that monitor the inlet temperature of adjacent racks and modulate cooling supply airflow to attempt to match the volume of air the IT equipment requires.

This approach is effective at reducing hot spots and improving system air temperature difference. The main limitation is the difficulty in balancing supply volumes to match the IT equipment air demand on a localized basis. If too much air is supplied from the tile in front of an IT rack, then some will bypass to the return and be wasted airflow that provided no useful cooling. But if not enough air is supplied, then hot air could be recirculated, sucked in from the IT exhaust area, usually into the pieces of equipment located at the top of the rack. In practice, it is very difficult to get and maintain this balance, but this limitation is primarily one of cost—operating cost and equipment cost due to capacity loss to bypass. A balanced distribution system can carry high loads if it is sized properly.

10.5.4.3 Partitioned For the higher IT equipment loads seen in modern data centers, control and efficiency can be significantly improved by capturing the hot air exiting IT equipment before it is mixed in the room. In high-density data centers, capturing the hot exhaust air is often the only way to ensure that it does not create a hot spot and overheat adjacent equipment. To control airflow, solid partitions can be created using anything from plastic curtains to solid walls. Fire codes and the need for future flexibility need to be considered in the design of partitions. If air distribution is being directed through the space by partitions, operation conditions such as an aisle being blocked by a new rack being moved in must be accommodated by design.

Partitions can eliminate the concern of localized overheating due to recirculation of hot exhaust air into the cooling intake area. Air balance is still a concern but is easier to achieve than in a Balanced Distribution approach since it is now averaged over a larger number of racks and area.

10.5.4.4 Rack Level Rack-level management of the hot exhaust can contain the hot air exhaust within the IT rack itself. This can take several forms, from incorporating a cooling coil directly in the rear door of the rack to a small traditional air conditioner integrated entirely in the rack itself, to a ducting system that collects hot air and directs it up a "chimney," and to an overhead return. These systems can carry very high loads without hot spots overheating adjacent systems. The power consumption associated with the system, in the form of internal fans, pumping, compressors, or other associated equipment, should be considered in the electrical system design and operational maintenance and power costs.

10.5.4.5 Active Fan Tiles The advent of low-cost digital controls and practical small variable flow fans has seen the rise of another approach to air management: supply fans integrated into individual 2 ft × 2 ft raised floor tiles that continuously vary the supply rate to match the demand for cooling airflow to a very local scale. These systems typically sense the temperature of air being drawn into the adjacent IT intakes. If the temperature is above set point, hot air is being recirculated and the active fan tile increases the speed of the integral fan(s) to increase the volume of cool air being supplied from the underfloor. The active and direct sensing of recirculation can make this an effective and robust method of air management.

10.5.4.6 Hybrid Using a single airflow management design often has long-term operational benefits from having a common operational and maintenance profile, but it is not a requirement. Mixing airflow design approaches is often the best solution when a data center is expected to carry IT equipment with significantly different load concentrations or airflow characteristics. Many of the aforementioned airflow methods may be combined within the same data center to achieve the best balance of system flexibility, reliability, first cost, and efficiency. A partitioned data center may utilize rack-level cooling to carry a few unusually high-load IT equipment racks. Legacy equipment such as a low-load old tape library system may be conditioned in a fully mixed airflow portion of data center floor, while high-load modern IT with standardized front-to-back airflow uses a hot/cold aisle arrangement with full partitions. Active fan tiles can offer short-term "fixes" to existing data centers with underfloor air supply plenums and hot spots while a severely congested underfloor plenum issue is addressed.

10.5.4.7 Future The airflow management design configurations discussed here range from common legacy to current leading-edge approaches. However, there is little doubt that new airflow management variations will appear in the future. New approaches should be evaluated on the basis of their reliability, flexibility, and efficiency. The served IT equipment needs should also be continuously evaluated; beyond efforts to push the maximum operational temperatures up, some new server designs incorporate integrated convective spaces, heat pipes to a common backplane heat sink, and other unusual features that radically change the airflow management requirements and opportunities. While the principle of isolating the waste heat exhaust from the cooling air intake is as fundamental as drawing drinking water from a river upstream of where raw sewage is dumped, the details of the various airflow management techniques to achieve this are still maturing.

10.6 CURRENT BEST PRACTICES

Data center design continues to evolve and is ultimately dictated by the needs of the client, but it is possible to identify current best practices. Good data center design must meet the specific needs of the location and the client and may not achieve all of these best practices, but it should consider all of these approaches.

10.6.1 Redundancy

Redundancy is a defining feature of data center design. Providing $N+1$ redundancy for all critical components is standard best practice. However, a best practice design will fully define and document the owner's needs through the design process. Opportunities to reduce the redundancy are sometimes available and appropriate to reduce construction and operation costs. Design decisions such as not providing fully redundant chilled water piping, designating a portion of the data center as nonredundant (most easily verified by the omission of UPS to the IT equipment), and depending on temporary rental equipment to provide servicing redundancy are examples of redundancy reductions that may occur through a best practice design process. Adding an air-side economization system to a data center or a cogeneration system with on-site fuel storage in addition to emergency generators is an example of design decisions that may be made to add redundancy beyond the standard $N+1$.

10.6.2 Reliability

The reliability impact of all system design decisions should be fully evaluated, rather than depending solely on assuming replication of past practice is adequate. The reliability benefit of free cooling can be significant and should be acknowledged in design evaluations. The reliability risk of a humidifier (with domestic water supply piping) in the space likewise should be evaluated. All design features should be similarly assessed for reliability impacts.

10.6.3 Layout and Air Management: Hot Aisle–Cold Aisle

In a high-density data center, the layout of the IT equipment is an integral part of best practice mechanical design. The layout must prevent hot air from being exhausted from one piece of equipment into the intake of another piece of equipment. Installing the IT equipment so that the hot air is exhausted to the same area as adjacent and opposite IT equipment is commonly referred to as creating a hot aisle–cold aisle configuration. For equipment with front intake and rear exhaust, this takes the form of arranging rows so that the hot air is exhausted into dedicated hot aisles while the intake side of the rack is served from dedicated cool aisles: a hot aisle–cold aisle arrangement. Equipment with side exhaust may employ shrouds to direct the exhaust to the hot aisle or incorporate vertical chimneys to exhaust it to a common overhead "hot" plenum. Partitions can be used to create a small hot aisle enclosure or the cold aisle enclosure could be smaller, but regardless of the specific form it takes, the ultimate function of aisle containment is to prevent recirculation causing hot spots that could damage equipment. In a best practice design, the hot aisle is also capitalized on to provide a return hot airstream that is 20°F or higher than the room set point temperature (Fig. 10.2).

Creating a hot aisle heat exhaust improves a number of free cooling design options that can significantly improve reliability. The capacity of the cooling equipment is typically increased by the higher temperature return, allowing it to support a higher IT load per air handler than lower temperature return designs. Rejection of heat can also be done more efficiently at higher temperature differences—the higher the temperature of the waste heat coming from the data center, the more easily it can be rejected to the heat sink (usually the outdoor ambient environment), allowing for higher compressor cycle efficiency. A high-temperature waste heat stream also can allow a higher-temperature differential between the supply and return streams (water or air), reducing flow requirements and energy consumption. In some cases, the high-temperature return airstream not only increases efficiency and system capacity but also offers a practical source of free heat for adjacent spaces.

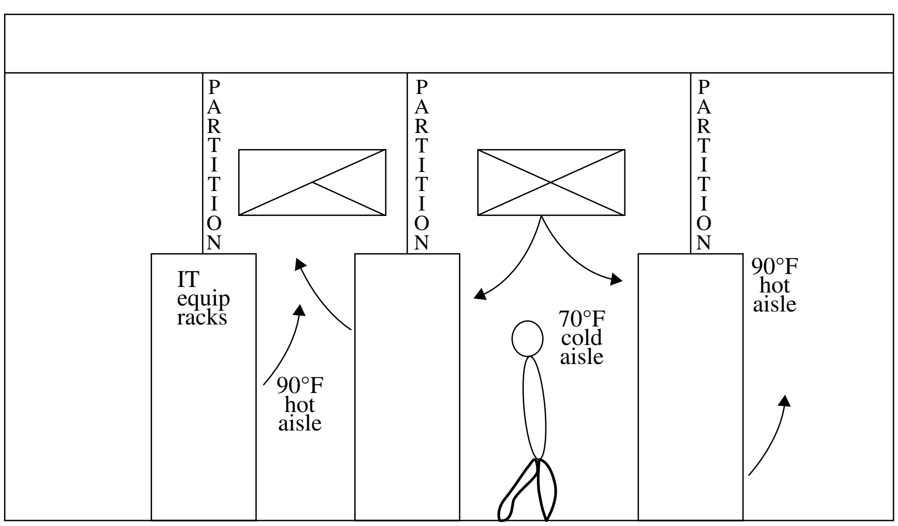


FIGURE 10.2 Hot aisle–cold aisle.

10.6.4 Liquid Cooling

Moving heat with liquid is far more efficient than moving it with air, with a small pump system having a heat moving capacity an order of magnitude bigger than an equivalent fan system. Typically, the closer to the IT rack that heat can be transferred to a liquid loop, the greater the efficiency. Liquid cooling can also offer more flexibility since a great deal of future capacity can be added by increasing a pipe diameter an inch or two; adding an equivalent amount of future capacity to an air-based system would entail adding a foot or two to the ducting dimensions.

Not all facilities are ultimately appropriate for a liquid cooling system, but it should be considered.

10.6.5 Optimize Space Conditions

With the notable exception of legacy equipment and some tape drives, modern IT equipment can allow operating temperatures significantly higher than legacy design, with an 80°F (27°C) inlet temperature within the ASHRAE TC9.9 recommended range for Tier 1 data centers. Best practice implements a high space temperature set point after implementing robust airflow management to prevent hot spots. Often, a low room operating temperature is used to compensate for poor airflow control or recirculation of hot exhaust air from the outlet of one IT rack to the inlet of an adjacent or opposite rack, space set point of 70°F (21°C) is required if there is a hot spot where the supply air is being mixed with exhaust air at 90°F (32°C) before reaching the worst-case IT rack inlet. But with proper layout and air management, such recirculation hot spots can be eliminated. With a well-designed air management system, there are no hot spots and the temperature of air supplied by the air handler is approximately equal to the temperature supplied to the IT equipment. Ideally, the supply air temperature is the space temperature and can even be set to 75°F (24°C).

10.6.6 Economization

Any best practice design fully investigates a careful implementation of economization to increase redundancy and allow for very low power cost cooling when it is cool outside—remember that even in a snowstorm the data center needs cooling. The normal operational benefits of economization are significant energy savings, sometimes reducing total data center power use by 25% or more, but the reliability benefits are potentially more valuable. It is best practice for even the most fail-safe design to consider the worst-case scenario. If the mechanical cooling system fails beyond the ability of redundancy to cover, a parallel economization system may still be functional and able to keep the space operating until a fix can be applied. Note that even if it is a summer day and economization can only hold a facility to 95°F (35°C), it may be enough to keep the housed IT equipment from dropping offline—which is a vastly preferable failure mode than the alternative of IT equipment hitting the temperature at which it automatically shuts down due to runaway overheating.

The energy benefits of economization are significant in almost every climate zone, particularly if air management offers the potential for a heat exhaust at over 90°F (32°C).

10.6.7 Cooling

For large data centers, over 1 MW, a centralized evaporatively cooled central plant is best practice—although climate and local regulations that may result in cooling tower shutdowns can dictate an air-cooled design. Smaller data centers often utilize more modular air-cooled equipment, which offers cost and control benefits but at higher operating costs. The ideal, which requires a rare combination of IT equipment, client, and climate, is elimination of mechanical cooling entirely through the use of economization and evaporative cooling.

Central control is best practice, whether in the form of a building DDC system or networked integrated controls, to avoid the common problem of adjacent cooling systems simultaneously humidifying and dehumidifying due to sensor drift over time. The system should provide comprehensive alarms to identify failures. Integrated power monitoring can pay for itself by identifying operational opportunities to reduce power bills and predicting equipment failures by alarming the increase in power consumption that often precedes them.

Best practice includes some form of economization, be it a properly filtered air-side economizer, a water-side economizer system configured to allow stable transition between chillers and water-side economizer, a dry cooler-supplied glycol coil, or a pumped refrigerant-based system integrated into some newer compressor-based systems. The added redundancy benefits of a second source of cooling (even if intermittent) combined with a constant internal cooling load 24 h a day during all weather conditions make a free cooling system best practice in all but the most severe and humid climates.

10.6.8 Humidity Control

Best practice is to minimize the humidity control in data centers to match the actual requirements of the IT equipment. ASHRAE TC9.9 guidelines offer a best practice starting range, to be adjusted to client requirements as necessary. A legacy control band of 45±5% is often requested, but control to these tolerances is rarely required and incurs significant operational and first costs. Consultation with and education of the client to select a more appropriate control band will yield the best final design product.

With the large amount of waste heat continuously available, if humidification cannot be eliminated, then adiabatic humidification is the ideal approach. Adiabatic humidification is the only solution appropriate for large humidification, such as may be seen with an air-side economizer system. Ultrasonic adiabatic humidifiers offer high precision and several market options, but an atomizing nozzle or media-based humidifier will typically be a lower-cost solution.

10.6.9 Efficiency

Accurate calculation of the system efficiency during design is required to identify and achieve the optimal design. Best practice is to "walk and chew gum at the same time" by providing a high-reliability system while also putting analytical emphasis on minimizing operating costs through higher-efficiency design. Efficiency is an often neglected area of data center design, which makes it one area where the best designers can differentiate themselves and offer premium value to the client. The operating cost of data centers is often small relative to the total revenue stream of the facility, but small efficiency improvements can offer annual energy savings of a magnitude that can often justify additional first costs.

10.7 FUTURE TRENDS

There is little doubt that data center design will change, driven by the rapid evolution of the computing equipment they house. Good mechanical design must continuously evaluate the current best practices in the field, the specific requirements of the project at hand, and the likely future demands. The IT equipment itself is a huge driver of data center design evolution for designers that pay attention to it, with changes in configuration, form factor, load density, and racks all influencing design requirements. Changes in the business environment have impacts as the size of data centers and the customization of IT equipment respond to market demands. Changes in the cost of power are continuously reweighting the balance between first construction cost and operating cost.

It is impossible to accurately predict the future, but no mechanical design or discussion of mechanical design would be complete without making a reasonable effort to identify future trends.

10.7.1 Water in the Data Center

Some data center operators do not want any chilled water or condenser water piping within the data center footprint out of concerns of a leak flooding the data center and causing a catastrophic failure. With many of the most efficient system design approaches based on a central chilled water plant approach, this can be a significant limitation on the design. Actual failure of operating chilled or condenser water piping is very rare, and many operating data centers do utilize chilled water piping across the data center floor.

Design approaches can reduce the risk inherent with water in a data center. Common techniques include placing the piping underfloor, using water sensors for alarm, and alarming any automatic water makeup system on the chilled water loop. Automatic shutoff and isolation of piping based on water sensors to automatically isolate leaks is feasible but should be approached with great caution since a false alarm may be more likely than an actual piping failure and may even become the cause of a catastrophic shutdown; even with a real leak, a data center may remain up and running with several inches of water in the underfloor but overheat and shutdown in minutes if cooling is lost.

10.7.2 Hot Data Centers

For many years, a key characteristic of data centers is that they were kept at low temperatures, cooler than office spaces. Most modern data center equipment can now operate at normal office temperatures or higher. Current design guidelines from ASHRAE Technical Committee 9.9 place the upper temperature bound for recommended Tier 1 data center operation at above 80°F (27°C). Computer chips themselves often have allowable operating temperatures of double that or more, making the current requirement for maintaining even an 80°F (27°C) data center clearly subject to future computer design. Exceptions certainly exist, with storage media more sensitive to temperature than memory chips, but the potential for IT equipment designed to operate with an ambient of 100°F (38°C)—no cooling at all, merely a (well-secured) ventilated warehouse—is easily possible using only current technology. The elimination of much of the mechanical engineer's design scope could occur through a mere market evolution demanding it, never mind where technology advances.

10.7.3 Questioning of Assumptions

Perhaps the safest prediction also serves as an overall final approach for all good quality design: traditional design assumptions will be challenged more frequently and overturned more often as data center design continues to evolve. The current maturity of the data center market has introduced more competition and a focus on finding cost reductions everywhere, including in the mechanical system design. The traditional data center configurations, such as stand-alone CRAC controlling to return air temperature and using raised floor distribution, are revered as "tried and true"—safe options in a field where reliability is job one. But the long-established legacy approaches are often not the most cost-effective or even the most reliable design for the current and future data center. A proven design that has worked in the past is a reliable option; it has been shown to work! However, the proven designs were proven on the IT equipment loads of years ago and often have implicit assumptions about IT equipment and mechanical equipment capabilities underpinning them that are no longer valid.

Often, even user functional requirements (such as the need for cabling space or control of electrostatic discharges) are incorrectly presented as system requirements (a raised floor or humidifier system). Good communication between the design team and the user group to separate the functional requirements from traditional expectations helps the mechanical designer identify the best system to meet the specific demand.

Now and in the future, the best mechanical design will be recognized for how it excelled at meeting the needs of the client, not the assumptions of the design team.

REFERENCE

[1] Mission critical facilities, data centers, technology spaces and electronic equipment. ASHRAE Technical Committee 9.9. Available at http://tc99.ashraetcs.org/. Accessed on May 22, 2014.

FURTHER READING

Data Center Mechanical. Open Compute Project. Available at http://www.opencompute.org/projects/mechanical/. Accessed on May 22, 2014.

Data center networking equipment—issues and best practices. ASHRAE, Technical Committee 9.9; 2012.

Economizer for Data Centers. ASHRAE. Available at https://www.ashrae.org/resources--publications/periodicals/ashrae-journal/features/economizer-for-data-center. Accessed on May 22, 2014.

Eubank H. et al. Design recommendations for high performance data centers. Snowmass: Rocky Mountain Institute; 2003.

Mechanical: Air Flow Management. Lawrence Berkeley National Laboratory. Available at http://hightech.lbl.gov/dctraining/strategies/mam.html. Accessed on May 22, 2014.

11

ELECTRICAL DESIGN IN DATA CENTERS

JAY S. PARK AND SARAH HANNA
Facebook, Inc., Menlo Park, CA, USA

In order to design an optimal data center, one must go through the process of determining its specific business needs. Planning and listing the priorities and the required functionality will help determine the best topology for the data center. Outlining the key ideas and concepts will help structure a focused and effective document.

To adequately define the basic functionality requirements, business needs, and desired operations of the data center, consider the following criteria:

- The facility's uptime
- The electrical equipment to be deployed
- The electrical design strategy

The basic requirements, business needs, and desired operations are collectively known as the *backbone requirements*.

11.1 UPTIME

First, determine the required uptime of the facility. Can the system incur some downtime?

If it can, you must address how much downtime can occur without affecting business operations. Due to the criticality of their businesses, financial institutions, colocation facilities, or institutions directly related to revenue generation require the highest levels of uptime. Less mission-critical organizations have the flexibility to lower their uptime requirements significantly.

11.2 ELECTRICAL EQUIPMENT TO DEPLOY

Next, consider the electrical equipment that will be deployed in the data center and used by the servers. It is necessary to answer the following questions:

- How is the power supply configured?
  - Single or dual
  - Line-to-Line or Line-to-Neutral voltage
- How much power will each server consume?
- What is the power factor of the server power supplies?
- What is the voltage/current total harmonic distortion (THD)?
- What is the power supply inrush current?

11.3 ELECTRICAL DESIGN

After clearly defining the backbone requirements as already mentioned, the next step is to develop one or more designs that will sufficiently accommodate your business needs.

There are three main hierarchies of electrical data center design: N, $N+1$, and $2N$. The N design system uses the exact number of equipment or systems without any built-in redundancy. $N+1$ designs have one additional system built in for redundancy, while $2N$ refers to designs that have double the equipment required, which provides maximum redundancy.

Table 11.1 outlines the most common data center topologies along with their pros and cons.

These configurations are described in greater detail in the following sections. Figure 11.1 illustrates the symbols used by the diagrams.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

TABLE 11.1 Data center electrical topologies

		$N+1$	$N+1$	$N+1$	
	N	Parallel UPS redundant	Block redundant	Distributed redundant	$2N$
Redundancy	No redundancy	One UPS capacity worth of redundancy	One system capacity worth of redundancy	One system capacity worth of redundancy	Maximum redundancy, two identical systems
Pros	• Less electrical equipment required • Lowest cost: initial build and maintenance	• Easier load management because the power is shared across UPS bus	• Reserve bus is always available in case of outages and maintenance • Easy load management	• All equipment is utilized • Cost-effective solution	• System separation provides true redundancy on every level
Cons	• Outages and failures will bring down server cabinets	• UPS bus is a single point of failure	• Requires installation of load transfer capability equipment • Low utilization of redundant system leading to decreased efficiency	• Requires installation of load transfer capability equipment • Strenuous ongoing load management exercises to ensure adequate distribution	• High equipment cost • Increased maintenance cost

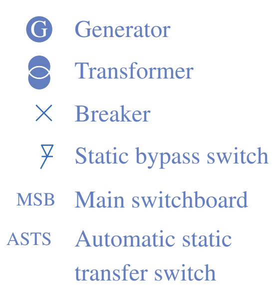


FIGURE 11.1 Diagram ledger.

11.3.1 Parallel UPS Redundant Configuration

In this topology, power flows from the utility through parallel uninterruptible power supply (UPS) system and power distribution units (PDUs). A UPS paralleling switchgear provides power to PDUs. PDUs distribute power to the servers. If the utility power source fails, generators will pick up the load, and the parallel UPS system will bridge the power outage gap during the utility-to-generator transition. A parallel UPS redundant topology accommodates single- or dual-corded rack configurations, providing redundancy at both the UPS ($N+1$) and PDU ($2N$) levels (Fig. 11.2).

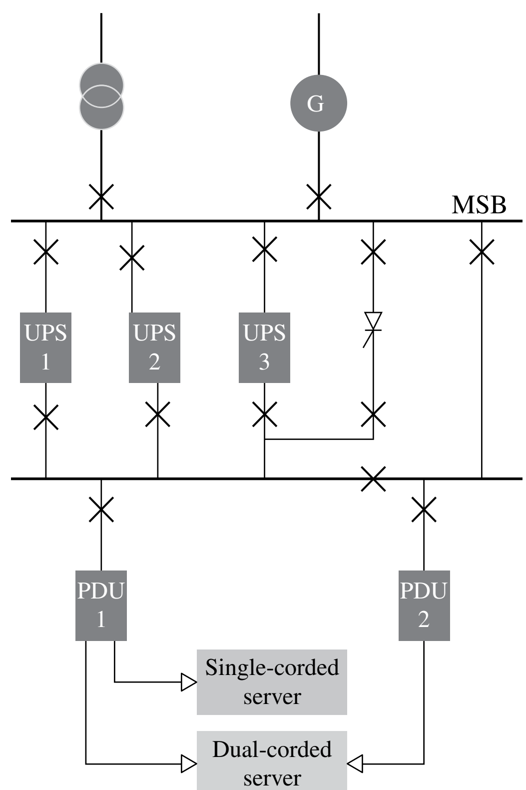


FIGURE 11.2 Parallel UPS redundant configuration.

11.3.2 Block Redundant Configuration

In this topology, also known as a *catcher system*, power flows from the utility through the UPS/PDU and connects to the server. Each set of PDUs has a UPS dedicated to it, with one reserve to provide power in case of an outage. A block redundant topology accommodates single- or dual-corded rack configurations, providing redundancy at both the UPS and PDU levels (Fig. 11.3).

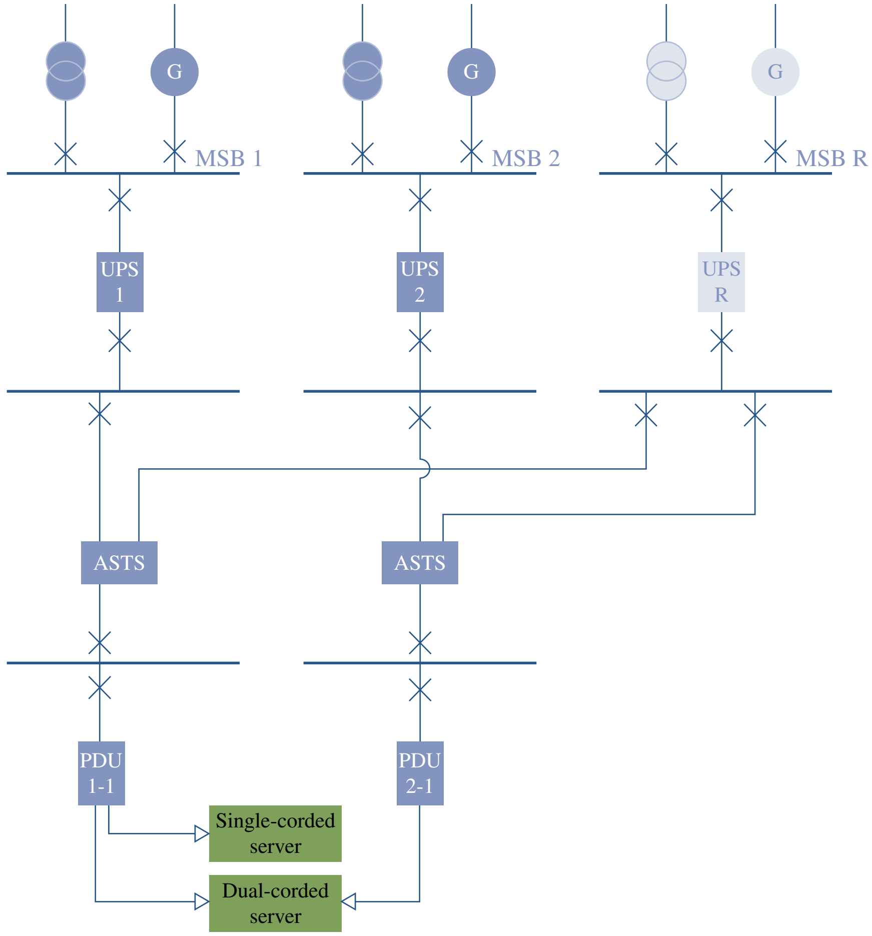


FIGURE 11.3 Block redundant configuration.

11.3.3 Distributed Redundant Configuration

In this topology, power flows from the utility through the UPS/PDU and connects to the server. The data center load is distributed across the PDUs, leaving enough capacity for the UPS.

For example, if there are three systems in the data center, each system should be loaded to 66%; if one system fails, 33% of the load can be transferred to each of the remaining live systems.

A distributed redundant topology accommodates single- or dual-corded rack configurations, providing redundancy at the system level (Fig. 11.4).

11.3.4 2*N* Configuration

In this topology, power flows from the utility through the UPS/PDU of two separate systems and connects to the server. A 2*N* configuration provides redundancy throughout the system, accommodating single- or dual-corded racks (Fig. 11.5).

11.3.5 *N* + 1 Topologies

Figure 11.6 displays the Parallel UPS, Block, and Distributed *N* + 1 redundant topologies in normal operation as well as in failure operation.

11.3.6 Facebook Inc. Electrical Design

These electrical topologies are not mutually exclusive; the key is to design a data center that satisfies business needs. Facebook designed a data center that merges these topologies, resulting in a solution satisfying their requirements. The data center comprises a mix of 208 and 277 V equipment as well as single- and dual-corded servers.

The Facebook data center design team developed a revolutionary design that does not require a centralized UPS, significantly reducing losses. In this design, power flows from the utility, connecting directly to the 277 V server. Battery backup cabinets are connected to the servers delivering DC power in case of an outage.

Overall, the Facebook data center follows the block redundant configuration with a reserve bus that provides power to one of the six independent systems if a failure occurs (Fig. 11.7).

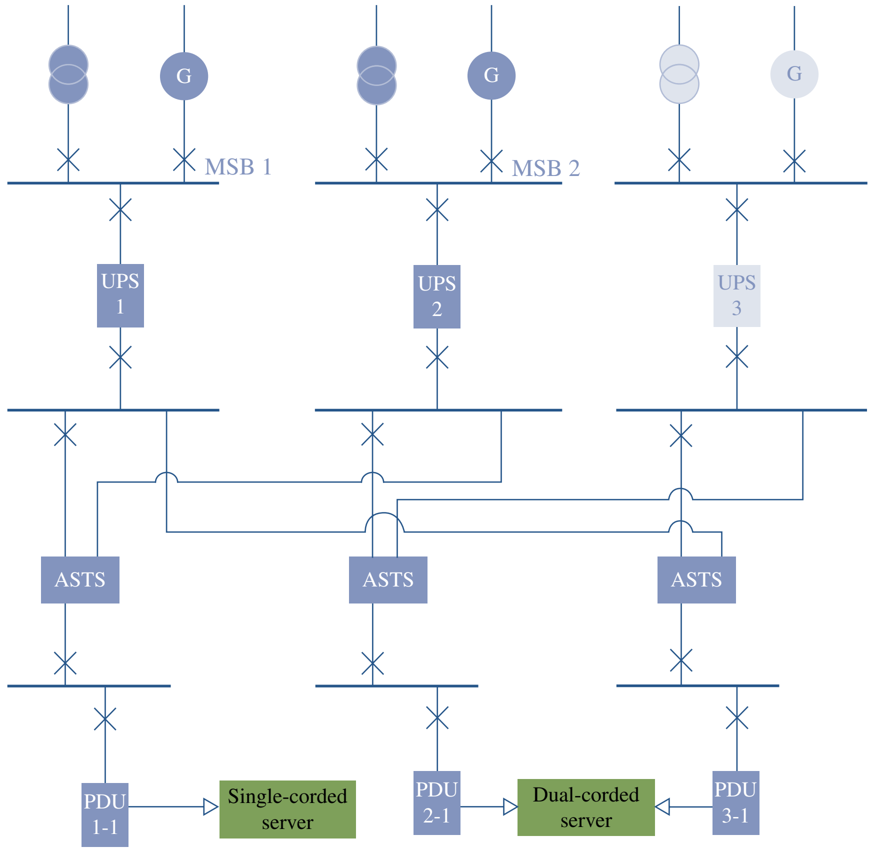


FIGURE 11.4 Distributed redundant configuration.

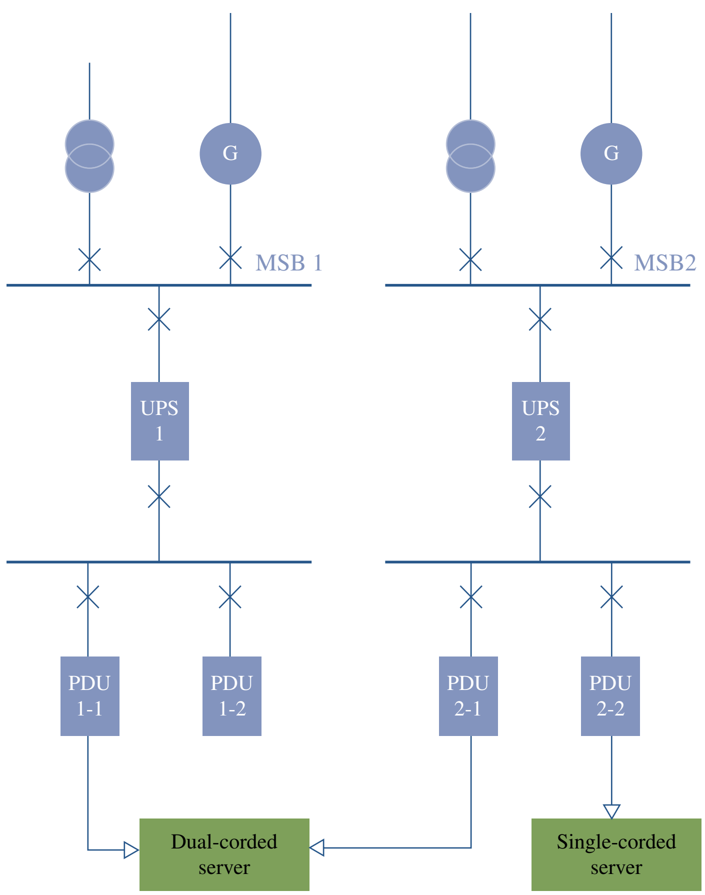


FIGURE 11.5 2*N* configuration.

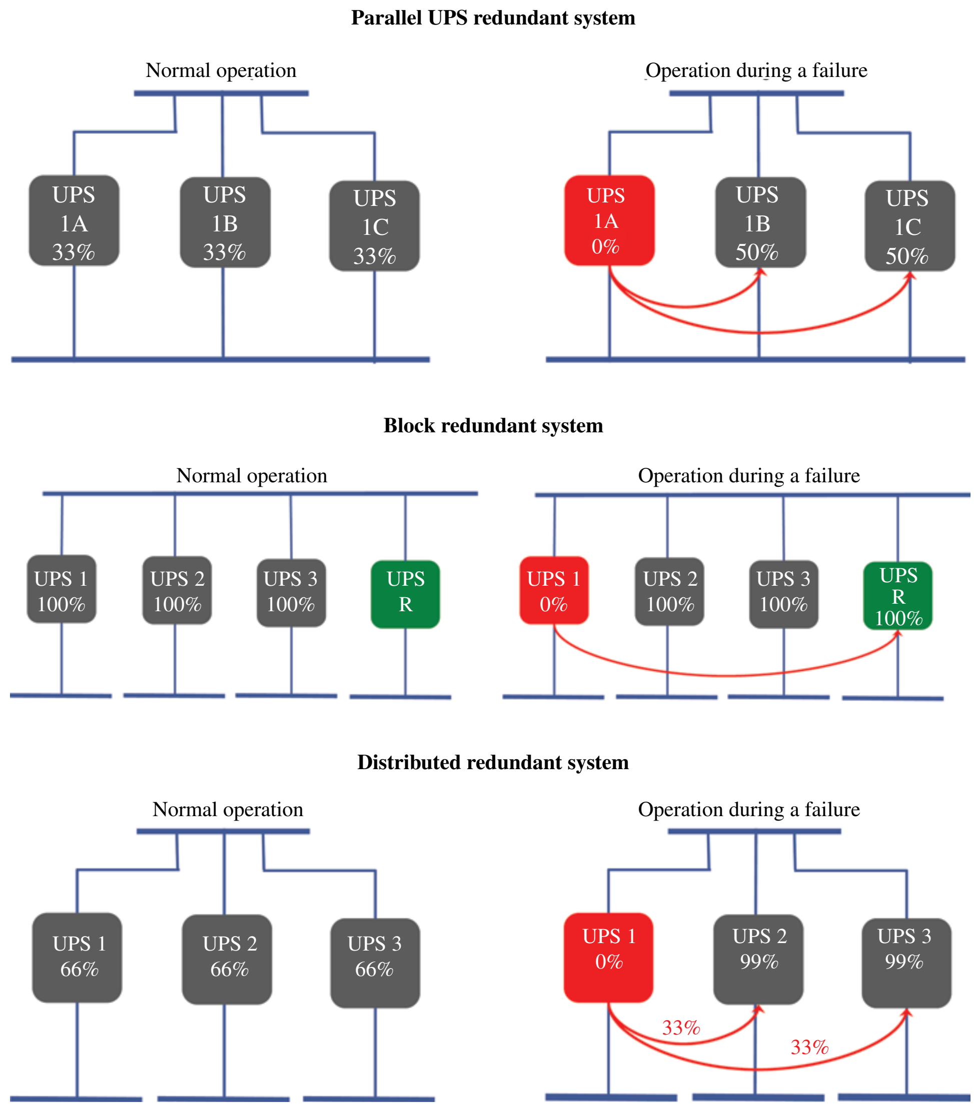


FIGURE 11.6 $N+1$ normal operation and failure scenarios.

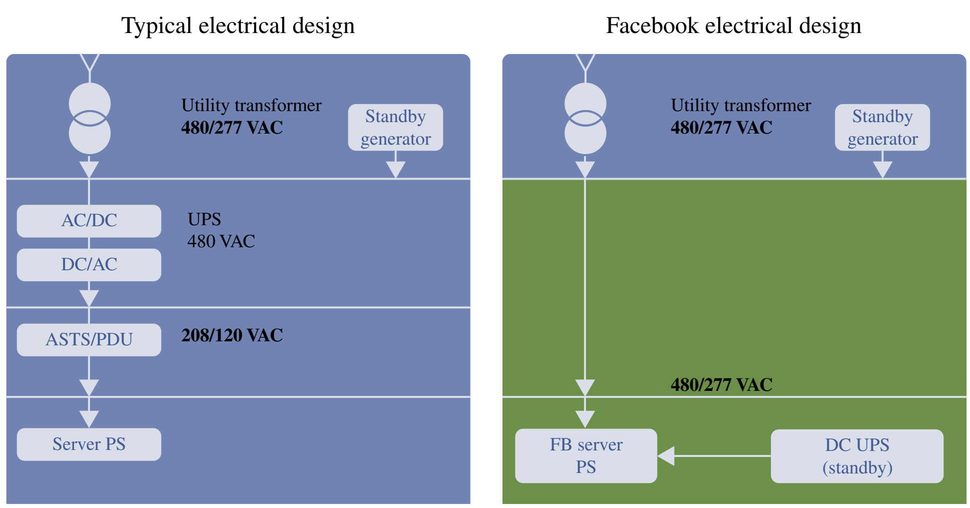


FIGURE 11.7 Typical versus Facebook electrical topologies.

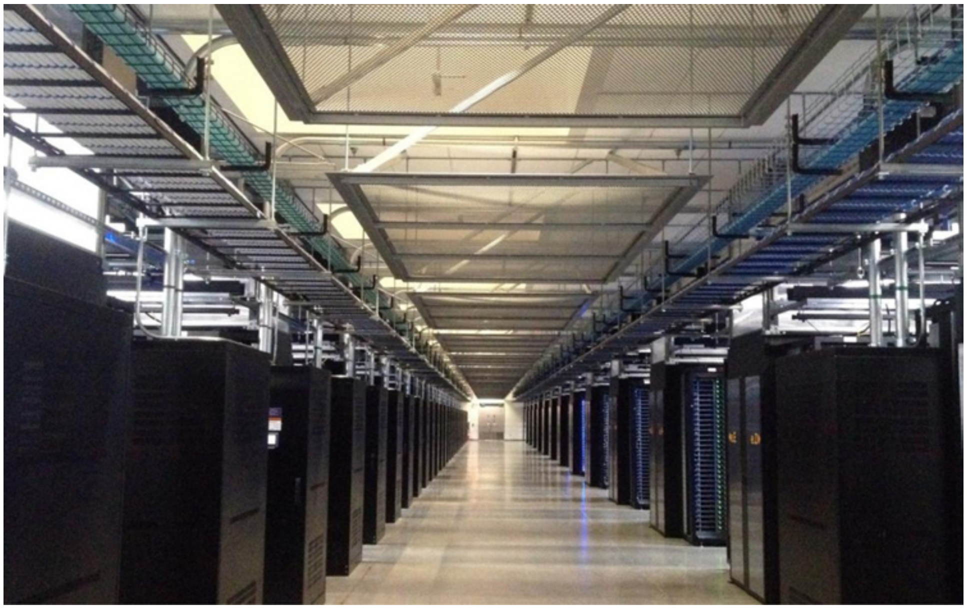


FIGURE 11.8 Facebook data center suite.

FIGURE 11.9 Facebook data center aisle.

Figures 11.8 and 11.9 illustrate a typical Facebook-designed suite. 277 V power is distributed to the Facebook Open Compute Project (http://www.opencompute.org/) servers.

Since there isn't centralized UPS, the DC UPS battery cabinet, in Figure 11.10, distributes power to the servers when failures occur.

Figure 11.11 is a diagram that goes into depth about the power configuration of a typical DC UPS battery cabinet and 277 V server.

11.4 AVAILABILITY

After successfully completing the initial designs based on specific business requirements, the next best practice is to calculate the system availability.

Availability calculations provide a means to understanding the predicted reliability of the data center design. These calculations will help the design team direct additional resources toward building adequate redundancy in the system, because the area with least redundancy is easy to identify.

In order to perform this calculation, you need to know the Mean Time to Fail (MTTF), the Mean Time between Failures (MTBF), and the Mean Time to Repair (MTTR); these values are available on the equipment manufacturer's data sheet or IEEE Gold Book.[1] Understanding the failure dependencies will help you maintain adequate operation of the data center through proactive preparation. The diagram in Figure 11.12 illustrates the failure rates, which contribute to availability calculations.

FIGURE 11.10 Facebook DC UPS battery cabinet.

Table 11.2 outlines the data that must be accumulated and the equations required to perform complete analysis of the data center topology. These calculations must be performed on individual electrical equipment. Then, the data can be built up to identify the entire data center's predicted availability.

11.4.1 Series versus Parallel Connections

After computing the failure rate, availability, and MTTF calculations for the individual pieces of electrical equipment, you need to identify the connection scheme in the various designed topologies to compare or enhance.

Equipment is connected either in series or in parallel. Series refers to a direct connection between two devices; parallel is when a bus connects two devices. Figure 11.13 depicts the differences between the two methods of connection.

The formulas in Table 11.3 show the calculations required for series versus parallel systems.

11.4.2 Example Availability Scenario

Table 11.4 shows an example of a system that may be deployed in a data center. It consists of a utility feed, transformer, generator, main switchboard (MSB), UPS, and a PDU. The first table shows the necessary data that is needed to perform availability calculations.

Note that this is a fictitious data used only for illustrating this example. When calculating for your data center, please refer to the IEEE Gold Book[2] and the equipment data sheets.

Next is a simple schematic of the power chain (Fig. 11.14). The // denotes a system in parallel, while + denotes a system in series:

Part 1 = [(Utility + Cable + Circuit Breaker + Transformer) // (Generator + Generator Controls + Cable + Circuit Breaker)] + MSB

Part 2 = Part 1 + Circuit Breaker + Cable + UPS + Cable + Circuit Breaker _ Distribution Panel

Part 3 = Part 2 + Circuit Breaker + Cable + PDU

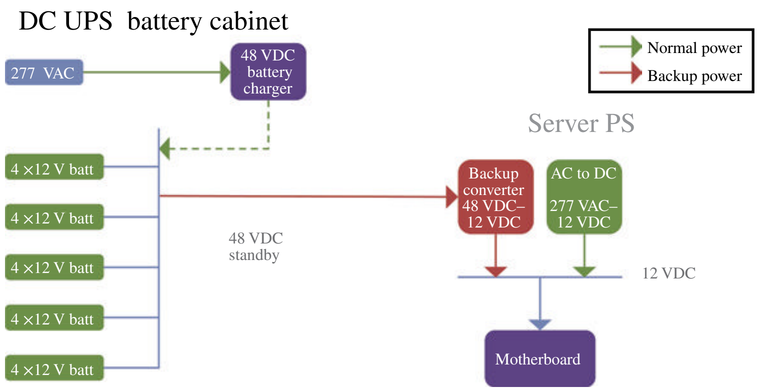
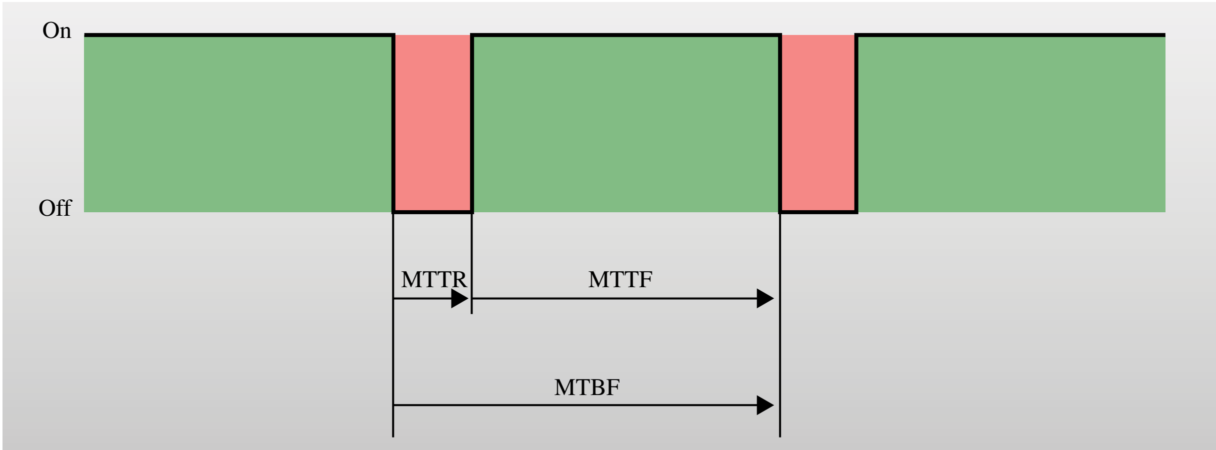


FIGURE 11.11 DC UPS backup scheme.

[1] Please refer to Appendix for a sample table.

[2] Please refer to Appendix for reference.



Mean Time To Fail = MTTF Mean Time Between Failure = MTBF Mean Time To Repair = MTTR

MTBF = MTTF + MTTR

FIGURE 11.12 Availability time diagram.

TABLE 11.2 Availability symbols, definitions, and equations

Symbol	Definition	Equations
$\boldsymbol{\lambda}$	Failure rate (failures/h)	$\boldsymbol{\lambda} = 1/\text{MTTF}$
MTTR	Mean time to repair (MTTR) per failure (h)	
MTBF	Mean time between failures (h)	**MTBF** = MTTF + MTTR
MTTF	Mean time to fail (h)	
A	System availability	**A** = MTTF/(MTTF + MTTR) = MTTF/MTBF
U	System unavailability	**U** = 1 − A
R	Reliability	**R** = $e^{-\lambda t}$
P	Probability of failure	**P** = 1 − $e^{-\lambda t}$
s	System in series	
p	System in parallel	

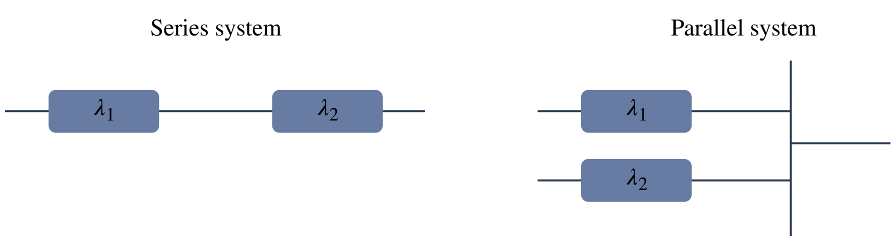


FIGURE 11.13 Series versus parallel connections.

TABLE 11.3 Series and parallel system equations

	Series equations	Parallel equations
Failure rate	$\lambda_s = \lambda_1 + \lambda_2$	$\lambda_p = [\lambda_1\,\lambda_2\,(\text{MTTR}_1 + \text{MTTR}_2)]/(1 + \lambda_1\,\text{MTTR}_1 + \lambda_2\,\text{MTTR}_2)$
Availability	$A_s = A_1 \times A_2$	$A_p = 1 - [(1 - A_1) \times (1 - A_2)]$
Mean time to repair	$\text{MTTR}_s = [(\lambda_1 \times \text{MTTR}_1) + (\lambda_2 \times \text{MTTR}_2)]/(\lambda_1 + \lambda_2)$	$\text{MTTR}_p = (\text{MTTR}_1 \times \text{MTTR}_2)/(\text{MTTR}_1 + \text{MTTR}_2)$

TABLE 11.4 Availability calculation data

Standard data input			Standard calculation	
Equipment	MTTF	MTTR	Failure rate (λ) (λ = 1/MTTF)	Availability (A) A = MTTF/ (MTTF + MTTR)
Cable	3,500,000	8.00	0.00000029	0.99999771
MSB	2,500,000	24.00	0.00000040	0.99999040
Generator	500,000	48.00	0.00000200	0.99990401
Generator controls	1,500,000	8.00	0.00000067	0.99999467
PDU	2,500,000	8.00	0.00000040	0.99999680
Transformer	2,000,000	250.00	0.00000050	0.99987502
UPS	1,000,000	0.00	0.00000100	1.00000000
Utility	7,500	6.00	0.00013333	0.99920064
Circuit breaker	2,500,000	8.00	0.00000040	0.99999680
Distribution panel	2,200,000	4.00	0.00000045	0.99999818

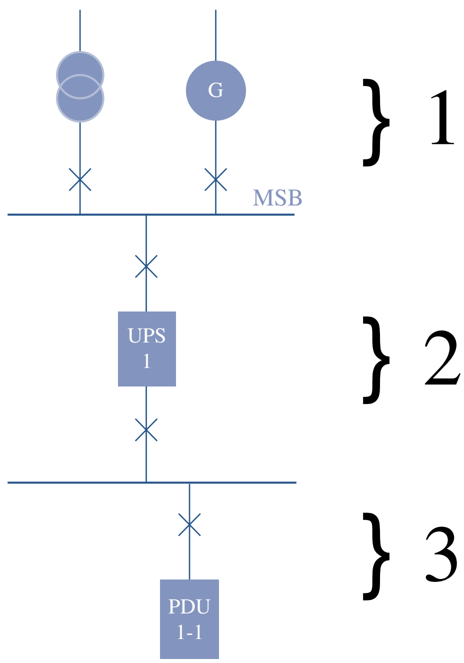


FIGURE 11.14 Power chain schematic.

Table 11.5 shows availability calculations; here, the // denotes a system in parallel, while + denotes a system in series.

11.4.3 Loading and Operations

Optimal data center operation requires efficient loading. Take these key factors into account when deploying equipment:

- Breaker size, rating, and trip settings
- Server information
- System power distribution
- Disaster recovery plans

Electrical equipment power information should be available through the manufacturer's data sheets. On the one-line diagram, you can usually find a combination of the Amperage (A), Voltage (V), Apparent Power (VA), and Real Power (W). In order to efficiently load the data center, it is necessary to convert all the equipment power to Real Power. Converting to Real Power avoids overloading and stranding power. You must also take Reactive Power into account because it affects the generator load (Table 11.6).

When loading the data center, it is important to understand the interdependencies of the deployed equipment. You must also consider the effects on all upstream and downstream machinery. It's very much a balancing game. You have to prioritize and find the best match for your ideal situation. More than likely, you will have to make some sacrifices in the final deployment.

11.4.4 Rack Information

First, you must define the amount of servers you plan on deploying and outline their characteristics. Make sure to identify these key metrics:

- Maximum power expected from the server
- Average power expected from the server
- Redundancy level
- Voltage compatibilities
- Phase balancing

Start with the most critical cabinets. Define where they will be deployed and their failover scenarios. For critical gear, it's vital that the system can withstand a failure without affecting the load. In order to make informed decisions about loading, you must have a deep understanding of the data center's power distribution. Identify the most reasonable row/cage/section to deploy dual-corded racks (based on redundancy level in the overall system) and where to place single-corded equipment. At Facebook, 208 and 277 V racks are used. Thus, in the planning phase, it is necessary to

TABLE 11.5 Availability calculations

Standard data input			Standard calculation		
Equipment	MTTF	MTTR	Failure rate (l)	MTBF	System availability
Utility					
Utility 1	7,500	7.00	0.00013333	7,507.00	0.99906754
Cable	3,500,000	8.00	0.00000029	3,500,008.00	0.99999771
Circuit breaker	2,500,000	8.00	0.00000040	2,500,008.00	0.99999680
Transformer	2,000,000	250.00	0.00000050	2,000,250.00	0.99987502
Series system (utility, cable, CB, TX)		**7.908315339**	**0.000134519**		**0.998937189**
Generator					
Generator	500,000	48.00	0.00000200	500,048.00	0.99990401
Generator controls	1,500,000	8.00	0.00000067	1,500,008.00	0.99999467
Cable	3,500,000	8.00	0.00000029	3,500,008.00	0.99999771
Circuit breaker	2,500,000	8.00	0.00000040	2,500,008.00	0.99999680
Series system (gen, gen controls, cable, CB)		**31.86363636**	**3.35238E−06**		**0.999893191**
Part 1 [(Utility, Cable, CB, TX)// (Gen, Gen Controls, Cable, CB)] + MSB					
Gen//utility		6.335813894	1.79355E−08		0.999999886
MSB	2,500,000	24.00	0.00000040		0.99999040
(Gen//utility) + MSB		**5.01**	**2.17635E−13**		**0.999999999998910**
Part 2 (Part 1 + CB + Cable + UPS + Cable + CB + DP)					
Part 1		5.01	2.17635E−13		0.9999999999989
Circuit breaker	2,500,000	8.00	0.00000040		0.99999680
Cable	3,500,000	8.00	0.00000029	3,500,008.00	0.99999771
UPS	1,000,000	0.00	0.00000100	1,000,000.00	1.00000000
Cable	3,500,000	8.00	0.00000029	3,500,008.00	0.99999771
Circuit breaker	2,500,000	8.00	0.00000040		0.99999680
Distribution panel	2,200,000	4.00	0.00000045		0.99999818
Series system (Part 1 + CB + Cable + UPS + Cable + CB + DP)		**4.525735332**	**0.00000283**		**0.99998721**
Part 3 (Part 2 + CB + Cable + PDU)					
Part 2		4.525735332	0.00000283		0.99998721
Circuit breaker	2,500,000	8.00	0.00000040		0.99999680
Cable	3,500,000	8.00	0.00000029	3,500,008.00	0.99999771
PDU	2,500,000	8.00	0.00000040		0.99999680
Series system (Part 1 + Part 2 + CB + Cable + PDU)		**5.49**	**0.00000391**		**0.99997852**

TABLE 11.6 Electrical Engineering Definitions and Units

Name	Definition	Unit	Formulas
Voltage	Measure of electrical potential	V (Volts)	
Current	Flow of electrical charge through a medium	A (Amperes)	
Apparent power	Total magnitude of power transmitted across an electrical power system	VA (Volt-Amps)	V × I
Power factor	The measure of how much real power is present in a AC power system	kW/kVA	Real Power/ Apparent Power
Reactive power	Energy that is stored in inductive and capacitive elements; it does no useful work in the electrical load and must be taken into account because it affects generator performance	VAR	
Real power	Power that does actual work	W (Watt)	V × I × Pf

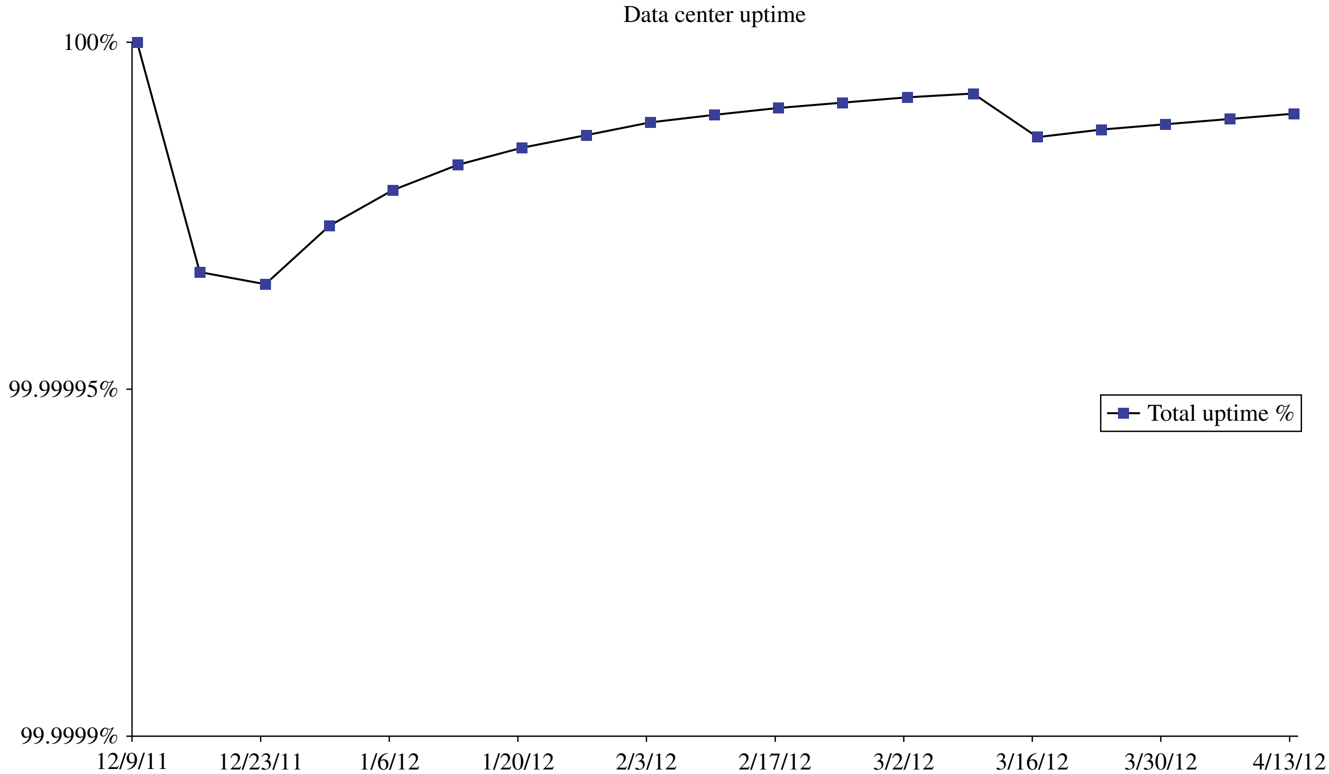


FIGURE 11.15 Field data validated uptime graph.

account for how the infrastructure will supply the necessary power to those racks. Then you must identify rows/cages/sections that are fed from the same source (PDU, SB, etc.). Ensure that you don't exceed the maximum kW load of the panels and upstream equipment. You must also not exceed the power of the backup generators. When placing the cabs, distribute the load evenly between the three phases. More than likely, you will iterate your layout several times until you design one that meets all your goals.

After the racks are deployed, it is important to monitor the health of the data center on an ongoing basis to proactively identify solutions for potential problems. Deploying Power Quality Meters (PQM) on the equipment and trending the data are key to preventing catastrophic events. More metering (BCM, smart iPDUs, etc.) will provide additional insight into the performance of the infrastructure. If all your metering tools use the same protocols to communicate with one another, it will be easier to pull them into a single interface. In addition, benchmarking on several metrics, such as peak power draw and kWh will be easier to obtain. Power Usage Effectiveness (PUE) is a unit-less metric in the industry that shows how much of the overall power is going to the racks versus support equipment (mechanical systems):

$$\text{PUE} = \frac{\text{Overall Power}}{\text{IT Power}}$$

Having meters at the utility level will provide you with the overall power usage of the data center. Metering the IT equipment separately simplifies this calculation into a simple division.

11.4.5 Data Center Uptime

Uptime is a metric that validates the availability calculations by trending live power consumption data from the data center. Figure 11.15 depicts a sample graph comparing uptime week by week. You can create a similar graph by maintaining records of cabinet uptime minutes. The sum of all the cabinet minutes validates the data center's total uptime.

11.5 DETERMINING SUCCESS

The determining success of a data center's design is ultimately driven by business requirements. To create an efficient design, you must define the needs, redundancy requirements, and desired uptime of the data center. Formulate designs that meet these needs and calculate the availability of every design to determine one that meets the needs and/or use the calculations to build more redundancy around weak areas.

APPENDIX 11.A Sample reliability data from IEEE Gold Book

IEEE Gold Book reliability data

Equipment category	λ failures per year	*r*, hours of down time per failure	λ*r* forced hours of down time per year	Data source in IEEE survey [B8] table
Protective relays	0.0002	5.0	0.0010	19
Metal-clad drawout circuit breakers				
0–600 V	0.0027	4.0	0.0108	5, 50
Above 600 V	0.0036	83.1[a]	0.2992	5, 51
Above 600 V	0.0036	2.1[b]	0.0076	5, 51
Power cables (1000 circuit ft)				
0–600 V, above ground	0.00141	10.5	0.0148	13
601–15,000 V, conduit below ground	0.00613	26.5[a]	0.1624	13, 56
601–15,000 V, conduit below ground	0.00613	19.0[b]	0.1165	13, 56
Cable terminations				
0–600 V, above ground	0.0001	3.8	0.0004	17
601–15,000 V, conduit below ground	0.0003	25.0	0.0075	17
Disconnect switches enclosed	0.0061	3.6	0.0220	9
Transformers				
601–15,000 V	0.0030	342.0[a]	1.0260	4, 48
601–15,000V	0.0030	130.0[b]	0.3900	4, 48
Switchgear bus—bare				
0–600 V (connected to 7 breakers)	0.0024	24.0	0.0576	10
0–600 V (connected to 5 breaker)	0.0017	24.0	0.0408	10
Switchgear bus insulated				
601–15,000 V (connected to 1 breaker)	0.0034	26.8	0.0911	10
601–15,000 V (connected to 2 breakers)	0.0068	26.8	0.1822	10
601–15,000 V (connected to 3 breakers)	0.0102	26.8	0.2733	10
Gas turbine generator	4.5000	7.2	32.4000	Appendix L, Table III

[a] Repair failed unit.
[b] Replace with spare.

FURTHER READING

Bitterlin IF. International Standards for Data Center Electrical Design. Chloride.

Data Center Energy Management Website. Lawrence Berkeley National Laboratory. Available at http://hightech.lbl.gov/DCTraining/. Accessed on June 12, 2014.

Open Compute Project. Available at http://www.opencompute.org/. Accessed on June 12, 2014.

Sawyer R. Calculating Total Power Requirements for Data Centers. APC; 2005. White Paper #3.

12

FIRE PROTECTION AND LIFE SAFETY DESIGN IN DATA CENTERS

Sean S. Donohue
Hughes Associates, Inc., Colorado Springs, CO, USA

12.1 FIRE PROTECTION FUNDAMENTALS

Fire is a risk every business must deal with. For data and telecommunications centers, that risk includes not only the safety of people in the building but continuity of operations and the value of the equipment and data. Today, these centers are the nervous system of businesses and organizations throughout the world; and the more critical the site, the less acceptable the risk of interruption or downtime. Fire protection comes in many forms, but the goals are simple:

1. Construct buildings and systems that guide people away from and protect them from harm.
2. Give the users and responders accurate information in order to make informed decisions.
3. Limit loss (life, downtime, equipment, data, or other).

This chapter will discuss life safety and active and passive fire protection and will present the choices available to the designer typically used in data centers.

12.1.1 Fire and Data Centers

Electronic equipment and data centers contain a variety of combustible fuel, from printed circuit boards to wiring insulation and cabinet enclosures, which increasingly contain more and more plastic. Furnishings, backboards, batteries, and floor tiles also contribute to fuel load.

In recent years, the trend has been to increase the rack power consumption density. With increased power density comes more heat and a higher risk of overheating if the ventilation systems cannot keep up. From a risk standpoint, it is critical to maintain good housekeeping within data centers and remove furnishings, paper, or other combustible load that does not contribute to the core function of the data center. Batteries and nonessential equipment should be housed in a separate room if possible.

When electronic equipment combusts, it produces many different gases generically referred to as smoke or products of combustion. These can include corrosive gases such as HCN and HCl that can do more damage to printed circuit boards than heat from a fire. Because of this, early detection is often desired so that staff can respond to an incipient condition before it becomes an emergency. Detection systems can continue to alert occupants to developing stages of a fire and can be programmed to provide suppression system activation.

When a fire grows beyond the ability of occupants to control, an automatic fire suppression system can extinguish or control the fire until the Fire Department arrives and completes extinguishment. Many buildings are required by building codes to be equipped with automatic fire sprinkler systems, based on the size and use of the building. Gaseous fire suppression systems are also used as alternatives to sprinklers, when permitted by the local authority having jurisdiction (AHJ).

The prime differentiator between the two systems is that sprinkler protection is considered a life safety system because it (very often) contains a fire to its room of origin, limits fire spread, and protects the remainder of the building, whereas a gaseous system is considered as equipment protection because it mitigates a specific loss other than life.

Table 12.1 illustrates how a fire in a data center may develop.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

TABLE 12.1 Stages of fire growth for an electrical fire in a data center

Fire growth stage	Description	Possible response
Incipient	Overheating of equipment/circuits; trace amounts of combustion gases equal to lowest amount detectable by an aspirating system. No other detection	Occupant alert Occupant action Pre-alarm
Smoldering (visible smoke)	Increased burning, detectable by human smell. Activation of spot-type smoke detection. Highest alert level for aspirating systems	Occupant action Fire alarm Initiate clean agent system countdown or release solenoid valve in a pre-action system
Flaming	Pyrolysis and flaming combustion. Activation of multiple spot-type detectors. Increased room temperature and development of an upper gas layer	Fire alarm Initiate clean agent system countdown or release solenoid valve in a pre-action system
Fire growth/spread	Copious production of smoke in quantities sufficient to quickly activate multiple spot-type detectors. Rapid acceleration of heat release and fusing of nearest sprinkler	Fire alarm Sprinkler system discharge

12.2 AHJS, CODES, AND STANDARDS

The term AHJ is often misconstrued to mean a government entity enforcing statutory or regulatory fire/life safety requirements within the site's geographic or jurisdictional area. While this group is certainly included, an AHJ can be any public or private entity to which ownership is subject to and can include the following:

- Local, state, or federal authorities
- Insurance companies
- Ownership (self-regulation)
- Industry groups

These groups either adopt national standards that address construction requirements or create their own. They also replicate much of the information for required compliance so the provisions will be similar, but not always the same. For example, the Telecommunication Industry Association (TIA) Level III requirements mirror FM Global requirements for 1 h rated rooms, whereas the building code does not. Sometimes, requirements can conflict so it is important to understand the priority. Statutory code requirements are legally required; insurance guidelines can have a financial impact, where ownership guidelines are a matter of internal policy.

12.3 LOCAL AUTHORITIES, NATIONAL CODES, AND STANDARDS

Data centers are highly specialized spaces with extensive technical demands, yet they represent a small percentage of what a typical jurisdiction reviews or inspects. As with any specialized system, it is important to communicate with the authorities early in the design process because requirements may not be the same from one jurisdiction to the other. This is true for site location, construction, power, ventilation, and fire protection among other requirements. Information that is typically available online or can be attained by contacting the planning, building, or fire department includes the following:

- Geographic area of jurisdiction
- Code edition
- Amendments and local policies
- Special interpretations

The local code reviewer will typically appreciate the designer contacting them early for a special project. For jurisdictions that are not as easily approached, a local designer may need to be brought on to assist the team.

In the United States, the International Building Code [1] (IBC) and International Fire Code [2] (IFC) apply in most jurisdictions as a base construction and maintenance code. Smaller rural jurisdictions will tend to adopt a code "straight up" or with little modifications, whereas large jurisdictions and cities will more heavily amend the code. An early code review is critical to ensure the design team understands all local constraints. An installation that was used in one location cannot always be repeated in another.

The National Fire Protection Association (NFPA) publishes hundreds of standards addressing topics ranging from storage of flammable and combustible liquids to protective gear for firefighters. NFPA standards that apply to data centers and are referenced by the IBC/IFC include:

- NFPA 10, Standard for Portable Fire Extinguishers
- NFPA 12, Standard on Carbon Dioxide Extinguishing Systems

- NFPA 12A, Standard on Halon 1301 Fire Extinguishing Systems
- NFPA 13, Standard for the Installation of Sprinkler Systems
- NFPA 20, Standard for the Installation of Stationary Fire Pumps for Fire Protection
- NFPA 70, National Electrical Code® (NEC)
- NFPA 72, National Fire Alarm and Signaling Code
- NFPA 2001, Clean Agent Fire Extinguishing Systems

Additional standards that are not referenced in the IBC or IFC but are applicable to the data center and telecommunications industry include:

- NFPA 75, Standard for the Protection of Information Technology Equipment
- NFPA 76, Standard for the Fire Protection of Telecommunications Facilities
- NFPA 101, Life Safety Code®
- NFPA 750, Standard on Water Mist Fire Protection Systems

NPFA 75 [3], for example, covers active and passive protection and risk analysis. As of this publication, NFPA 75 is not referenced by the IBC or NFPA 101; therefore, it is not enforceable unless specifically adopted. It is referenced by the NEC in Article 645, but not as a required standard; therefore, designers must choose to use this standard unless required by some other AHJ. Among other provisions, NFPA 75 requires fire separation of IT rooms, sprinkler protection if the room is located within a sprinkler protected building, and automatic detection.

12.3.1 Insurance Companies

The goals of insurance companies are clear: mitigate loss and reduce risks. In order to keep premiums low, insurance companies will often place requirements on their customers. Some companies such as FM Global have created their own list of standards known as FM Data Sheets. Examples include FM Data Sheet 5–32, Electronic Data Processing Systems, or FM Data Sheet 4–9, Clean Agent Fire Extinguishing Systems. These data sheets prescribe compliance that may exceed that found in building/life safety codes.

The user should be aware that ownership may be held to these standards in the future and should incorporate any discrepancies into the design.

12.3.2 Ownership Standards

Ownership (e.g., federal and state governments and large companies) may have specific requirements that exceed code or insurance requirements for their own protection including and many times based on their own experience with previous installations or loss. Some examples include the following:

- No wet piping above the data center
- Security measures that must still allow code-complying egress

12.3.3 Tiering System

Lastly, industry groups such as the TIA [4] and the Uptime Institute [5] have published standards based on a level or tiering system that affect, among other requirements, passive fire protection, fire detection, and suppression. Tiering describes various levels of availability and security for the data center infrastructure; the higher the tier, the stricter the requirement. Tier I and II facilities are typically only required to meet minimum code requirements, whereas Tier III and IV facilities often exceed minimum code requirements. For example, Tier III and IV facilities may require both sprinkler and clean agent fire suppression, whereas Tier I and II facilities do not specify clean agent systems. Examples of the topics covered by the TIA are shown in Table 12.2.

12.4 LIFE SAFETY

The first goal of building and fire codes is to safeguard the lives of people within a building. When it comes to data center layouts, rooms are more equipment intensive than occupant intensive. Some data centers are designed to be maintained remotely, and only rarely is on-site presence required. In either case, the building and life safety codes address the typical life safety concerns appropriate for the intended use of the building. In the following are some highlights that will assist the designer in addressing occupant-specific code requirements.

12.4.1 Occupancy

Occupancy classification describes the use or function of a space and sets in motion different requirements for different hazards. A tire storage warehouse will have very different construction and protection requirements than a hospital, for example. Data centers have historically been defined as business occupancies because of their accessory function to a business that employs people and provides a service. Data centers can also be considered as storage occupancies especially when they are constructed as stand-alone buildings, because of their function of providing data *storage.*

TABLE 12.2 Summary of fire protection and life safety topics for different levels*

Level reference guide topics (including but not limited to)	
Architectural	• Type of construction • Exterior wall ratings • Structural/interior bearing walls • Roofs and floor/ceiling assemblies • Shafts • Computer room partition walls • Noncomputer room partition walls • Meet NFPA 75 • Fire separation from computer room and support areas • Corridor width • Door and window fire rating • Multiple tenancy within same building
Mechanical/electrical	• Automatic fire suppressant release after computer and telecommunications system shutdown for Emergency Power Off (EPO) • Fire alarm system activation with manual EPO shutdown • Battery monitoring system • Fire detection system • Fire sprinkler system • Gaseous suppression system • Early warning smoke detection system • Water leak detection system

Courtesy of the TIA.

*Content from the ANSI/TIA 942-A-2012, *Telecommunications Infrastructure Standard for Data Centers*, standard is reproduced under written permission from the Telecommunications Industry Association. Note: All standards are subject to revision, and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent editions of the standards published by them.

12.4.2 Occupant Load

Occupant load is the number of people the code considers to be in a space at the same time. This is a conservative number meant to represent a "worst case" and is used to determine egress width, number of exits, plumbing fixture count, and ventilation (although for data centers, ventilation load is driven by equipment).

Occupant load is derived as a function of gross floor area based on function of the space as follows:

$$\frac{\text{Floor Area}\left(\text{m}^2 \text{ or } \text{ft}^2\right)}{\text{Occupant load factor}\left(\text{m}^2 \text{ or ft}^2/\text{occupant}\right)} = \text{Number of occupants in space}$$

The actual use should be openly discussed to most closely match the highest number of people anticipated during the normal use of the space. The trend for data centers is to employ fewer and fewer personnel. Spaces such as "lights-out" data centers are designed to eliminate personnel entirely, except in emergency circumstances. The applicable building or life safety code should be consulted; but the typical occupant load factor for a data center will range from 9.3 gross m^2 (100 gross ft^2) per occupant to 46.5 gross m^2 (500 gross ft^2) per occupant depending on the occupant density. The designation of "gross" indicates that the entire floor area must be used to calculate the occupant load including the following:

- Space used by equipment
- Interior walls and columns
- Supporting spaces such as corridors and restrooms

12.4.3 Egress

Building and life safety codes should be consulted for the full set of requirements regarding egress. Examples include the IBC and the Life Safety Code [6] (NFPA 101). A few of the more common egress design concerns are presented in the following.

12.4.3.1 Number of Exits All occupied rooms require at least one means of egress. A business occupancy will require a second means of egress when the occupant load reaches more than 50 occupants. Using a conservative occupant load factor of 9.3 m^2 (100 ft^2)/occupant, this means the designer should be concerned about a second exit when the data center exceeds 465 m^2 (5000 ft^2). If the data center is provided with sprinkler protection, the exits need to be placed at least one-third of the diagonal distance of the room apart from each other. If the room is not sprinkler protected, this separation increases to half the diagonal distance. When the occupant load exceeds 500 or 1000, a minimum of three and then four exits are required, respectively.

12.4.3.2 Egress Width A 915 mm (36 in.) wide door provides at least 813 mm (32 in.) of clear width. Using a width capacity factor of 5 mm (0.2 in.)/occupant required by code, this equates to about 160 occupants per door. For code compliance purposes, assuming 9.3 m^2 (100 ft^2)/occupant with two exits, the occupant load would need to exceed 320 occupants, or a floor area of 2,973 m^2 (32,000 ft^2) before a width of more than 2 typical 915 mm (36 in.) doors would need to be considered.

12.4.3.3 Travel Distance Travel distance is a function of the occupancy type discussed earlier and whether or not a building is sprinkler protected. Travel distance is the maximum distance a person should travel before reaching an exit; it is measured to the closest exit from the most remote location in a room and should be measured orthogonally to account for equipment and furnishings. The applicable building or life safety code should be consulted for these requirements, but these typically range from 61 to 91 m (200 to 300 ft).

12.4.4 Aisles

Equipment placement is a function of operational needs; however, occupants need to fit in between pieces of equipment for maintenance and for egress. Based on disability requirements, aisles should be maintained at 813 mm (32 in.) clear minimum. In large data centers, primary aisles will need to be larger to accommodate the additional occupant load and number of exits required, but not smaller than 1118 mm (44 in.).

12.5 PASSIVE FIRE PROTECTION

Walls, floors, and ceilings of rooms and buildings are required to be fire-resistance rated for a variety of reasons, in accordance with building codes, including separation of hazards, protection of the means of egress, or to allow larger buildings. Often, the building code does not require any rating at all, especially in the case of data centers, but the sensitivity of the equipment, process, or data may drive the insurer or owner to require fire-resistance rating as previously discussed. Additional hazards, such as UPS batteries, may require fire-resistance rating per the fire code.

The goal of passive fire protection is to delay the spread of fire from an adjacent space to allow time for egress and to give firefighters time to contain a fire. The higher the hourly rating of the assembly, the higher the thermal resistance. The hourly rating assigned to fire-resistance-rated assemblies should not be construed to imply a guarantee against adjacent fire events for the duration of the stated rating, but represents the minimum time an assembly is capable of resisting a predetermined fire curve. Actual fires may burn cooler or hotter than the ASTM E-119 standard curve [7] because heat output is heavily dependent on the type of fuel burning. A fire in a data center, started by the overheating of electrical insulation, could actually smolder for quite some time before developing into a flaming fire, meaning that it would likely not be as severe a fire exposure as the ASTM E-119 standard fire curve.

Typically, 1 or 2 h assemblies are encountered through the model building codes. Some standards, such as TIA, require bearing walls to have a fire-resistance rating as high as 4 h for Level IV centers. An example of a 1 h assembly from the UL Online Certifications Directory is provided in the following (Fig. 12.1). This type of wall is one of the more common 1 h assemblies and is composed of light-gage metal studs, insulation, and 5/8 in. Type "X" gypsum board. Refer to the full UL listing for complete information concerning all the materials permitted with this assembly.

The designer may consult several sources for examples of assemblies that provide the requisite fire-resistance rating. Popular sources include the IBC the UL Online Certifications Directory, and the US Gypsum Manual [8].

Openings in fire-resistance-rated walls such as doors and windows require intrinsic rating, closers, or shutters to maintain the intended fire rating of the room; this is addressed by

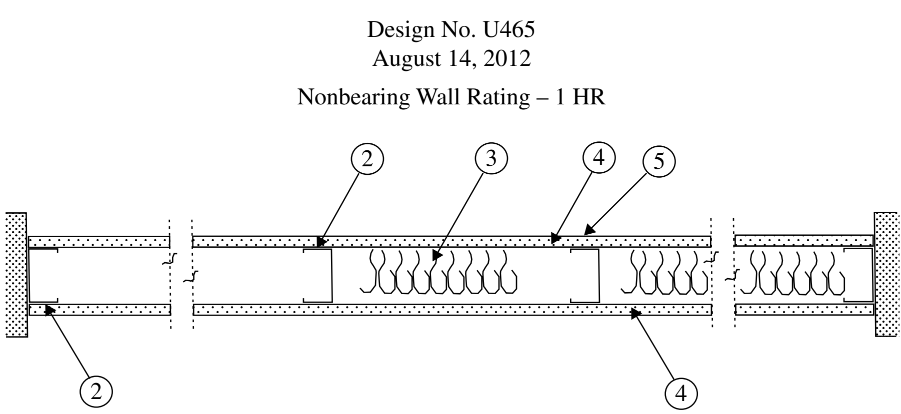


FIGURE 12.1 UL Design U465. Courtesy: Underwriters Laboratories. Reprinted from the Online Certifications Directory with permission from UL. © 2012 UL LLC.

the building code. Penetrations such as ducts, pipes, and conduit through fire-resistance-rated construction must be protected when the assembly is required to have fire-resistance rating, and again, codes and standards address how to do this. Fire and smoke dampers serve to protect the duct penetration into the protected room in case of fire. While fire dampers are activated by a fusible link, smoke dampers are activated via duct mounted or area smoke detection. It is important to evaluate user goals and ensure that HVAC flow to a room not be shut off unless there truly is a fire.

12.6 ACTIVE FIRE PROTECTION/SUPPRESSION

Automatic fire suppression is often required in buildings housing data centers; therefore, designers need to be aware of the choices, risks, and costs involved for each type of suppressing agent. Halon 1301 used to be synonymous with data center fire protection, but the use of that agent is now limited to maintenance of existing locations. A number of chemical extinguishing and inerting agents offer alternatives to Halon 1301, although automatic sprinklers still remain a viable option for low-risk installations.

12.6.1 Automatic Sprinkler Systems

Water has long been a fire suppressant of choice. It is readily available, relatively inexpensive, and nontoxic and has excellent heat absorption characteristics. That being said, water is electrically conductive and will damage energized equipment. However, automatic sprinkler systems are the fire suppression system of choice for the majority of built environments including occupancies that may be located in the same building as a data center.

Sprinkler activation is often misunderstood due to frequent misrepresentation by the entertainment industry. Contrary to popular belief, sprinklers are only activated by thermal response (not smoke) and only activate one at a time. Although there are many types of thermal elements, a popular one is the frangible glass bulb. A bulb filled with a proprietary alcohol-based fluid keeps pressurized water, air, or nitrogen from being released. When the fluid in the bulb reaches a predetermined temperature, it expands to fill the volume and breaks the glass bulb enclosure. Water or air then escapes the piping network via the new opening created (Fig. 12.2).

Due to the excellent track record sprinklers have achieved in controlling the spread of fire, current building codes offer many incentives when designers specify sprinkler protection, including the following:

- Larger buildings
- More lenient egress requirements
- Less restrictive passive fire protection

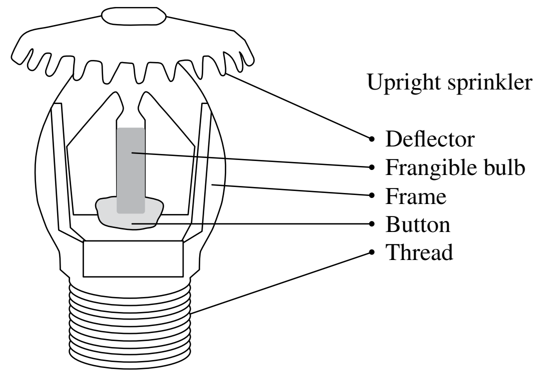


FIGURE 12.2 Standard upright sprinkler.

It is imperative that the design team discusses the use of sprinkler protection for the building. When the incentives are taken as indicated earlier, sprinklers are required throughout a building, regardless of whether an alternative system is installed, unless specific omission is permitted by all AHJs.

NFPA 13 [9] covers the installation requirements for sprinkler systems. It should be noted that this standard along with many of the installation standards promulgated by the NFPA tells the user “how” to install a system and its components. Building and life safety codes tell the designer “when” these systems are required.

When sprinkler protection is required, it is required in all occupied spaces but may also be required in accessible interstitial spaces depending on the fuel load and combustibility of those spaces. Thought should also be given to how water will drain after actuation. The activation of a sprinkler system can produce several hundred gallons of water before the fire is deemed controlled. Provisions to limit the spread of sprinkler water in a building should be incorporated into the building construction wherever possible.

12.6.2 Wet Pipe Sprinkler Systems

As the name implies, a wet pipe sprinkler system is filled with water, which is connected to a water supply system so that water discharges immediately from sprinklers opened by heat from a fire. Wet sprinkler systems are the simplest and most common of automatic fire protection systems. They make up over 80% of sprinkler systems installed [10].

Buildings in which wet systems are installed must be maintained at or above 40°F (4.4°C) and should be coordinated to avoid proximity to cooling systems operating below this temperature.

Wet systems are not typically used in data centers, but can be used where the risk of loss due to accidental release of water is low and/or cost of system installation is an issue.

12.6.3 Dry Pipe Sprinkler System

A dry pipe sprinkler system is a sprinkler system employing automatic sprinklers that are attached to a piping network containing air or nitrogen under pressure, the release of which (as from the opening of a sprinkler) permits the water pressure to open a valve known as a dry pipe valve and then allows water to flow into the piping and out the opened sprinklers. Dry systems are typically reserved for unheated buildings or portions of buildings. The designer will encounter dry pipe systems in exterior environments such as docks or canopies.

12.6.4 Preaction Sprinkler System

A preaction sprinkler system employs automatic sprinklers attached to a piping network that contains air or nitrogen that may or may not be under pressure, with a supplemental detection system installed in the same areas as the sprinklers. Preaction systems are frequently specified for data centers because they reduce the risk of accidental, nonfire release of water over the electronic equipment.

A preaction system requires the prior activation of a detection system to open a control valve and allow water into the piping. This is most typically accomplished with smoke or heat detection but can be done with any fire alarm signal including a manual fire alarm box. Preaction systems may be set up to incorporate additional strategies to prevent accidental, nonfire release of water. There are three fundamental types of preaction systems, as illustrated in the following.

12.6.4.1 Noninterlock A noninterlock system uses a deluge valve where either a sprinkler or fire alarm signal such as a smoke detector will open the valve. Without detection activation, the system behaves like a dry pipe system. If a sprinkler fuses, water will flow. If a smoke detector activates, the piping network will fill with water, but no water will flow until a sprinkler fuses.

12.6.4.2 Single Interlock A single interlock system requires the activation of a smoke detection system to operate a solenoid valve. Once the solenoid valve opens, water will enter the sprinkler piping network, but discharge will not occur until the sprinkler fuses. Operation of the solenoid valve alone turns the system into a wet pipe sprinkler system.

A data center may fall into this category if a faster response time from the sprinkler system is desired or it will take too long for water to reach a sprinkler once the sprinkler fuses.

12.6.4.3 Double Interlock A double interlock system requires the activation of a smoke detection system and the fusing of a sprinkler to allow water into the sprinkler piping. Both the solenoid valve and the deluge valve must open to admit water. Operation of the solenoid valve alone turns the system into a dry pipe sprinkler system.

This application includes conditions in which it would be hazardous to have water in the piping for an extended amount of time such as an unoccupied or remote site where response will be delayed or for sites that cannot tolerate overhead water except in an emergency condition. Both noninterlock and single interlock systems admit water; therefore, the sprinkler system could remain charged for some time before it is drained and reset. A double interlock system will not admit water into piping until a sprinkler fuses.

12.6.4.4 Galvanized Piping Although galvanized piping has historically been used in dry and preaction sprinkler system piping, a recent study [11] suggests that galvanized steel corrodes more aggressively at localized points in the piping compared to unprotected steel, which corrodes over a more uniform distribution. This can result in pinhole leaks in locations that are precisely designed to avoid water except in a fire condition.

When an air compressor is used, oxygen is continually fed into the system as the system maintains pressure, interacting with trapped water to corrode the system from the inside out. To combat the effects of corrosion, nitrogen or “dry air” can be used in lieu of air. When using a nitrogen-inerting system, the same study suggests that corrosion is virtually halted and that performance between galvanized and black pipe is roughly identical.

12.6.5 Water Mist

Water mist systems are based on the principle that water is atomized to a droplet size of no larger than 1 mm (0.04 in.). The large surface area to mass ratio results in a highly efficient heat transfer between hot gases and the water droplets, and a large amount of heat is absorbed with a relatively small amount of water. Water mist systems were initially researched in the 1950s as “fine water sprays” [12, 13] but resurfaced in the 1990s in response to the search for halon system alternatives.

Water mist systems have been tested and approved for use in computer room subfloors and for in-cabinet suppression systems. One advantage is that a properly designed water mist system can achieve fire protection equivalent to standard sprinklers but using a third or less water than a sprinkler system. Therefore, if a fire occurs, the collateral damage that could be caused by discharged water may be reduced, compared to sprinklers. However, accumulated water droplets are conductive, and accumulated moisture on circuit boards will cause problems for electronics. Where electronic equipment is likely to suffer irreversible damage due to water deposition, clean agent suppression systems are typically preferred over water mist.

Water mist systems utilize higher pressure to generate smaller water droplets than standard sprinklers. Water mist systems are designed to operate at pressures of anywhere between 175 psi (12 bar) and 2300 psi (158 bar). Pressures in this range require positive displacement pumps and high-pressure stainless steel tubing and incur higher materials and installation costs than standard sprinkler systems. Corrosion-resistant tubing, such as stainless steel, and low-micron media filters are critical to prevent the plugging of small orifices in water mist nozzles.

NFPA 750 [14] should be consulted for the application of these types of systems. The most important requirement of NFPA 750 is that the water mist system design must be based on fire testing to a test protocol that matches the actual application. Therefore, a water mist system that has been tested and approved for machinery spaces would not be approved for application in a data center.

The current status of water mist systems is accurately summarized by J.R. Mawhinney as follows: “Although FM Global has shown support for the use of water mist for telecommunication central offices, general acceptance by end users has been slow in North America. Similarly, the use of water mist as a halon replacement for computer rooms has been mixed. The fundamental issue has to do with comparing the performance of total flooding gaseous agents that can penetrate into electronic cabinets with water mist that cannot extinguish a fire inside a cabinet, at least not in a total compartment flooding mode” [15].

12.6.6 Clean Agents and Gaseous Fire Suppression

Water and electricity don’t mix; and today’s centers are so critical that they often need to keep running even during a fire event. Since the early 1900s, several forms of gaseous fire suppression have been explored and tested to the point that their use is now very well documented and fairly well understood. A gaseous fire suppression system acts on several branches of the fire tetrahedron (Fig. 12.3).

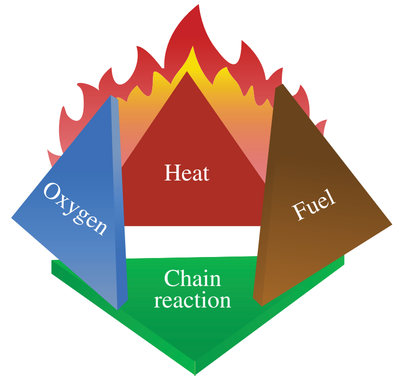


FIGURE 12.3 The fire tetrahedron. Courtesy of Wikipedia Commons, created by Gustavb, public domain.

Primarily, most agents displace oxygen, which slows down or halts the combustion process. A high specific heat allows many agents to remove heat, which would otherwise continue to accelerate combustion. Lastly, a more recent discovery has been the ability of some agents to interrupt the flame chain reaction. In reality, all three modes work together to suppress fire.

Clean agents are covered by NFPA 2001 [16], and the definition of a clean agent states that the material be nonconductive and nontoxic at concentrations needed for suppressing a fire and do not leave a residue. These properties make clean agents very attractive from an owner standpoint because these systems can protect a data center while allowing a relatively quick resumption of operations after an event.

Clean agents are typically designed as a total flooding system, meaning that upon detection or manual activation, a system will discharge the contents of pressurized cylinders into a defined volume to create a predesigned concentration necessary to extinguish a fire. Clean agents can also be used in manual fire extinguishers for local streaming applications.

Here are a few design considerations to be aware of regarding a total flooding application:

1. Total flooding is a one-shot approach; once the agent has been discharged, it will either suppress the fire or it won’t. If a fire begins to grow again, another suppression method will be required, such as an automatic sprinkler system or manual suppression. As previously stated, gaseous suppression is not a replacement for sprinkler protection.
2. NFPA 2001 requires that discharge be designed such that the agent reaches at least 95% of the design concentration within 10 s with a 35% safety factor for Class C (electrical) hazards and within 60 s for inert agents. This can require high-pressure systems depending on the agent. Systems require regular maintenance to ensure pressure requirements.
3. At least 85% of the design concentration must be maintained in the protected space for 10 min unless otherwise approved. This means either an extended discharge of agent or a very tight room that can maintain the concentration. Additional agent discharge is sometimes used to maintain mechanical mixing for the heavier agents.
4. Not meeting the room leakage requirement is one of the top modes of failure during commissioning. Further, rooms can be compromised by future construction and penetrations.
5. Most clean agents are super pressurized in liquid phase in containers and then expand to gas in the piping network under discharge. The design of these systems requires a balanced hydraulic design for two-phase flow. Many manufacturers provide their own proprietary software and design services as part of the installation costs.

6. A design involving HVAC shutdown can include the protected room as well as any above-ceiling or below-floor volumes. The volume perimeter must be established and maintained via tight construction and dampers.
7. In order to maintain cooling load, shutdown may not be desired; therefore, designs can include the air handling equipment and ductwork within the design volume of the protected space; however, the following should additionally be considered:
   a. Agent volume must be increased to include the volume of and leakage through the air handling equipment and associated ductwork. This can be a substantial increase.
   b. System sequence must be modified to keep the equipment running with outside air dampers closed to maintain concentration; the equipment should not need to be rated for elevated temperatures since the agent will prevent combustion throughout.

A chief benefit to this approach is that the room does not lose cooling during an event and the air handling system provides the necessary mechanical mixing to maintain concentration.

There are many gaseous suppressing agents on the market, each making independent claims of superiority. The best way to become educated regarding specific applications is to consult with manufacturers, vendors, or fire protection consultants. Although it is not possible to cover each agent in use today, common agents are discussed in the following.

12.6.7 Halon

Halon refers to a family of chemical compounds using halogens (predominantly Fluorine, Chlorine, and Bromine) to replace the hydrogen atoms in a typical hydrocarbon structure. NFPA 12A [17] covers Halon system installation requirements. Halon 1301 is to this day one of the best clean agents ever discovered because it requires such a low extinguishing concentration and provides deep vapor penetration. Its biggest drawback, however, is its classification as an ozone-depleting agent and the ensuing international regulations banning its manufacture. Since the issuance of the Montreal Protocol in 1987 and decree in 1993, halon has not been a viable choice for new installations.

That being said, the installation and maintenance of Halon 1301 systems are not banned, only its manufacture. Therefore, designers may encounter existing systems with halon from time to time. In fact, several vendors worldwide buy, store, and sell halon for reuse and many of these can be readily found online. When a halon system discharges, the owner faces the serious choice of whether to pay for replacement halon cylinders or to pull out the system and replace it with a similar clean agent.

When modifying an existing data center with a halon system, the right choice may be to modify the halon system in lieu of replacing it.

12.6.8 Hydrofluorocarbons

In response to the regulation of halon, DuPont developed a family of hydro fluoro carbon (HFC)-based clean agents that have a zero ozone depletion potential, the first of which was HFC-227ea, also known as FM-200. Other common HFC agents include HFC-125, sold under the trade name Ecaro-25, and HFC-23, also known as FE-13.

FM-200 is not as efficient at suppressing fire as halon, requiring on average about twice as much agent for the same hazard. It is also heavier than halon requiring additional mechanical mixing to maintain concentration. Due to the larger quantities required, existing piping systems and nozzles designed for halon cannot be reused. However, FM-200 and its family of other HFC agents have remained a popular alternative because of their similar attributes.

One of the downsides of using a halogenated clean agent is the possible production of Hydrogen Fluoride (HF), which is a by-product of HFC thermal decomposition. This means that if a fire is not knocked out quickly, the suppression agent could break down under heat. When reacting with water, HF turns into hydrofluoric acid, which is corrosive and highly toxic. For this reason alone, it is imperative that the agent achieve quick knockdown through conscientious design and that occupied spaces be evacuated during discharge.

12.6.9 Inert Gases

Inergen is one of the more widely used of the inert gases, so called because it is primarily made up of physiologically inert species including nitrogen and argon. Further, carbon dioxide is added to increase breathing rate.

Inergen suppresses fire through oxygen depletion. At 14% oxygen concentration by volume and lower, flaming ignition is no longer supported. Normally at this concentration, the physiological effects of hypoxia include confusion and loss of mental response; however, a small amount of carbon dioxide has been shown [18] to allow occupants to function for a period of time necessary for egress. The effects dissipate when the occupant is introduced to normal atmospheric conditions.

Inergen is relatively inexpensive when compared to other clean agents, and because it is made up of inert compounds, it will not break down into hazardous species or harm the environment. However, because it is so light and requires such a high extinguishing concentration, the volume of agent required is among the highest. Inert gases also require the highest delivery pressure among the clean agents, increasing the cost of piping.

Lastly, a long-term cost that should be taken into consideration for inert gases is the regular hydro testing of the system including cylinders, hoses, and piping. Due to the high pressures required, the system must be taken out of service for extended periods of time.

12.6.10 Novec 1230

Novec 1230 was released in 2004 by 3 M as a new type of clean agent known as fluoroketone. While Halon 1301 came under scrutiny in the 1990s for its ozone depletion potential, the HFCs such as FM-200 have come under fire for enhancing global warming. Novec 1230 goes by the technical designation of FK-5-1-12 and is also marketed under the trade name Sapphire. It has a zero ozone depletion potential due to its lack of bromine and advertizes a global warming potential (GWP) of 1 or less.

Like other clean agents, Novec 1230 absorbs heat and displaces oxygen. Design concentrations range from 4 to 8% depending on type of fuel. For data centers, the typical design concentration is approximately 5%. Pre-engineered systems are designed to meet the NFPA 2001 requirement of achieving 95% of design concentration in 10 s. Although Novec 1230 is a liquid at standard atmosphere and temperature, it readily vaporizes when discharged. As a heavy gas, it requires high discharge velocities and mechanical mixing during the 10 s release to achieve a uniform concentration throughout the enclosure.

A primary benefit of Novec 1230 is that, stored as a liquid, it can be transported via air and can be hand pumped from one container to another without significant agent loss to atmosphere. All other clean agents must be transported over ground and delivered under pressure.

12.6.11 Hypoxic Air (Reduced Oxygen)

An emerging technology that is used in some countries is hypoxic air, also known as an oxygen reduction system. A compressor/membrane system is used to reduce the concentration of oxygen in air to approximately 14%. At that concentration, studies [19] have shown that flaming ignition can be prevented. The low oxygen content, however, causes concern especially over prolonged exposure. Notably in the United States, the Occupational Safety and Health Administration (OSHA) does not permit anyone to work in a permanently hypoxic environment below 19% oxygen [20].

Unlike other agents, hypoxic air is not discharged during a fire condition, but is the constant atmosphere maintained within the protected space. The chief benefit is that there is no need for integration of fire detection systems to initiate the system. Also by reducing the oxidation process, products of combustion are not produced in the same quantity or at the same rate as they are in a normal 21% oxygen environment. Lastly, a hypoxic air system may have lower installation costs than the piping and cylinder network needed for a clean agent system.

Conversely, hypoxic air has not been accepted in the United States, and there are no U.S. standards that cover its use as a fire suppression agent. British Standard BSI PAS 95:2011 does address its use in the United Kingdom. Second, hypoxic air must be constantly generated, so it consumes energy to maintain the low oxygen concentration where other agents are stored in pressurized cylinders until needed. Oxygen sensors must be installed to control the system as the O_2 concentration seeks equilibrium with adjacent spaces. To be economical, the enclosure must be even more "airtight" than is required for total flooding clean agent systems, because the hypoxic air generation rate must exceed the leakage rate from the enclosure. Therefore, service and long-term operational costs must be considered. Furthermore, the systems require many hours to bring the oxygen concentration in a space down from normal ambient 21% oxygen to 14%. During the hours required to achieve the desired low oxygen level, the space is left unprotected. Lastly, the OSHA requires all the safety precautions associated with confined space entry for personnel who might need to enter the space. It may be necessary to ventilate the enclosure to bring the oxygen level up to at least 19% before anyone can enter the space. It may then require up to 24 h to reestablish the hypoxic level needed for fire prevention.

Hypoxic air is not a viable option in the United States for fire suppression in occupied spaces, but may be an option elsewhere.

12.6.12 Cabinet-Specific Suppression

A recent trend has been to provide cabinet-specific suppression in the form of clean agents or carbon dioxide (CO_2). CO_2 is another excellent suppressing agent; however, its use as a total flooding agent is limited by the fact that the concentrations needed for extinguishment far exceed human survivability levels. However, it remains a viable agent for localized suppression.

Cabinet suppression combines the one-shot principle of a suppressing agent without the high cost of a total flooding system. For equipment that may be susceptible to higher energy loads and serving as a possible source of ignition, these systems provide quick knockdown and local application before other systems could be activated. The downside is that, similar to portable fire extinguishers, these systems cannot maintain any extended concentration. If the equipment is still energized, the source of ignition and fuel has not been removed, and the fire will continue to grow. These systems are best used when the subject equipment can be de-energized as a portion of the sequence of operation and when system activation will be quickly investigated.

12.6.13 Portable Fire Extinguishers

NFPA 10 [21] is the standard governing selection and placement of portable fire extinguishers, regardless of suppressing agent. The standard breaks fire hazards into four categories:

- Class A: Cellulosic, combustible materials
- Class B: Flammable and combustible liquids
- Class C: Electrical fires
- Class D: Metal fires

Data centers represent a combination of Class A and C fire hazards. For sensitive electronic equipment, the standard requires selection from types of extinguishers listed and labeled for Class C hazards. Dry chemical fire extinguishers are expressly prohibited because the solid powder can irreversibly damage sensitive electronic equipment. That is why the most common portable extinguishers for data centers are those containing clean agents or carbon dioxide.

As with any manual suppression component, a portable fire extinguisher is only as effective as the person using it. Factors that limit the value of portable fire extinguishers in data centers include training, human sensitivity to an incipient stage fire, the length of time to reach an extinguisher, and access to equipment.

12.6.14 Hot/Cold Aisle Ventilation

The recent trend to provide hot and cold aisle containment for energy efficiency can have negative effects on fire suppression systems. For sprinkler systems, this primarily includes obstructing the sprinkler pattern at the ceiling. For total flooding systems, this could mean a delay to reach concentration at the source of ignition if the agent is released in a different aisle.

Aisle partitions or curtains are sometimes designed to drop in a fire condition either via thermal or signal response. This can blanket equipment upon release and slow extinguishment if not properly designed. It is imperative that the design team coordinate this prior to installation.

12.6.15 Summary

There are many gaseous fire-suppressing agents and blends on the market. Table 12.3 highlights the agents discussed in this section, and Figure 12.4 provides a rough comparison of agent needed based on the same volume of protection.

12.7 DETECTION, ALARM, AND SIGNALING

A traditional fire alarm system provides early warning of a fire event before the fire becomes life threatening. In the case of data centers, the threat to life is relatively low

TABLE 12.3 Comparison of common gaseous fire suppression agents [22]

Agent	Halon 1301	HFC-227ea	HFC-125	IG 541	FK 5-1-12
Trade name	Halon 1301	FM-200	FE 25	Inergen	Novec-1230
Type	Halogenated	Halogenated	Halogenated	Inert	Fluoro ketone
Manufacturer	NA	DuPont	DuPont	Ansul	3M
Chemical formula	CF_3Br	C_3HF_7	C_2HF_5	$5N_24ArCO_2$	$C_6F_{12}O$
Molecular weight (g/mol)	149	170	120	34.4	316
Specific volume: m^3/kg (ft^3/lb) at 1 atm, 20°C	0.156 (2.56)	0.137 (2.20)	0.201 (3.21)	0.709 (11.36)	0.0733 (1.17)
Extinguishing concentration[a]	5%	7%	9%	38.5%	4.7%
NOAEL concentration	5%	9.0%	7.5%	52%	10%
LOAEL concentration	7.5%	10.5%	10%	62%	>10%
Vapor weight required kg/100 m^3 (lb/1000 ft^3)[b]	44 (28)	74 (46)	66 (41)	66 (656)[c] $m^3/100\ m^3$ (ft^3)	91 (57)
Minimum piping design pressure at 20°C, bar (psi)	42 (620)	29 (416)	34 (492)	148 (2175)	10 (150)
Ozone Depletion Potential (ODP)	12.0	0	0	0	0
Global warming potential (100 years) relative to CO_2	7030	3220	3500	0	1
Atmospheric lifetime (years)	16	34	32	NA	0.038

LOAEL, lowest observable adverse effect level; NOAEL, no observable adverse effect level.

[a] Actual extinguishing concentration depends highly on fuel. Values represent common concentrations associated with fuels typical of a data center and are listed for comparison purposes. Engineering evaluation should be performed for each case.

[b] Design calculations per NFPA 2001, 2012 edition, Chapter 5 at sea level, including a safety factor of 1.35 for Class C fuels; values are for comparison only and not to be used for design applications.

[c] Inert gases use a different formula and are represented as a volume fraction including a 1.35 safety factor.

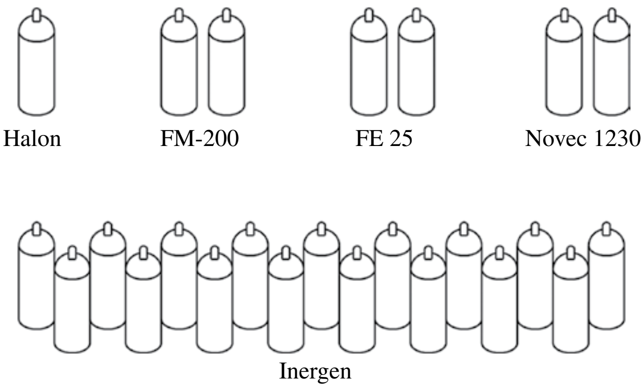


FIGURE 12.4 Comparison of clean agent quantities required for the same protected volume.

because of sparse occupancy and moderate fuel load; therefore, the building and fire codes don't typically require fire alarm systems unless occupant loads are higher than 500 for single-story buildings or 100 for multistory buildings. So fire alarm and signaling systems have evolved to provide early warning not only for life-threatening fire events but for property and operational loss events as well. In fact, this is part of the reason for the name change of NFPA 72 [23] from National Fire Alarm Code® to the National Fire Alarm *and Signaling* Code as of the last 2010 code cycle.

Documents such as TIA 942 and the Uptime Institute's *Tier Standard: Topology* recommend detection and fire alarm for Tier III and IV centers, as do FM Global data

TABLE 12.4 Sample portion of a fire alarm matrix

	System outputs	Notification					Suppression control				
		Actuate audible/visual alert	Actuate audible/visual alarm	Actuate abort alert signal	Actuate supervisory signal	Actuate trouble signal	Activate 30 s delay	Release clean agent	Release preaction solenoid	Interrupt countdown cycle	Resume countdown cycle
	System inputs	A	B	C		D	E	F	G	H	I
1	Air sampling Alert	●									
2	Air sampling FIRE		●				●				
3	Termination of delay cycle		●					●			
4	Smoke detector activation		●				●				
5	Manual fire alarm box		●								
6	Manual clean agent switch		●					●			
7	Activation of abort switch			●						●	
8	Release of abort switch										●
9	Water flow switch		●								
10	Tamper switch				●						
11	Lockout valve				●						
12	Preaction Alarm		●						●		
13	Preaction supervisory				●						
14	Preaction trouble					●					

sheets. Therefore, for data centers, fire alarms provide emergency signals to serve the following functions:

1. To detect a fire event, system trouble, or supervisory event at some predetermined risk level. This could include any of the following:
   a. Smoke detection
   b. Heat detection
   c. Sprinkler system activation
2. To alert occupants, owners, and off-site monitoring services to varying degrees of a fire event, system trouble, or system supervisory event via signals such as the following:
   a. Prealarm alert to on-site personnel
   b. Notification appliances in the building
   c. Wired and wireless communication to off-site personnel and responders
3. To initiate various sequences of suppression system activation, as programmed, such as the following:
   a. Releasing a solenoid valve for a preaction sprinkler or clean agent system
   b. Initiating a countdown for a clean agent system

In practice, the type of system is driven by the perception of risk. Items to consider include risk to life safety, operational continuity and business interruption, and property/equipment loss. A small server room wherein the aggregate risk is low may not need any form of detection. If the owner wants some form of alert prior to sprinkler system activation, a single smoke detector can be installed to sound alarm.

For more complex and critical sites, a dedicated smoke aspirating system will provide very early warning of smoldering conditions before alerting occupants of a prealarm condition. If combustion continues, the system will go into alarm and may initiate an optional 30 s delay prior to opening a solenoid valve to a preaction system. If the sequence is not aborted or the conditions in the room deteriorate, then either clean agent or sprinkler water will discharge.

12.7.1 Heat Detection

Heat detection is used in spaces where smoke detection would not be practical, such as dirty or freezing locations. When a room is sprinkler protected, it is typically not necessary to additionally provide heat detection as the water flow switch from a sprinkler system will initiate alarm, unless earlier thermal warning is desired.

12.7.2 Smoke Detection

The most popular mode of smoke detection within buildings is the spot-type photoelectric or ionization smoke detector. These detectors have a low cost and high level of reliability and can be programmed for different sensitivity levels. When combined with an addressable fire alarm system, specific responses can be programmed including the following:

- Prealarm/alert
- Fire alarm
- Timer initiation
- System discharge

These detectors also have their drawbacks, most notably that the initial sensitivity level is quite high compared to other technologies; by the time a spot-type detector detects smoke, a fire originating from electrical equipment may have been already caused damage to that equipment. Second, dirt and dust accumulate in the sensing element and further affect the sensitivity of the detector over time. For this reason, spot-type detectors must be located such that they are always accessible for servicing, which can have an impact on operational costs. Lastly, ionization-type detectors are sensitive to air velocity and are typically not listed for use in environments exceeding 300 ft/min (1.5 m/s). Listings exist for different ranges of velocities, but designers need to be aware of placing ionization detectors in plenums or under floor spaces.

A popular smoke detection system for data centers is the aspirating or air sampling smoke detector also known as high-sensitivity smoke detection (HSSD). An early manufacturer of this type of system and one that has become synonymous with the technology is the very early smoke detection apparatus (VESDA). There are many manufacturers and specialized applications for these systems, and this document will only address the basics of how these systems work.

In contrast to spot-type detectors, aspirating systems provide a much earlier warning because the sensitivity of the detector is so much higher. A network of piping is arranged somewhat similar to how a sprinkler system would be laid out, except that this network carries no agent; instead, it continuously aspirates the air through a series of sampling ports and feeds this air into a detector. Sampling points can also be positioned at air inlets or outlets to measure air quality at those points. The systems are modular in that a single detector can accommodate a certain length of pipe with a certain number of sampling ports and additional detectors can be added to accommodate larger spaces or separate zones, such as under floor or interstitial spaces.

The systems use proprietary laser technology to analyze the reflection that particles make when they pass through a laser chamber. The detector measures a percentage of obscuration per lineal foot depending on the manufacturer's algorithm. For example, if a detector is set to initiate an alert condition at an obscuration of 0.25%/ft and that level of obscuration were consistent across the entire room, an occupant 40 ft away would perceive an obscuration of 10%. In reality, an overheating circuit only produces a localized level of obscuration, which is not typically perceptible to occupants, but would be picked up by the closest sampling port. Most manufacturers offer a number of preset obscuration levels that will initiate a certain signal. These preset levels can be manipulated based on the geometry and contents of the protected space. One example may look like this:

- Alert – 0.25%/ft
- Action – 0.50%/ft
- Fire 1 – 0.625%/ft
- Fire 2 – 1.0%/ft

In addition to providing a higher level of sensitivity, air sampling systems are not susceptible to high-velocity flows the way ionization-type detectors are. Depending on the manufacturer, air sampling systems have differing levels of reliability in terms of segregating dust particles from actual combustion particles. One complaint is that after brief use in a new installation, the user experiences nuisance alarms and requests the sensitivity levels changed or ignores the alert and action warnings intended to protect the equipment. A thorough understanding of what these systems are intended for is highly recommended before purchasing. Maintenance includes cleaning sampling points and changing out filters. Frequency ranges in between 6 and 60 months depending on the environment of the space.

Another type of detection is gas detection. For example, some types of batteries produce hydrogen when they charge. Battery rooms are typically ventilated to avoid a buildup of hydrogen; however, the owner may install hydrogen detection as an additional precaution. Gas detection is also used in the semiconductor industry for rooms using or storing hazardous production materials (HPM).

12.7.3 Sequence of Operation

The sequence of operation is the set of logical functions that a fire alarm will follow based on inputs from devices. A portion of a sample fire alarm event matrix for a data center is shown in Table 12.4. This is not to be construed as a complete sequence of operation for a fire alarm system.

12.7.4 HVAC Shutdown

As a basis of design and in accordance with mechanical codes and standards, most HVAC systems with a return air system larger than 944 l/s (2000 cfm) are required to shut down when smoke is detected via a duct smoke detector. The duct detector is not required when the entire space served by that unit is protected with smoke detection. It is important to coordinate this code requirement with the operational need to keep the data center equipment ventilated to avoid damage. When an HVAC system is dedicated to a single room, it can be argued that automatic shutdown is not necessarily the best choice, because the intent of system shutdown is primarily to keep a fire from spreading to other portions of the building. Shutdown also protects the HVAC equipment from damage, but in some cases, it may make sense to leave the equipment running, at least for a period of time. Examples include the following:

- An alert-type alarm that does not deteriorate to a fire condition
- While data center or telecommunications equipment is being shut down
- Where the HVAC system is used to assist in mixing a clean agent

12.8 FIRE PROTECTION DESIGN

A good fire protection strategy will include a thorough evaluation of anticipated risk and will continue to evaluate that risk through the life of the building. Starting with the user and working through all the stakeholders, the strategy would determine how critical the site is and what effect a complete loss would have on the organization. The effect may be minimal due to other systems in place. In this case, the designer's role is to determine the appropriate authorities having jurisdiction and meet minimum code requirements.

On the other side of the spectrum, a fire event may have an enormous consequence in lost production or service. In this case, the right system may include a combination of fire suppression, detection, and alarm approaches. It will give the user the right information at the right time and focus protection on the most critical components.

Communication and consensus on goals with all vested parties will result in a successful protection strategy. Once the goals are agreed upon, the technical portion of design can start, a summary of which follows. The design team should determine the following:

- Site requirements (e.g., separation from other buildings on a campus)
- Type of construction
- Occupancy classification
- Occupant load
- Egress requirements
- Hazards analysis (e.g., batteries, generator fuel, storage)
- Suppression systems requirements
- Detection, fire alarm, and emergency communication systems requirements
- Interconnection with mechanical and electrical systems

Each data center or telecommunications room carries a certain level of risk that must be mitigated. As described in this chapter, the level of fire protection in these rooms can often exceed code minimums. Risk is evaluated in terms of life safety, operational continuity, and property/equipment value. Then the tolerance for that risk is addressed. This is where the various standards and previous experience of the design team will drive the design.

Fire protection design should be incorporated early into the design along with other systems. A coordinated team including vendors, manufacturers, and fire protection consultants experienced with life safety codes and integrated fire protection systems will help the team make the best decision based on established design criteria.

REFERENCES

[1] 2012 International Building Code. Country Club Hills: International Code Council, Inc.; 2011.

[2] 2012 International Fire Code. Country Club Hills: International Code Council, Inc. 2011.

[3] NFPA 75. Standard for the Protection of Information Technology Equipment. Quincy: National Fire Protection Association; 2008.

[4] Telecommunications Infrastructure Standard for Data Centers. Annex G, Tables 9-11. Arlington: Standards and Technology Department; 2012. ANSI/TIA-942-A-2012.

[5] Uptime Institute professional Services. Data Center Site Infrastructure Tier Standard: Topology. New York: Uptime Institute; 2012.

[6] NFPA 101. Life Safety Code. Quincy: National Fire Protection Association; 2008.

[7] ASTM E-119, 2012. *Standard Test Methods for Fire Tests of Building Construction and Materials*. West Conshohocken: ASTM International; 2012.

[8] United States Gypsum Corporation. *Gypsum Construction Handbook*. 6th ed. Kingston: R.S. Means Company, Inc.; 2009.

[9] NFPA 13. Standard for the Installation of Sprinkler Systems. Quincy: National Fire Protection Association; 2009.

[10] Hall JR. U.S. Experience with Sprinklers. Quincy: National Fire Protection Association Fire Analysis and Research Division; 2012. p 6.

[11] Kochelek J. White Paper, Mission Critical Facilities—Is the Use of Galvanized Pipe an Effective Corrosion Control Strategy in Double Interlock Preaction Fire Protection Systems. St. Louis: Fire Protection Corrosion Management, Inc.; 2009.

[12] Braidech MM, Neale JA, Matson AF, Dufour RE. *The Mechanism of Extinguishment of Fire by Finely Divided Water*. New York: National Board of Fire Underwriters; 1955. p 73.

[13] Rasbash DJ, Rogowski ZW, Stark GWV. Mechanisms of extinction of liquid fuels with water sprays. Combust Flame 1960;4:223–234.

[14] NFPA 750. Standard on Water Mist Fire Protection Systems. Quincy: National Fire Protection Association; 2009.

[15] Mawhinney JR. *Fire Protection Handbook*. 20th ed. Quincy: National Fire Protection Association; 2008. p 6–139.

[16] NFPA 2001. *Clean Agent Fire Extinguishing Systems*. Quincy: National Fire Protection Association; 2011.

[17] NFPA 12A. Standard on Halon 1301 Fire Extinguishing Systems. Quincy: National Fire Protection Association; 2008.

[18] Research Basis for Improvement of Human Tolerance to Hypoxic Atmospheres in Fire Prevention and Extinguishment. Philadelphia: Environmental Biomedical Research Data Center, Institute for Environmental Medicine, University of Pennsylvania; 1992. EBRDC Report 10.30.92.

[19] Brooks J. Aircraft Cargo Fire Suppression Using Low Pressure Dual Fluid Water Mist and Hypoxic Air. Gaithersburg: National Institute of Standards and Technology. NIST SP 984-2; NIST Special Publication 984-2.

[20] Occupational Safety and Health Administration, Respiratory Protection Standard, 29 CFR 1910.134; 2011. Available at https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=standards&p_id=12716. Accessed on June 19, 2014.

[21] NFPA 10. Standard for Portable Fire Extinguishers. Quincy: National Fire Protection Association; 2009.

[22] DiNenno PJ. Halon replacement clean agent total flooding systems. In: *SFPE Handbook of Fire Protection Engineering*. Quincy: Society of Fire Protection Engineers, National Fire Protection Association; 2008.

[23] NFPA 72. National Fire Alarm and Emergency Signaling Code. Quincy: National Fire Protection Association; 2009.

13

STRUCTURAL DESIGN IN DATA CENTERS: NATURAL DISASTER RESILIENCE

David Bonneville and Robert Pekelnicky

Degenkolb Engineers, San Francisco, CA, USA

13.1 INTRODUCTION

Natural hazards pose special design challenges for buildings that house data centers because of the value and fragility of the contents. Since building codes have as their primary purpose the protection of life safety (occupant safety or public safety), as opposed to business continuity, special structural design consideration is desirable for many data center buildings. This chapter provides an overview of the risks to buildings and contents due to natural hazards and addresses design considerations for critical buildings that may exceed those needed in the typical commercial building.

In the United States, all states currently adopt some edition of the International Building Code (IBC) for the regulation of structural design requirements. The IBC is based on a collection of adopted standards, including a load standard adopted from ASCE/SEI 7-10, *Minimum Design Loads for Buildings and Other Structures* [1], and various construction material standards, addressing, for example, design requirements for steel, concrete, masonry, and wood. ASCE 7 is primarily intended to provide protection against structural failure, which it does through four sets of performance objectives, known as *Risk Categories*, discussed in more detail later. Inherent within each of these categories are an assumed probability of failure and a resulting level of reliability that is considered appropriate for the intended occupancy of buildings assigned to that category. There is inherent uncertainty associated with each category due to uncertainties in the loading intensity, material strengths, and construction quality. While the protection against structural failure is the primary basis for design requirements within each category, there is also intent to provide some level of protection of property, at least at lower force levels, although this protection is not clearly defined.

13.1.1 Structural and Nonstructural Components

Referring to a building's resilience to natural disasters, the building is typically broken down into its structural system and the nonstructural components. The structural system consists of all the floor and roof-decks or slabs, beams, columns, foundations, and any load-bearing walls. The nonstructural components are everything else. Exterior cladding; mechanical, electrical, and plumbing equipment; piping; ductwork; access floors; and server racks are all considered nonstructural elements.

Although the structural system largely controls the overall building performance under natural hazards, it represents a small percentage of the total building value. In the case of an office building, the structural system may, for example, represent about 20% of the total building shell value, the remainder being controlled by architectural, mechanical, and electrical components and systems. Tenant improvements reduce the percentage further. In the case of a data center building, the structural system may represent as little as 5% of the total cost of the shell and core and a fraction of that when fully equipped. Since relatively modest increases in costs associated with enhancements in the structural system and anchorage of nonstructural components can have major impacts on performance, the return on this investment can be significant.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

13.1.2 Environmental Design Hazards

Obviously, a building must support its own self-weight and the weight of all the equipment and people inside the building. Those loads are typically referred to as gravity loads. In addition to gravity loads, buildings are designed to safely resist loads imposed by their anticipated exposure to natural hazards. Earthquake effects (ground motion, ground failure, and tsunamis) probably have the greatest impact on data centers, followed by wind loading including hurricane, typhoon, and tornado. The resulting flooding associated with both hazards also poses a significant risk. Snow and rain can also impose significant loadings on the building's roof. Both seismic and wind loads impose forces on the primary structural system and on nonstructural components, such as exterior cladding. In addition, earthquake loads affect interior equipment and utility systems.

13.1.2.1 Earthquake Effects In an earthquake, the amount of damage to the structural system and to the various nonstructural systems and components will vary significantly from building to building based on the intensity of the ground motion, the type and quality of structural system, the quality of the anchorage and bracing of nonstructural systems, the strength and ruggedness of the components internally and externally, and the quality of construction. As discussed in the following, it is possible to influence the damage and associated financial loss, as well as the loss of operations (repair time), through design processes that consider hazard level and performance more directly.

13.1.2.2 Tsunamis Many coastal regions are affected by tsunamis (ocean waves) created by earthquakes. There are two manners by which a tsunami affects a building. The first is the impact of the waves on the exterior walls and the forces that are imparted to the structure due to that. The walls can be significantly damaged or even blown out. The wave can generate a huge force that could also cause permanent lateral deformations to the structure or in the most extreme cases push the entire structure over. The second is the flooding that occurs due to the waves, which can cause significant damage to many mechanical and electrical components.

13.1.2.3 Wind Effects Wind, whether resulting from a hurricane, tornado, cyclone, or storm, affects data centers in a similar manner. As the wind blows against the exterior of the building, pressure is generated against the exterior cladding, which translates into the supporting structure. Additionally, there is typically an upward force generated on the roof as the wind blows over it. Typically, wind-induced damage affects isolated areas of the exterior or the roof, where local failures of the cladding, roof, or supporting framing occur. In more extreme cases, the entire structure can be deformed laterally or in rare cases completely blown over. An additional issue in windstorms is that strong winds can pick up objects and then blow them into buildings. The object that is made airborne is termed a missile, and its impact, or "strike," can damage exterior cladding, in particular windows.

13.1.2.4 Rain Effects Rain from storms only affects a data center if there is insufficient drainage on the roof to allow for ponding to occur or the envelope of the building is damaged, allowing water to seep into the building. Ponding on the roof occurs when there is a pooling of water due to insufficient or blocked drainage systems. The accumulation of water can become significant enough to overstress the roof framing, resulting in local collapse.

13.1.2.5 Snow Effects Snow, similar to rain ponding, primarily affects structures by overloading the roof framing. Snow drifting, where an uneven accumulation of snow occurs, or rain on snow, which increases the mass of the snow, can result in increased loading on the roof framing, potentially leading to local collapse.

13.1.2.6 Flooding Among the most destructive effects of a hurricane or tropical storm is the flooding in coastal regions due to storm surge and heavy rainfall. Flooding is the most common natural disaster in the United States and also results from other causes such as dam failures or river overflow. The Thailand Flood of 2011 resulted in over \$45 billion in economic damage, mostly associated with manufacturing facilities, and the 2012 flooding from Superstorm Sandy resulted in over \$30 billion in economic losses. The most significant effect of flooding is the water damage affecting the nonstructural components. However, in very significant floods, the floodwaters can act like waves and impact the structure, causing damage similar to tsunami waves.

13.2 BUILDING DESIGN CONSIDERATIONS

13.2.1 Code-Based Design

Because of geologic, environmental, and atmospheric conditions, the probabilities and magnitudes of natural disasters vary. In order to develop design and evaluation standards, the engineering community has selected "maximum" probabilistically defined events for each hazard, which are felt to capture a practical worst-case scenario for the specific region. For example, it is theoretically possible that a Magnitude 9 earthquake could occur in some parts of the country. However, the probability of that occurring is so remote, and the associated forces on the structure so great, that it becomes impractical to consider. On the other hand, it is not impractical to consider a Magnitude 8 earthquake in the San Francisco Bay Area given that there was a Magnitude 7.9 earthquake that occurred in 1906 and that geologic studies of the region indicate that the probability of such an occurrence, while small, is high enough that it should be used as the “maximum considered earthquake” for parts of that region. Conversely in Phoenix, the probability of a significant large earthquake is so remote that it is unnecessary to require consideration of a large-magnitude earthquake. Buildings in both regions, and the rest of the United States, are designed considering earthquake ground shaking that has a 2% probability of being exceeded in 50 years.

For structural design purposes, U.S. building codes and standards group buildings into Risk Categories that are based on the intended occupancy and importance of the building. Minimum design requirements are given within each of four such categories, designated as Risk Category I, II, III, and IV. The general intent is that the Risk Category numbers increase based on the number of lives affected in the event of failure, although higher risk categories can also offer greater protection against property damage and downtime. Risk Category II is by far the most common category and is used for most commercial and residential construction and many industrial buildings. Risk Category III is used for buildings that house assemblies of people, such as auditoriums; for buildings housing persons with limited mobility, such as K-12 schools; and for buildings containing hazardous materials up to a specified amount. Risk Category IV is used for buildings housing essential community services, such as hospitals, police and fire stations, and buildings with greater amounts of hazardous or toxic materials. Risk Category I is used for low occupancy structures such as barns. The ASCE 7 standard attempts to provide the higher performance intended with increasing risk category by prescribing increasing design forces and stricter structural detailing requirements for the higher risk categories. Naturally, construction costs tend to increase in higher risk categories as they do with increasing hazard level. Loads associated with each natural hazard are addressed separately in each Risk Category with the intent being to provide improved performance in the higher categories.

Data centers in accordance with code requirements generally fall into Risk Category II, suggesting that special design consideration is not warranted. This is consistent with the primary code purpose of protecting life safety while leaving the consideration of business risk to the owner. This introduces a building performance decision into the design process that is sometimes overlooked in the case of high-value or high-importance facilities, such as data centers. The desire to provide protection against substantial damage to the building and contents and the desire to protect ongoing building function would need to be addressed through performance-based design that exceeds prescriptive code requirements. Much of the desired protection of data center buildings, equipment, and contents can be achieved by treating them as Risk Category IV structures and by using performance-based rather than prescriptive code requirements.

13.2.2 Performance-Based Design Considerations

Most buildings are designed based on the design requirements specified in the standard for the applicable risk category of the building, and this is the approach that is taken where performance goals are not articulated by the building owner. In such cases, the performance expectations are not actually assessed by the design team, meaning that the owner’s performance goals relative to financial loss and downtime may not be addressed. Where building performance goals are more clearly understood, as is often the case for data centers, performance-based design requirements may be appropriate. For new building design, performance-based design may be used in two different ways. First, ASCE 7 permits alternative performance-based procedures to be used to demonstrate equivalent strength and displacement to that is associated with a given Risk Category without adhering to the prescriptive requirements. Such procedures may facilitate the use of more creative design approaches and allow the use of alternative materials and construction methods, resulting in a more economical design.

The second way that performance-based design is used may be more relevant to data centers since it relates more directly to the consideration of expected financial loss associated with damage to the building and contents and to the facility’s loss of operations after the event. Recently, a comprehensive methodology was developed for performance-based seismic design called FEMA P-58, *Seismic Performance Assessment of Buildings, Methodology and Implementation* [2]. The FEMA P-58 methodology involves simulating the performance of a given design for various earthquake events and characterizing the performance in terms of damage consequences, including life safety, financial loss, and occupancy interruption. The design can then be adjusted to suit the objectives of the owner.

13.2.3 Existing Buildings

Data centers are often housed in existing buildings that may have been constructed under older building codes using structural standards that have been superseded. This results in a broader range of performance expectations than defined for new buildings considering their risk categories. Many existing buildings do not meet the current Risk Category II requirements so that lower performance is expected. However, existing buildings may be upgraded to provide performance similar to new buildings designed to various risk categories.

The building codes are evolving documents. Every major disaster provides engineers with new information on how buildings perform and what did or did not work. Over the years, code requirements have become significantly more robust. In some cases, methods of design and construction that engineers once thought were safe and thus were permitted by code were found to be unsafe, and later editions of the building code reflected those findings. Also, as scientists study natural disasters, a greater understanding of their impact is realized, and this has often translated into improved design requirements.

This is not to say that all modern buildings pose little risk and all older buildings pose great risk. Performance of buildings, new or old, can vary considerably and is influenced by many factors. The type of structure chosen, the quality of initial design and construction, modifications made after the initial construction, and the location of the building can all affect the performance of the building. Because of that, the risk to a data center of a natural disaster, both in terms of life safety and financial loss, requires special attention.

Even in cases involving modern buildings, the design generally will not have specifically considered enhanced performance. Therefore, during the acquisition of a building for data center usage, it is important that the due diligence process includes for budgeting purposes an understanding of the vulnerability of the building to natural hazards, just as electrical and mechanical system requirements are understood.

13.3 EARTHQUAKES

13.3.1 Overview

The ground shaking, ground failures, and ensuing fires caused by major earthquakes have rendered parts of entire cities uninhabitable, as was the case in San Francisco in 1906, Port-au-Prince, Haiti in 2010, and Christchurch, New Zealand, in 2011. The vast majority of earthquakes, however, do not destroy entire cities, but still do considerable damage to buildings, transportation structures, and utility infrastructure. This can render a data center inoperable, either through damage to the physical structure or loss of critical utility services like power. The occurrence of major damaging earthquakes is relatively rare when compared to the lifetime of a data center facility but also unpredictable. Therefore, earthquakes exemplify the high-consequence, low-probability hazard. The occurrence may be rare, but the consequences are too great to be ignored in the design of a specific facility. Earthquake effects may represent the most challenging natural hazard for data centers that are exposed to them. This is due to the high value of equipment and contents that are susceptible to damage from shaking and the possibility that such damage may cause a loss of operation of the facility.

Building codes recognize the need to provide specific provisions to reduce the impact of a major earthquake on communities in locations prone to damaging events. Because earthquakes occur infrequently but produce extreme forces, our codes recognize that it is cost-prohibitive to design for buildings to remain damage-free in larger events. Therefore, provisions have been developed that will reasonably ensure life safety in a relatively large earthquake while accepting that there might be a very rare earthquake in which the building would not be safe. There has always been some consideration given to protecting function and property in lesser earthquakes, but no explicit provisions have been provided.

13.3.2 Earthquake Hazard

While individual earthquakes are unpredictable, there has been much study to document locations where earthquakes will have a greater likelihood of occurring and their maximum potential magnitude. Earthquakes most frequently occur on regions adjacent to boundaries of tectonic plates. The most common example of this is the region known as the Pacific "Ring of Fire," which runs along the western coast of North and South America, the eastern coasts of Japan, the Philippines, and Indonesia, and through New Zealand. Due to the relative movement of tectonic plates, stresses build up to the point where the earth crust fractures, resulting in a sudden release of energy. Typically, this occurs along previously fractured regions, which have been classified as fault lines. The San Andreas Fault, which runs along the western coast of northern California and then inland through Southern California, is an example of this. The Great 1906 San Francisco Earthquake occurred along this fault.

In addition, there is the potential for earthquakes to occur in regions within tectonic plates. These earthquakes, called intraplate earthquakes, occur because internal stresses within the plate cause fractures in the plate. An example of this is the New Madrid Seismic Zone near the tip of southern Illinois, which produced massive earthquakes in 1811 and 1812.

These regions, and specifically the faults, have been studied by geologists and seismologist to the extent that maximum estimates of earthquake magnitudes and probabilities of those earthquakes occurring in a given time frame have been established. That information is then translated into parameters that engineers can use to assess how an earthquake would affect structures, typically a representation of the maximum acceleration of the ground during the event. From that, information maps of earthquake hazard are produced, which provide engineers with information on the earthquake hazards that they should consider in the design or evaluation of buildings.

Since it is impossible to predict earthquake occurrence precisely, the hazard maps are typically based on some assumed probability of an earthquake occurring within a given time frame. In some locations, there is a possibility of not only a very large earthquake but also the possibility of more frequent, smaller, but still damaging events. In other locations, there is mainly the likelihood of having a rare but extremely damaging earthquake. The San Francisco Bay Area is an example of the former, and the middle Mississippi Valley (Memphis and St. Louis) is an example of the latter. This may be a factor to be considered in siting a new data center or evaluating an existing one.

13.3.3 Common Effects on Buildings

The release of energy in an earthquake translates through the ground and is expressed as horizontal and, to a lesser extent, vertical shaking at the surface. How the shaking propagates up through a structure affects how the structure will respond to the earthquake. Ideally, the structure when shaken would be undamaged, and the nonstructural components would remain securely fastened against shifting or toppling. Unfortunately, in the most seismically active regions, it is generally not economically feasible to design a structure to be so robust that it can withstand the largest possible earthquakes without damage. Therefore, engineers have recognized the need to design structures allowing damage to occur but in a controlled manner that does not compromise their overall stability.

As a structure is shaken, it may either deform in a ductile manner that absorbs energy or in a brittle manner allowing portions of the structure to fail. Older structures, particularly those designed before the 1970s or 1980s, often lack the features required to ensure ductile behavior. Consequently, those buildings and some marginally designed modern buildings can experience failures of structural connections or columns, which can result in partial or even large-scale collapse. Even when the failure is not significant enough to cause a collapse, it can damage a building to the point where the structure would be sufficiently weakened to make it vulnerable to aftershocks.

Even for structures that deform in a ductile manner during earthquakes, there is still the potential for damage. In some cases, it may be extensive enough to require repair prior to reoccupancy. Because of this, the engineering profession has adopted common performance states to describe a building's postearthquake state. They are illustrated in Figure 13.1.

The description of building earthquake performance is incomplete without discussion of how the nonstructural systems—the architectural cladding, finishes, furnishings, and mechanical, electrical, and plumbing equipment—are affected by the earthquake. Earthquakes can cause equipment to shift and topple if not anchored to the structure. Building codes provide requirements for anchorage design based on expected floor accelerations. Additionally, the swaying of the building can cause distribution systems, such as piping or ductwork, to break their seals, allowing contents to leak. Movement can also cause the building's envelope to break its watertight seals, allowing water intrusion in a rainstorm. Another damage consequence of building drift is the swaying of suppression systems, resulting in sprinkler head impact and resulting water damage.

Nonstructural damage can, and historically does, occur at earthquake intensities much lower than those that cause significant structural damage. This is of significant concern for a data center because most of the value and importance of the building is related to the equipment inside the structure.

13.3.4 New Building Design Considerations

For design purposes, earthquake ground motions are defined in terms of design accelerations given in hazard maps developed by the USGS and provided in ASCE 7. ASCE 7 requirements are based on seismic concepts developed in the *FEMA 750, NEHRP Recommended Seismic Provisions for New Buildings and Other Structures* [3], which is updated on a 5-year cycle. The maps provide *Maximum Considered Earthquake* ground motions that are adjusted (reduced) for design purposes and combined with factors related to dynamic response and system performance to provide design seismic parameters. Because buildings cannot be practically designed to resist large (design-level) earthquake ground motions while the structural system remains elastic, seismic design provisions incorporate ductility requirements that are intended to permit postelastic energy dissipation that may be accompanied by damage to the structural system, as well as nonstructural components. The general intent is to provide reasonable assurance that life safety is protected in Risk Category II buildings and that an enhanced level of safety is achieved in Risk Category III and IV structures, along with some level of protection against damage and loss of function. These inherent levels of performance are assumed within the standard, although the more clearly defined objective for typical buildings (Risk Level II) is collapse prevention in the very rare earthquake event, which is assumed to be achievable in most cases. The assumption is that the collapse probability is 1% or less in a 50-year period for Risk Category II buildings designed in accordance with the ASCE 7 Standard and that lower probabilities are associated with the higher Risk Categories.

Because a common building is only designed to prevent collapse and not to protect against damage in a large earthquake, new data centers should ideally be designed for higher performance. That higher performance should consider, at a minimum, a level that controls structural damage. Controlling damage is essential for ensuring that a data center does not get flagged as unsafe following a major

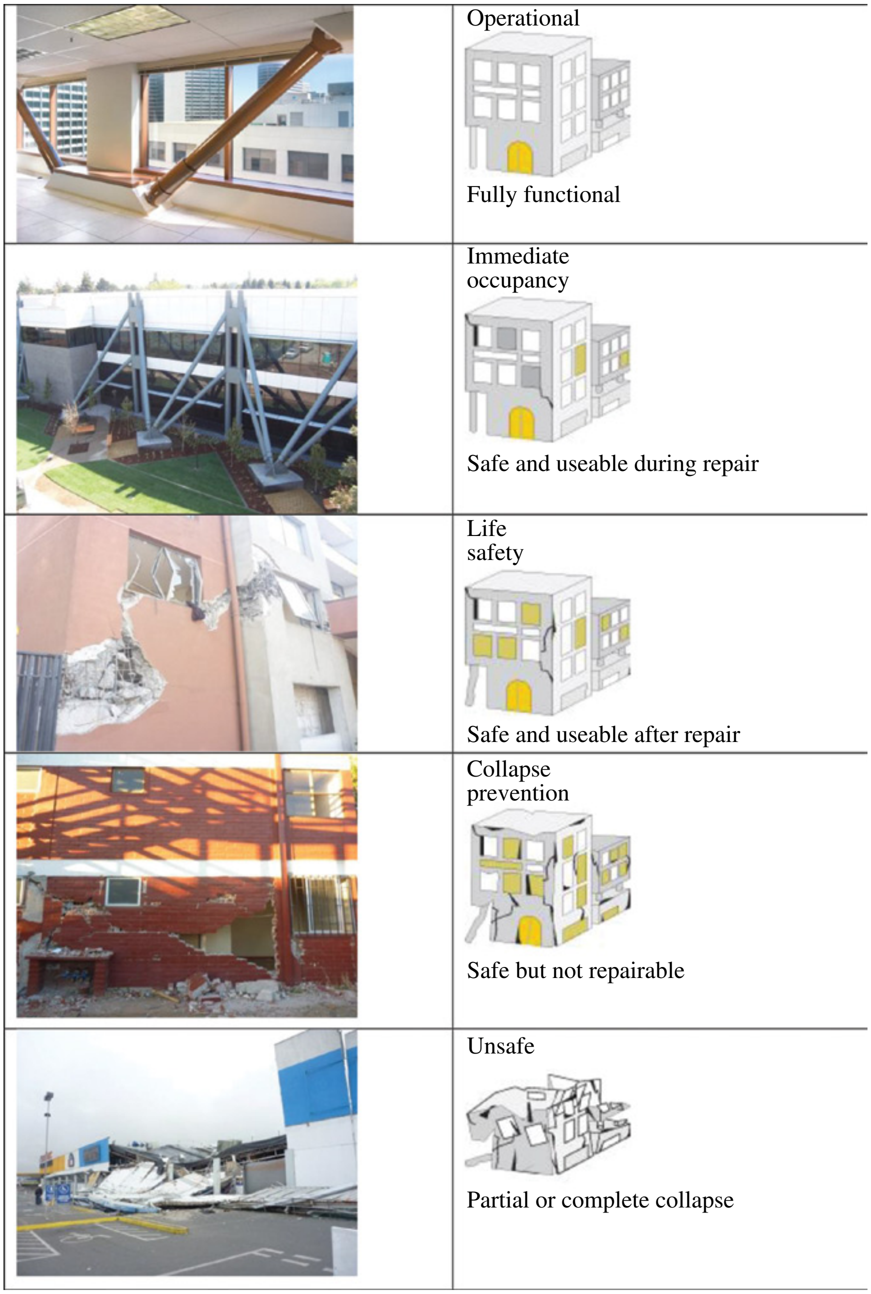


FIGURE 13.1 Building performance states. Courtesy and © Degenkolb Engineers, 2013.

disaster, something that could hinder reoccupancy. The repair time following a major earthquake may be quite long, also suggesting the need for enhanced design. In many instances, it is prudent to design a data center structure to the same level as essential facilities, such as hospitals and emergency operations centers. That would involve designing for the provisions of a Risk Category IV, as opposed to a Risk Category II, facility.

As previously stated, the equipment within a data center is more critical than the structure and therefore may require as much attention to design as the structure. In a typical building, most equipment is required to be seismically anchored if the building is located in a region of moderate or high seismicity. However, while such anchorage prevents the equipment from toppling, it does not provide assurance that the equipment will be functional following the event. U.S. codes now require seismic certification by the manufacturer for equipment in essential facilities (e.g., hospitals) if required to be operable after an earthquake and for equipment containing hazardous substances. Equipment that has undergone this testing is seismically certified and should be considered for use in a data center.

The most critical pieces of equipment in data centers, the servers and server racks units, are not normally seismically certified. Often, the electronic components inside are sensitive to large floor accelerations. One way to protect the servers is to place the server racks on isolated platforms or to isolate the entire access floor. Using seismic isolation technology decouples the equipment from the shaking of the floor it is situated on, greatly reducing the accelerations imparted to the equipment.

13.3.5 Existing Building Mitigation Strategies

When an existing building is considered for a potential data center site or if a data center is housed in an existing building, one of the first steps should be to ascertain the seismic performance expectation of the building. A commonly used standard for this evaluation is ASCE 41-13, *Seismic Evaluation and Retrofit of Existing Buildings* [4]. Unlike standards for new building design, which contain prescriptive design requirements, ASCE 41 recognizes that an existing building contains a combination of structural elements of varying strength and ductility and that many of those elements would not meet the standards for new buildings. In many instances, the evaluation will require detailed analysis procedures, beyond the scope of those used in the design of a new building.

ASCE 41 has five structural performance levels: Immediate Occupancy, Damage Control, Life Safety, Limited Safety, and Collapse Prevention. Life Safety is the standard commonly associate with a typical new building, while Immediate Occupancy is typically associated with a new Risk Category IV essential facility. When assessing an existing building, Life Safety should be the minimum standard, while Immediate Occupancy may be the desired level and Damage Control an acceptable level. It will be rare to find an existing building that meets Damage Control of Immediate Occupancy; therefore, the choice will be to accept a building that only meets Life Safety and contains no protection against long downtime following an earthquake or to seismically upgrade the building.

If upgrade is chosen, it is ideal to construct the retrofit before the building is outfitted as a data center. Retrofit of a vacant structure is significantly less expensive and has less risk. If the structure is already occupied, greater care is required in the retrofit design. It may be advisable to consider structural upgrade approaches that place new structural elements on the outside of the building and above the roof, so as to limit the amount of construction occurring over the servers. If work must be performed inside the facility, temporary protective boundaries should be constructed over the servers, and the retrofit measures should be laid out in a manner that avoids or at least minimizes conflicts with existing mechanical and electrical components.

Like the structure, existing nonstructural systems are typically not constructed to new seismic codes. While this is not as significant an issue, because most equipment in a data center is relatively new (<10 years old), it is still common for it to have been installed without proper consideration for seismic bracing and likely no consideration for ruggedness and postearthquake function. Therefore, it is likely that the equipment will need to be properly anchored. Because existing equipment cannot typically be tested for ruggedness, the choice will have to be made whether to keep the existing equipment in place and accept the risk of function loss, replace the equipment with certified equipment, or isolate the existing equipment. If preventing loss of function is paramount, then isolating the equipment would generally be less expensive than replacing the equipment.

13.4 HURRICANES, TORNADOES, AND OTHER WINDSTORMS

13.4.1 Hazard Overview

There are typically three different atmospheric phenomena that produce wind gusts of sufficient magnitude to affect structures. They are storms, hurricanes/cyclones, and tornadoes. Similar to earthquakes, regional conditions dictate the magnitudes of these hazards. Also similar to earthquakes, wind hazard occurs transiently and the magnitude cannot be precisely predicted. Because of this, scientists have developed models that can predict the probability of occurrence of a given wind gust of a certain size in a specified time frame or recurrence interval. For example, the current design wind for a typical building is based on a maximum wind gust that has a mean recurrence interval of 700 years.

Wind loads are defined in ASCE 7 by uniform recurrence interval wind speed contour maps that are provided separately for each of the four Risk Categories. The maps address all geographic areas of the United States including regions affected by hurricanes. The wind speeds provide pressures that are combined with factors related to exposure, topography, and directionality to provide design loading. The hurricane-prone region of the United States includes the Southern Gulf Coast region continuing up along the east coast. In wind design, the code-specified design pressures represent loads that the building could be expected to experience during a maximum wind event. So unlike seismic design, in which the maximum expected earthquake forces are reduced to account for energy dissipation through inelastic response, it is intended that building structural systems, including components and cladding, experiencing design wind loading remain elastic and are not substantially damaged. An exception is made in the case of tornadoes, which can generate extreme wind loads that are not generally covered by building codes.

Tornadoes are not generally addressed by the building code because the probability that a building will be located within the path of a tornado of sufficient strength to be destructive is very low. Currently, the only structures where tornado wind speeds are considered are major high-risk structures like nuclear power plants. This is not to say that tornado effects cannot be considered in the design of a data center. There are maximum tornado wind speed maps available for the United States. For areas in "Tornado Alley," the Great Plains states, the tornado wind speeds are approximately 75% greater than those typically considered in building design.

13.4.2 Common Effects on Buildings

As the wind blows against a building, pressure is generated against the exterior cladding, which is transferred to the supporting structure. Additionally, there is typically a suction force generated against the roof and the side and back walls of the building. These pressures must be resisted by the cladding and roof elements and the structural members supporting them. Failures of cladding panels, windows, and doors are common in extreme wind events. Not as common, but still observed, are failures of the roof sheathing and roof-decks. This occurs when the suction pressure is strong enough to cause the sheathing to pull upward.

In more extreme cases, the entire structure can be deformed laterally or for lighter buildings even completely blown over. For most engineered buildings, these types of failures are rare. However, if a data center is housed in a light-framed metal building, the potential for this does exist and should be evaluated.

Another design consideration in windstorms relates to strong winds picking up objects and blowing them into buildings. The object that is made airborne is termed a missile, and its impact, or "strike," can damage exterior cladding and, in particular, windows. If a missile breaks a window or part of a wall, wind can rush into the building, increasing the internal pressure, increasing the demands on the roof and walls, possibly leading to failure.

Rooftop equipment can be susceptible to damage in windstorms. If the equipment is not sufficiently anchored, it can be blown over. The equipment can also be damaged if a large missile hits it.

Other than structural collapse, the main causes of function loss following a major windstorm are damage to the building cladding and damage to the rooftop equipment. If the building cladding is damaged, the building is no longer "watertight." Therefore, rainwater can enter the building and damage equipment. Because maintaining function of the air-conditioning system is so important in a data center, the loss of rooftop air handling equipment may result in shutdown of the facility.

13.4.3 Mitigation Strategies

For new buildings, the most logical mitigation of wind hazards is to design the building assuming it is a Risk Category IV Essential Facility. This will require that the building be designed for higher wind forces than an ordinary building. If the building is located in an area with a high potential for tornadoes, it may be prudent to consider even higher wind forces in the design of the building and its cladding. Another consideration is to provide windows and doors that are resistant to impact from missile strikes. There is an ASTM Standard for these types of windows and doors, E1966. If the building must have rooftop-mounted equipment, the rooftop equipment could be surrounded by wind screens that are designed for the maximum wind forces and are of a material that can protect the equipment from missile strikes.

Existing buildings should be evaluated for the wind forces that a new building would be designed to. If the roof, framing members, cladding, or associated connections are overstressed, they should be strengthened. It is likely that the windows and doors are not resistant to missile impact and should be replaced if the tornado or hurricane hazard is high. Rooftop equipment will likely be unscreened or the existing wind screen inadequate. A new, compliant screen should be added or the existing screen upgraded.

13.5 SNOW AND RAIN

13.5.1 Hazard Overview

Similar to wind and earthquake, the snow and rain hazards vary based on the location of the building. The meteorological climate of the area will dictate the potential for major snowstorms or rain events. In both cases, the hazards are defined probabilistically. Maps can be found in building codes, which provide the design snow and rain levels, or those parameters can be obtained by site-specific studies.

The design rain hazard is commonly taken as the water accumulated on the roof from a storm with a 100-year mean recurrence interval. The rain accumulation is determined based on the slope of the roof, the types of primary and secondary roof drains, and whether those drains are blocked. There are a number of features on the roof, such as the presence of rooftop equipment, depressions in the roof, and the flexibility of the roof framing, which can lead to localized ponding of rainwater, causing greater than anticipated demands. Presently, U.S. codes and standards do not provide for increased rain loads for essential facilities.

For snow, the 50-year mean recurrence (or 2% annual probability of exceedance) interval ground snow fall is used as the basis for the snow hazard and is augmented based on the height of the roof and the presence of roof features such as parapets that can allow for snow drifts to form. Additionally, there are factors such as rain or snow surcharge that should be taken into account because the snow traps rainwater, causing an increase in the density of the snow. U.S. codes provide a factor that increases the design snow load for higher risk category facilities.

13.5.2 Mitigation Strategies

Addressing rains and snow hazards in new buildings is simply a matter of following the building code and applicable structural design standards, such as ASCE 7. Because of the critical nature of a data center, it is recommended that, as in the case of wind and earthquake, the provisions for Risk Category IV essential facilities be used. It is also important to put in place a maintenance plan that has the roof drains regularly checked for blockage and cleaned. Also to address the issue of power loss creating a "cold roof" condition, the heating system could be placed on emergency power that can operate for at least 3 days.

Many existing buildings that are desirable for data centers, industrial warehouses, and big box-type buildings are the most likely to have roofs that have very little reserve capacity and inadequate consideration of snow drifting or are flexible enough to create ponding instabilities. Therefore, it is essential to evaluate the roof framing of a building during a due diligence study. Augmenting the roof framing can be a very significant cost and will be difficult to accomplish if the building has already been outfitted with the systems, piping, ducts, and cable trays. All of those items will obstruct access to the roof framing. If new rooftop is equipment is planned, consideration of its effect on the roof drainage and the ability of snow to drift adjacent to it should be considered. The equipment may need to be relocated, placed on elevated platforms to allow for drainage, or have the roof locally strengthened under it.

13.6 FLOOD AND TSUNAMI

13.6.1 Hazard Overview

While the mechanisms that cause a flood and a tsunami are different, they are similar hazards and their effects on buildings are similar. Floods and tsunamis are characterized by the uncontrolled flow of water into a region. The force of the water impacting the building can damage the building's façade and, if the water flow is high enough and violent enough, even damage the building structure. Once the building envelop is compromised, water may flow into the building and can damage equipment and leave the building unoccupiable due to debris and environmental hazards such as mold.

Like all environmental hazards, flood and tsunami risk differs from region to region. The magnitude of risk is based on whether the building is situated in an inundation zone. That is a region that, for a given mean recurrence interval or probability of occurrence, might be subject to flooding due to the high water level. Based on the inundation height, it can be determined if the building is located at an elevation where floodwaters will impact it.

U.S. building codes address flood by requiring consideration of a 100-year mean recurrence interval flood. The Federal Emergency Management Agency (FEMA) publishes maps that provide flood inundations zones. If a building is located in an inundation zone, the designer or a consultant to the designer needs to determine if the loads from the floodwater impacting the building are significant. Currently, U.S. building codes do not require higher mean recurrence intervals be considered for essential facilities. There is some debate about this and professional opinions that the 500-year mean recurrence interval flood should be used for essential Risk Category IV buildings, instead of the 100-year flood.

At the present time, U.S. building codes do not explicitly address tsunami risk, though efforts are underway to develop prescriptive and performance-based requirements in the near future (ASCE 7-16). Tsunami risk and inundation maps published by the National Oceanic and Atmospheric Administration (NOAA) and some state agencies, such as the California Department of Conservation, show coastal regions subject to tsunami hazard. Many other countries with coastal regions subjected to tsunami risk also have such hazard maps. To assess the vulnerability of a site, the extent of inundation, height of run-up, and velocity of flow are needed. Where maps are not available for a specific location, site-specific studies can be performed.

13.6.2 Common Effects on Buildings

The most common effect of tsunami and flood on buildings is the inundation of the basement and first floor. The water damage can be significant and immediately renders a building or much of its equipment nonfunctional. For example, the Fukushima Daiichi nuclear disaster was initiated by tsunami waters flooding the rooms housing the emergency generators, rendering the plant without power and unable to maintain the coolant system. Additionally, after the water subsides, there can be issues with mold and other environmental hazards that would need to be mitigated before it is safe for people to reoccupy the building.

As stated previously, the floodwaters flowing into the building can damage the building's façade and even the structure if the floodwaters are high enough and have a sufficient flow velocity. Weak points in the building envelope such as windows and doors are the most susceptible. In the most extreme cases, a flood or tsunami can produce such a violent flow of water as to literally push a building off its foundation. Debris in the flowing water can also damage buildings significantly.

13.6.3 Mitigation Strategies

The most straightforward way to address the risk of tsunami or flood is simply to not build the data center in the inundation zone or locate a data center in an existing building in an inundation zone. When this is not possible, consideration can be given to the inundation height, the maximum water flow height and velocity, and the presence of surrounding elements, which could become waterborne debris hazards.

In many cases, the flood or tsunami flows will not be very high or violent. In those cases, mitigation may simply involve locating all equipment required for continued operation of the data center at a level above the inundation height. Locating critical equipment in the basement should be avoided. It is common in flood regions to build new buildings on platforms, so the first occupiable floor is above the inundation height. If the data center is of critical importance, consideration of the 500-year as opposed to the 100-year flood inundation should be given.

13.7 COMPREHENSIVE RESILIENCY STRATEGIES

13.7.1 Predisaster Planning

The first step in any natural disaster risk mitigation plan is to understand what the natural disaster hazards are at each facility in the company's inventory. Federal and local government agencies publish information on earthquake, hurricane, flood, snow, and tornado hazards. A simple but effective approach to begin a plan would be to develop a matrix of all the facilities' sites and rank the hazard for earthquake, tornado, hurricane, and flood as high, medium, or low.

Once an understanding of the hazard at each site is known, then the process of assessing the resilience of the different facilities on each site can occur. As noted previously, not all buildings, even those designed to the same code, will perform the same in a natural disaster. The code sets forth a minimum standard, but not explicit performance objectives related to downtime and damage control. Therefore, it is helpful to have a common approach to define building performance.

For any disaster, there are three main concerns related to facility performance: loss of life, physical damage, and downtime following the disaster. These three metrics may or may not be related to each other. For example, a facility could be "Life Safe" but sustain damage to the point of being not economically feasible to repair. Conversely, a building can have very little damage overall, but localized collapsed, resulting in several people killed or injured.

For many buildings, Life Safety is the only performance level that needs to be considered. Most office buildings fall into this category. The lives of the people inside need to be protected, but the building could be significantly, even irreparably damaged from a disaster. The workers may be able to work off-site until a new facility is found or the building is repaired. On the other hand, the building may house a critical, nonredundant data center, or manufacturing operation, the loss of operation of which would cause significant business impact. In those cases, the target performance level may be significantly higher than simply Life Safety, and considerations for minimal post-disaster downtime may need to be addressed.

It is important to understand what the needs are for each facility in the organization's inventory so the right performance level can be selected for each specific building. Standards can be tailored to a company's specific needs and should be agreed upon, at least in concept before any evaluation is to begin. Once the standards are set for which buildings need only to be Life Safe, which ones have critical functions and need to be immediately occupiable, and which ones need some level of damage control, then evaluations of the buildings can be carried out and mitigation strategies developed.

Evaluating building performance for different natural disasters does not have to be a major undertaking for each building. There are methods that can be employed to provide cursory assessments of all the buildings within a portfolio. Following the preliminary evaluation, it can be determined which buildings warrant more in-depth evaluation. Typically, more detailed evaluations are done for critical buildings and for buildings where the cursory evaluation indicates there may be a problem, understanding the range of conservatism built into the cursory evaluation methods.

Following or concurrent with the initial evaluation of the facilities, it is recommended that company-specific natural disaster guidelines be developed. These guidelines should set forth the minimum performance level for each type of facility in the company's inventory. The guidelines can then be used to direct new construction projects, set forth standards for prelease and prepurchase evaluations, and determine which current facilities are not up to the required performance.

It is important to avoid adding buildings to the inventory that represent moderate or high risk. Therefore, the guidelines should be used for all new construction projects and for assessment of any building the company plans to purchase or lease. Depending on the type of facility, the cost of added natural disaster resilience may be very small. However, without guidelines, building designers are frequently not aware that greater resilience is appropriate for a given facility.

For a typical office building, the structural cost makes up only about 20% of the total building cost, so making the structure more resilient may only add 5% or less to the total building cost. For manufacturing and data centers, the structural cost is an even smaller portion of the total building cost, and thus, added disaster resilience would cost much less as a percentage of the total cost.

When considering acquisition or lease a new building, a proper natural disaster due diligence study should be performed during the due diligence process. The risk study should focus on assessing the building's risk to life safety, damageability, and potential loss of function with respect to any significant natural disaster that may be present at the site. The company's risk guidelines should have a section that addresses acceptable risk levels for owned and leased buildings based on the occupancies and functions of the buildings.

For existing buildings currently in the company's inventory that do not meet the facility standards, there are four options: Retrofit, Replace, Insure, and Accept. Retrofitting to bring a facility up to the required performance may require significant structural modifications or may only involve addressing isolated deficiencies or bracing equipment. Structural retrofits can vary from modifying the structure in isolated areas to the addition of exterior buttress, augmenting existing member connections, or even additions of new structural elements to the interior of the building. Nonstructural elements, such as mechanical and electrical equipment, piping and ducts, and architectural elements, may need to be braced so they can stay intact during earthquake shaking or not be blown over by strong winds. Some nonstructural elements may need to be relocated so they are not located in an area that will be inundated with water if a flood occurs. In any retrofit, it is advantageous to perform the work when the building is vacant or in conjunction with a major tenant improvement. For cases in which that is not feasible, a retrofit can be designed to minimize the amount of temporary relocation and be constructed in phases or the new structural elements concentrated at the exterior of the building.

In some instance, the cost of a retrofit may be excessive and approach that of building a new facility. In certain cases, that may still be the appropriate business decision. In other cases, several options should be explored. One is to build a new, disaster-resilient facility. The other might be to build a second facility in another location, which can create sufficient redundancy so the loss of one does not significantly impact the company's business operations.

The last two options—insure and accept—are predicated on the cost of retrofit or replacement being too large to justify in conjunction to the risk exposure. Natural hazard insurance can be costly but in some cases may be sufficient to mitigate the natural disaster risk. In some cases, the time between the disaster and when the insurance claim is fully paid can be quite long. On the other hand, if the facility is redundant and does not pose a threat to the lives of the people inside, the company may choose to accept the risk and self-insure.

13.7.2 Postdisaster Planning

The moments after a major natural disaster can be chaotic. However, a well-developed postdisaster plan can serve to allow the immediate recovery to begin in spite of the chaos. There are a few important concepts that every postdisaster plan should have. The first one is to educate all employees to protect themselves during the disaster. For example, it is common for people to run out of the building during an earthquake. However, the more appropriate response, advocated by the FEMA among others, is to drop, cover, and hold. The second is having on-site personnel trained on how to properly inspect buildings to determine if there are any glaring safety hazards. The default position should be to evacuate and wait for an engineer or building official to evaluate the building to determine it is safe.

It can take weeks for a jurisdiction to inspect a specific facility. This is because the demands on the local building department, even when bolstered by volunteer engineers, are so great that response times are unpredictable. Additionally, finding a consulting engineer to hire may be difficult because of the increased demands on their time due to the disaster. Therefore, it is important to have in place prearranged retainer agreements with an engineer who can inspect the facility or multiple facilities. It is ideal for the retained engineer to have previously evaluated the facilities so as to have an understanding of them and where the potential damaged areas may be. This will make their evaluation much more effective and can also be used to pretrain the on-site personnel for specific hazards to be aware of.

In San Francisco, following the 1989 Loma Prieta Earthquake, a program was enacted in conjunction with the Structural Engineers Association of Northern California called the Building Operation Resumption Program (BORP). In this program, the building owner contracts with the evaluating engineer, who then prepares a postearthquake inspection plan that is submitted to the jurisdiction. The city officials then approve the plan and that engineer is registered and required to post the safety rating of the building within 3 days of the disaster: Green, Safe for Reoccupancy; Yellow, Only safe for limited reoccupancy by trained personnel and Red, Unsafe. While other cities do not have a specific program like the BORP, many have been willing to adopt building-specific BORP-like programs if the building owner brings a proposed program to the building official or planning department.

REFERENCES

[1] American Society of Civil Engineers. Minimum Design Loads for Buildings and Other Structures. Reston: ASCE; 2010. ASCE/SEI 7-10.

[2] Applied Technology Council (ATC). Seismic Performance Assessment of Buildings, Methodology and Implementation, FEMA P-58, developed for the Federal Emergency Management Agency, Washington, DC; 2012.

[3] Building Seismic Safety Council (BSSC). NEHRP Recommended Seismic Provisions for New Buildings and Other Structures, FEMA 750, developed for the Federal Emergency Management Agency, Washington, DC; 2009.

[4] American Society of Civil Engineers. Seismic Evaluation and Retrofit of Existing Buildings. Reston: ASCE; 2013. ASCE/SEI 41-13.

14

DATA CENTER TELECOMMUNICATIONS CABLING

Alexander Jew
J&M Consultants, Inc., San Francisco, CA, USA

14.1 WHY USE DATA CENTER TELECOMMUNICATIONS CABLING STANDARDS?

When mainframe and minicomputer systems were the primary computing systems, data centers used proprietary cabling that was typically installed directly between equipment. See Figure 14.1 for an example of a computer room with unstructured nonstandard cabling designed primarily for mainframe computing.

With unstructured cabling built around nonstandard cables, cables are installed directly between the two pieces of equipment that need to be connected. Once the equipment is replaced, the cable is no longer useful and should be removed. Although removal of abandoned cables is a code requirement, it is common to find abandoned cables in computer rooms.

As can be seen in the figure, the cabling system is disorganized. Because of this lack of organization and the wide variety of nonstandard cable types, such cabling is typically difficult to troubleshoot and maintain.

Figure 14.2 is an example of the same computer room redesigned using structured standards-based cabling.

Structured standards-based cabling saves money:

- Standards-based cabling is available from multiple sources rather than a single vendor.
- Standards-based cabling can be used to support multiple applications (e.g., local area network (LAN), storage area network (SAN), console, wide area network (WAN) circuits), so the cabling can be left in place and reused rather than removed and replaced.
- Standards-based cabling provides an upgrade path to higher-speed protocols because they are developed in conjunction with committees that develop LAN and SAN protocols.
- Structured cabling is organized so it is easier to administer and manage.

Structured standards-based cabling improves availability:

- Standards-based cabling is organized so tracing connections is simpler.
- Standards-based cabling is easier to troubleshoot than nonstandard cabling.

Since structured cabling can be preinstalled in every cabinet and rack to support most common equipment configurations, new systems can be deployed quickly.

Structured cabling is also very easy to use and expand. Because of its modular design, it is easy to add redundancy by (copying) the design of a horizontal distribution area (HDA) or a backbone cable. Using structured cabling breaks the entire cabling system into smaller pieces, which makes it easier to manage, compared to having all cables in one big group.

Adoption of the standards is voluntary, but the use of standards greatly simplifies the design process, ensures compatibility with application standards, and may address unforeseen complications.

During the planning stages of a data center, the owner will want to consult architects and engineers in order to develop a functional facility. During this process, it is easy to become confused and perhaps overlook some crucial aspect of data center construction, leading to unexpected expenses or downtime. The data center standards try to

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

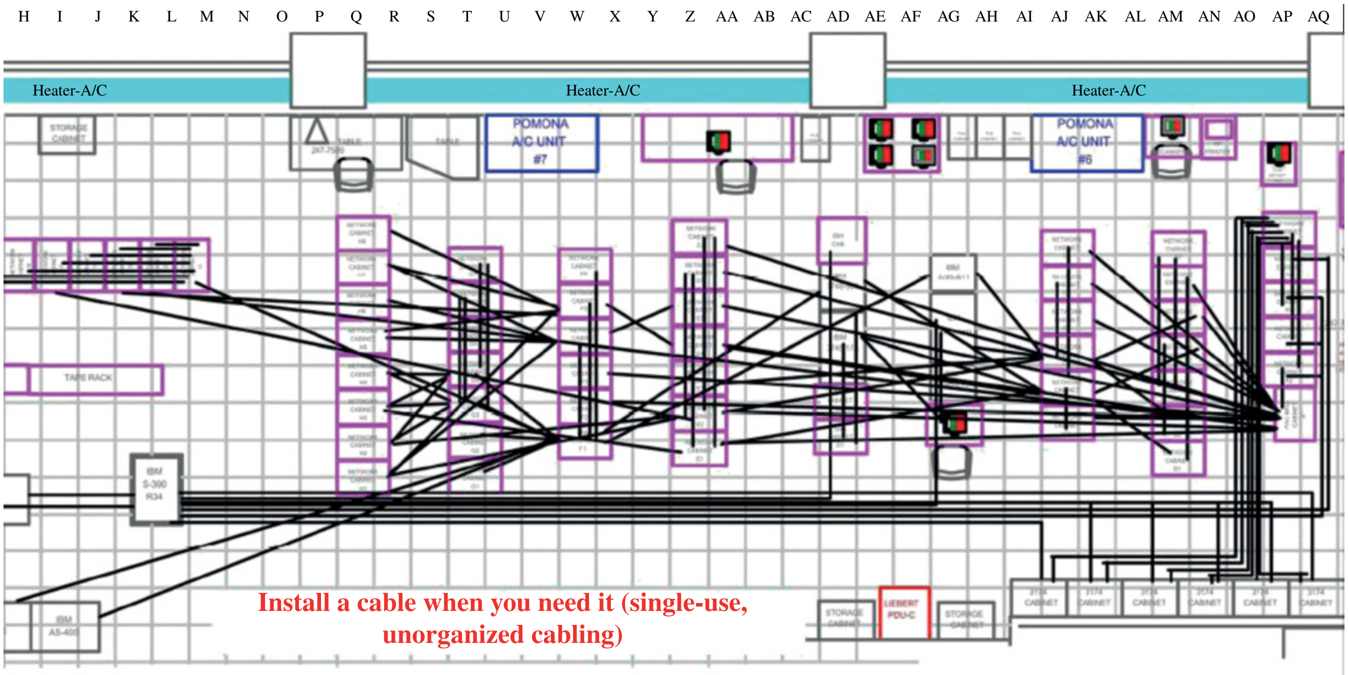


FIGURE 14.1 Example of computer room with unstructured nonstandard cabling.

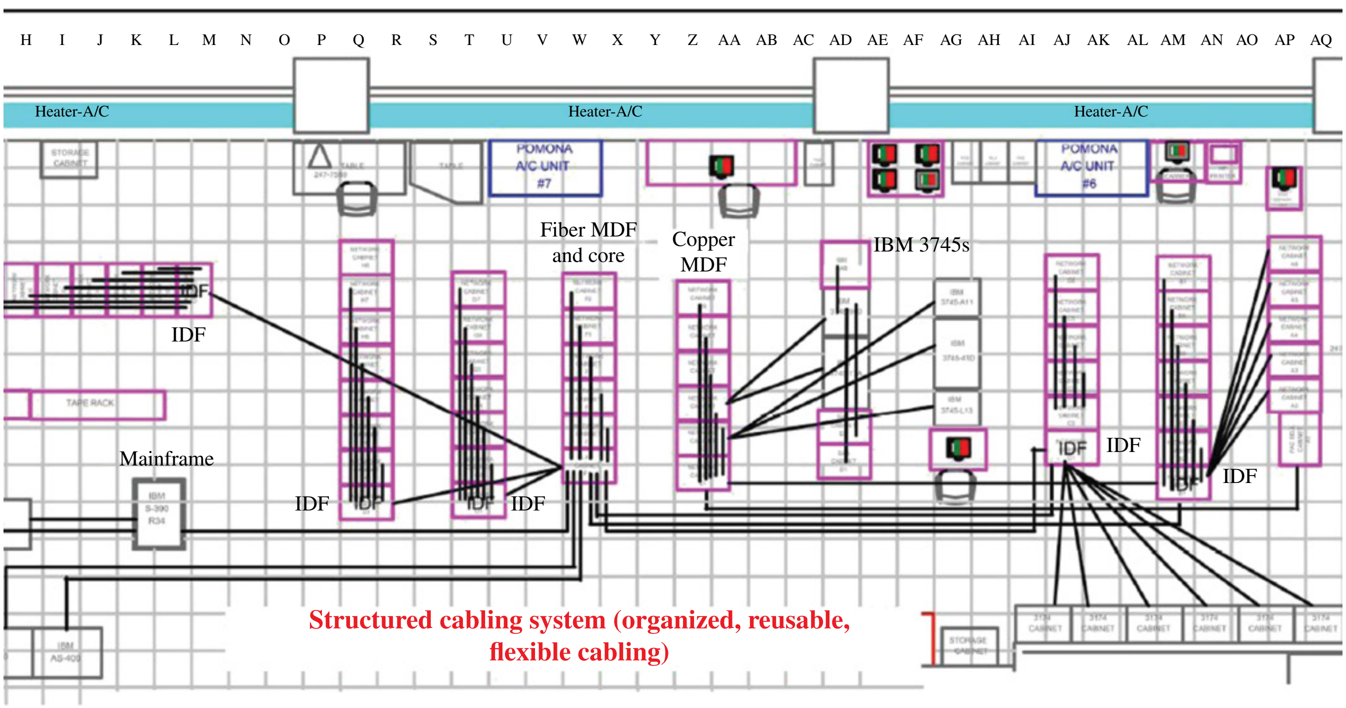


FIGURE 14.2 Example of computer room with structured standards-based cabling.

avoid this outcome by informing the reader. If a data center owner understands their options, they can participate during the designing process more effectively and can understand the limitations of their final design. The standards explain the basic design requirements of a data center, allowing the reader to better understand how the designing process can affect security, cable density, and manageability. This will allow those involved with a design to better communicate the needs of the facility and participate in the completion of the project.

Common services that are typically carried using structured cabling include LAN, storage area network (SAN), WAN, system console connections, out-of-band management connections, voice, fax, modems, video, wireless access points, security cameras, and other building signaling systems (fire, security, power controls/monitoring, HVAC controls/monitoring, etc.). There are even systems that permit LED lighting to be provisioned using structured cabling.

14.2 TELECOMMUNICATIONS CABLING STANDARDS ORGANIZATIONS

Telecommunications cabling infrastructure standards are developed by several organizations. In the United States and Canada, the primary organization responsible for telecommunications cabling standards is the Telecommunications Industry Association, or TIA. The TIA develops information and communications technology standards and is accredited by the American National Standards Institute and the Canadian Standards Association to develop telecommunications standards.

In the European Union, telecommunications cabling standards are developed by the European Committee for Electrotechnical Standardization (CENELEC). Many countries adopt the international telecommunications cabling standards developed jointly by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC).

These standards are consensus based and are developed by manufacturers, designers, and users. These standards are typically reviewed every 5 years, during which they are updated, reaffirmed, or withdrawn according to submissions by contributors. Standards organizations often publish addenda to provide new content or updates prior to publication of a complete revision to a standard.

14.3 DATA CENTER TELECOMMUNICATIONS CABLING INFRASTRUCTURE STANDARDS

Data center telecommunications cabling infrastructure standards by the TIA, CENELEC, and ISO/IEC cover the following subjects:

- Types of cabling permitted
- Cable and connecting hardware specifications
- Cable lengths
- Cabling system topologies
- Cabinet and rack specifications and placement
- Telecommunications space design requirements
- Telecommunications pathways (e.g., conduits and cable trays)
- Testing of installed cabling
- Telecommunications cabling system administration and labeling

The TIA data center standard is ANSI/TIA-942-A, *Telecommunications Infrastructure Standard for Data Centers*. The ANSI/TIA-942-A standard is the first revision of the ANSI/TIA-942 standard. This standard provides guidelines for the design and installation of a data center, including the facility's layout, cabling system, and supporting equipment. It also provides guidance regarding energy efficiency and provides a table with design guidelines for four levels of data center reliability.

ANSI/TIA-942-A references other TIA standards for content that is common with other telecommunications cabling standards. See Figure 14.3 for the organization of the TIA telecommunications cabling standards.

Thus, ANSI/TIA-942-A references each of the common standards:

- ANSI/TIA-568-C.0 for generic cabling requirements including cable installation and testing
- ANSI/TIA-569-C regarding pathways, spaces, cabinets, and racks
- ANSI/TIA-606-B regarding administration and labeling
- ANSI/TIA-607-B regarding bonding and grounding
- ANSI/TIA-758-B regarding campus/outside cabling and pathways
- ANSI/TIA-862-A regarding cabling for building automation systems including IP cameras, security systems, and monitoring systems for the data center electrical and mechanical infrastructure

Detailed specifications for the cabling are specified in the component standards ANSI/TIA-568-C.2, ANSI/TIA-568-C.3, and ANSI/TIA-568-C.4, but these standards are meant primarily for manufacturers. So the data center telecommunications cabling infrastructure designer in the United States or Canada should obtain ANSI/TIA-942-A and the common standards ANSI/TIA-568-C.0, ANSI/TIA-569-C, ANSI/TIA-606-B, ANSI/TIA-607-B, ANSI/TIA-758-B, and ANSI/TIA-862-A.

The CENELEC telecommunications standards for the European Union also have a set of common standards that apply to all types of premises and separate premises cabling standards for different types of buildings (Fig. 14.4).

A designer that intends to design telecommunications cabling for a data center in the European Union would need to obtain the CENELEC premises-specific standard for data centers—CENELEC EN 50173-5—and the common

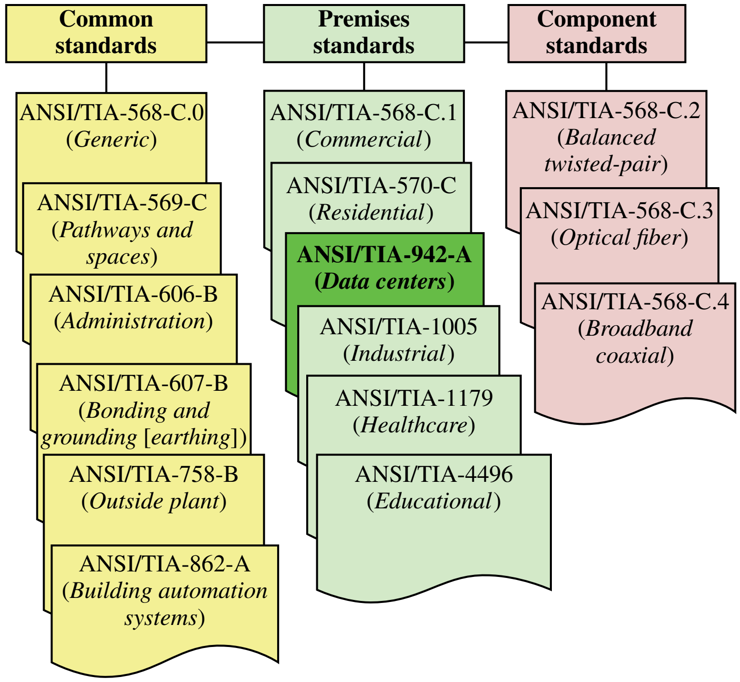


FIGURE 14.3 Organization of TIA telecommunications cabling standards.

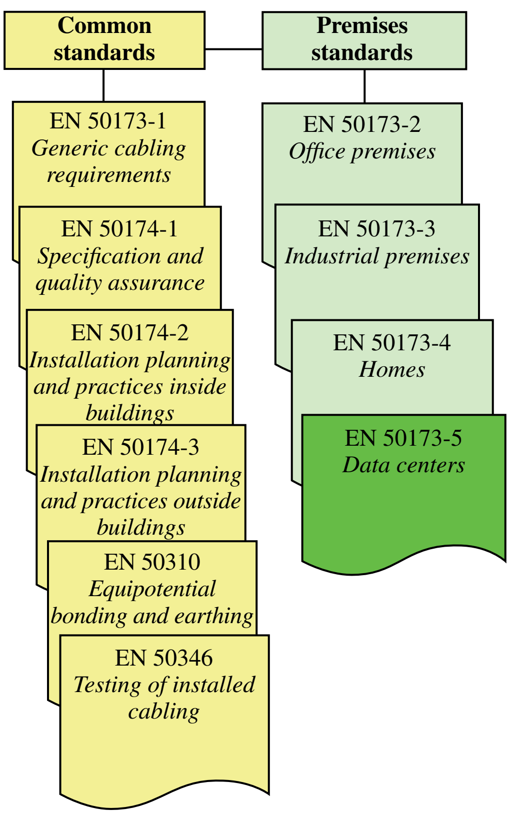


FIGURE 14.4 Organization of CENELEC telecommunications cabling standards.

standards CENELEC EN 50173-1, EN 50174-1, EN 50174-2, EN 50174-3, EN 50310, and EN 50346.

See Figure 14.5 for the organization of the ISO/IEC telecommunications cabling standards.

A designer that intends to design telecommunications cabling for a data center using the ISO/IEC standards would need to obtain the ISO/IEC premises–specific standard for data centers—ISO/IEC 24764—and the common standards ISO/IEC 11801-1, ISO/IEC 14763-2, and ISO/IEC 14763-3. The ISO/IEC premises-specific standards will be renumbered in the next revision to correspond more closely to the scheme used in the CENELEC cabling standards. ISO/IEC 24764 will be renamed ISO/IEC 11801-5 in the next revision.

The data center telecommunications cabling standards use the same topology for telecommunications cabling infrastructure but use different terminology. This handbook uses the terminology used in ANSI/TIA-942-A. See Table 14.1 for a cross-reference between the TIA, ISO, and CENELEC terminology.

ANSI/BICSI-002 Data Center Design and Implementation Best Practices Standard is another useful reference. It is an international standard meant to supplement the telecommunications cabling standard that applies in your country—ANSI/TIA-942-A, CENELEC EN 50173-5, ISO/IEC 24764, or other—and provides best practices beyond the minimum requirements specified in these other data center telecommunications cabling standards.

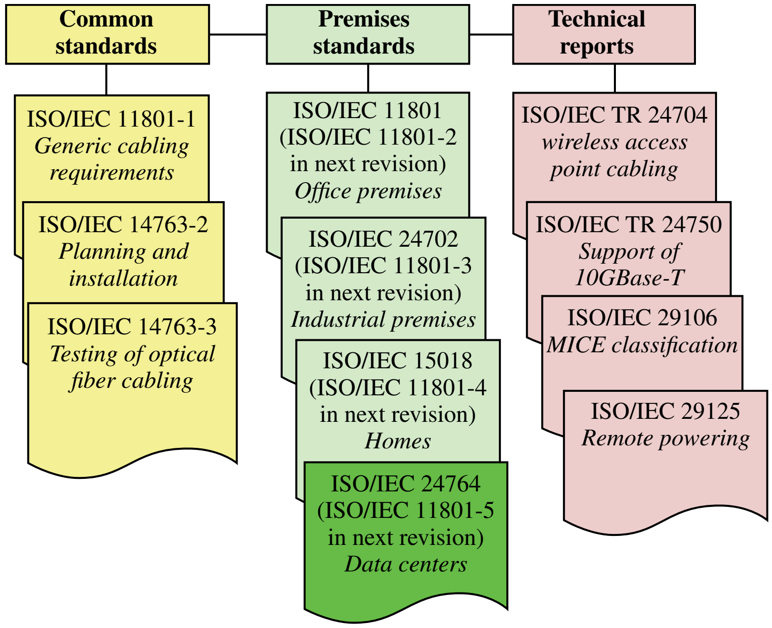


FIGURE 14.5 Organization of ISO/IEC telecommunications cabling standards.

TABLE 14.1 Cross-reference of TIA, ISO/IEC, and CENELEC terminology

ANSI/TIA-942-A	ISO/IEC 24764	CENELEC EN 50173-5
Telecommunications distributors		
Telecommunications entrance room (TER)	Not defined	Not defined
Main distribution area (MDA)	Not defined	Not defined
Intermediate distribution area (IDA)	Not defined	Not defined
Horizontal distribution area (HDA)	Not defined	Not defined
Zone distribution area (ZDA)	Not defined	Not defined
Equipment distribution area (EDA)	Not defined	Not defined
Cross-connects and distributors		
External network interface (ENI) in TER	ENI	ENI
Main cross-connect (MC) in the MDA	Main distributor (MD)	MD
Intermediate cross-connect (IC) in the IDA	Intermediate distributor (ID)	ID
Horizontal cross-connect (HC) in the HDA	Zone distributor (ZD)	ZD
Zone outlet or consolidation point in the ZDA	Local distribution point (LDP)	LDP
Equipment outlet (EO) in the EDA	EO	EO
Cabling subsystems		
Backbone cabling (from TER to MDAs, IDAs, and HDAs)	Network access cabling subsystems	Network access cabling subsystems
Backbone cabling (from MDA to IDAs and HDAs)	Main distribution cabling subsystems	Main distribution cabling subsystems
Backbone cabling (from IDAs to HDAs)	Intermediate distribution cabling subsystem	Intermediate distribution cabling subsystem
Horizontal cabling	Zone distribution cabling subsystem	Zone distribution cabling subsystem

14.4 TELECOMMUNICATIONS SPACES AND REQUIREMENTS

14.4.1 General Requirements

A computer room is an environmentally controlled room that serves the sole purpose of supporting equipment and cabling directly related to the computer and networking systems. The data center includes the computer room and all related support spaces dedicated to supporting the computer room such as the operation center, electrical rooms, mechanical rooms, staging area, and storage rooms.

The floor layout of the computer room should be consistent with the equipment requirements and the facility providers' requirements, including floor loading, service clearance, airflow, mounting, power, and equipment connectivity length requirements. Computer rooms should be located away from building components that would restrict future room expansion, such as elevators, exterior walls, building core, or immovable walls. They should also not have windows or skylights, as they allow light and heat into the computer room, making air conditioners work more and use more energy.

The rooms should be built with security doors that allow only authorized personnel to enter. It is also just as important that keys or pass codes to access the computer rooms are only accessible to authorized personnel. Preferably, the access control system should provide an audit trail.

The ceiling should be at least 2.6 m (8.5 ft) tall to accommodate cabinets up to 2.13 m (7 ft) tall. If taller cabinets are to be used, the ceiling height should be adjusted accordingly. There should also be a minimum clearance of 460 mm (18 in.) between the top of cabinets and sprinklers to allow them to function effectively.

Floors within the computer room should be able to withstand at least 7.2 kPa (150 lb/ft^2), but 12 kPa (250 lb/ft^2) is recommended. Ceilings should also have a minimum hanging capacity, so that loads may be suspended from them. The minimum hanging capacity should be at least 1.2 kPa (25 lb/ft^2), and a capacity of 2.4 kPa (50 lb/ft^2) is recommended.

The computer room needs to be climate controlled to minimize damage and maximize the life of computer parts. The room should have some protection from environmental contaminants like dust. Some common methods are to use vapor barriers, positive room pressure, or absolute filtration. Computer rooms do not need a dedicated HVAC system if it can be covered by the building's and has an automatic damper; however, having a dedicated HVAC system will improve reliability and is preferable if the building's might not be on continuously. If a computer room does have a dedicated HVAC system, it should be supported by the building's backup generator or batteries, if available.

A computer room should have its own separate power supply circuits with its own electrical panel. It should have duplex convenience outlets, for noncomputer use (e.g., cleaning equipment, power tools, and fans). The convenience outlets should be located every 3.65 m (12 ft) unless specified otherwise by local ordinances. These should be wired on separate power distribution units/panels from those used by the computers and should be reachable by a 4.5 m (15 ft) cord. If available, the outlets should be connected to a standby generator, but the generator must be rated for electronic loads or be "computer grade."

All computer room environments including the telecommunications spaces should be compatible with $M_1I_1C_1E_1$ environmental classifications per ANSI/TIA-568-C.0. MICE classifications specify environmental requirements as M, mechanical; I, ingress; C, climatic; and E, electromagnetic. Mechanical specifications include conditions such as vibration, bumping, impact, and crush. Ingress specifications include conditions such as particulates and water immersion. Climatic specifications include temperature, humidity, liquid contaminants, and gaseous contaminants. Electromagnetic specifications include electrostatic discharge (ESD), radio-frequency emissions, magnetic fields, and surge. The CENELEC and ISO/IEC standards also have their own similar MICE specifications.

Temperature and humidity for computer room spaces should follow current ASHRAE TC 9.9 and manufacturer equipment guidelines.

The telecommunications spaces such as the main distribution area (MDA), intermediate distribution area (IDA), and HDA could be separate rooms within the data center but are more often a set of cabinets and racks within the computer room space.

14.4.2 Telecommunications Entrance Room

The telecommunications entrance room (TER) or entrance room refers to the location where telecommunications cabling enters the building and not the location where people enter the building. This is typically the demarcation point—the location where telecommunications access providers hand off circuits to customers. The TER is also the location where the owner's outside plant cable (such as campus cabling) terminates inside the building.

The TER houses entrance pathways, protector blocks for twisted-pair entrance cables, termination equipment for access provider cables, access provider equipment, and termination equipment for cabling to the computer room.

The interface between the data center structured cabling system and external cabling is called the external network interface (ENI).

The telecommunications access provider's equipment is housed in this room, so the provider's technicians will need access. Because of this, it is not recommended to put the entrance room inside a computer room and that it is housed within a separate room, such that access to it does not compromise the security of any other room requiring clearance.

The room's location should also be determined so that the entire circuit length from the demarcation point does not exceed the maximum specified length. If the data center is very large:

- The TER may need to be located in the computer room space.
- The data center may need multiple entrance rooms.

The location of the TER should also not interrupt air flow, piping, or cabling under floor.

The TER should be adequately bonded and grounded (for primary protectors, secondary protectors, equipment, cabinets, racks, metallic pathways, and metallic components of entrance cables).

The cable pathway system should be the same type as the one used in the computer room. Thus, if the computer room uses overhead cable tray, the TER should use overhead cable tray as well.

There may be more than one entrance room for large data centers, additional redundancy, or dedicated service feeds. If the computer rooms have redundant power and cooling, TER power and cooling should be redundant to the same degree.

There should be a means of removing water from the entrance room if there is a risk. Water pipes should also not run above equipment.

14.4.3 MDA

The MDA is the location of the main cross-connect (MC), the central point of distribution for the structured cabling system. Equipment such as core routers and switches may be located here. The MDA may also contain a horizontal cross-connect (HC) to support horizontal cabling for nearby cabinets. If there is no dedicated entrance room, the MDA may also function as the TER. In a small data center, the MDA may be the only telecommunications space in the data center.

The location of the MDA should be chosen such that the cable lengths do not exceed the maximum length restrictions.

If the computer room is used by more than one organization, the MDA should be in a separate secured space (e.g., a secured room, cage, or locked cabinets). If it has its own room, it may have its own dedicated HVAC system and power panels connected to backup power sources.

There may be more than one MDA for redundancy.

Main distribution frame (MDF) is a common industry term for the MDA.

14.4.4 IDA

The IDA is the location of an intermediate cross-connect (IC)—an optional intermediate-level distribution point within the structured cabling system. The IDA is not vital and may be absent in data centers that do not require three levels of distributors.

If the computer room is used by multiple organizations, it should be located in a separate secure space—for example, a secured room, cage, or locked cabinets.

The IDA should be located centrally to the area that it serves to avoid exceeding the maximum cable length restrictions.

This space also typically houses switches (LAN, SAN, management, console).

The IDA may contain an HC to support horizontal cabling to cabinets near the IDA.

14.4.5 HDA

The HDA is a space that contains an HC, the termination point for horizontal cabling to the equipment cabinets and racks (equipment distribution areas (EDAs)). This space typically also houses switches (LAN, SAN, management, console).

If the computer room is used by multiple organizations, it should be located in a separate secure space—for example, a secured room, cage, or locked cabinets

There should be a minimum of one HC per floor, which may be in an HDA, IDA, or MDA.

The HDA should be located to avoid exceeding the maximum backbone length from the MDA or IDA for the medium of choice. If it is located in its own room, it is possible for it to have its own dedicated HVAC or electrical panels.

To provide redundancy, equipment cabinets and racks may have horizontal cabling to two different HDAs.

Intermediate distribution frame (IDF) is a common industry term for the HDA.

14.4.6 Zone Distribution Area

The zone distribution area (ZDA) is the location of either a consolidation point or equipment outlets (EOs). A consolidation point is an intermediate administration point for horizontal cabling. Each ZDA should be limited to 288 coaxial cable or balanced twisted-pair cable connections to avoid cable congestion. The two ways that a ZDA can be deployed—as a consolidation point or as a multiple outlet assembly—are illustrated in Figure 14.6.

The ZDA shall contain no active equipment, nor should it be a cross-connect (i.e., have separate patch panels for cables from the HDAs and EDAs).

ZDAs may be located in underfloor enclosures, overhead enclosures, cabinets, or racks.

14.4.7 EDA

The EDA is the location of end equipment, which comprises the computer systems, communications equipment, and their racks and cabinets. Here, the horizontal cables are

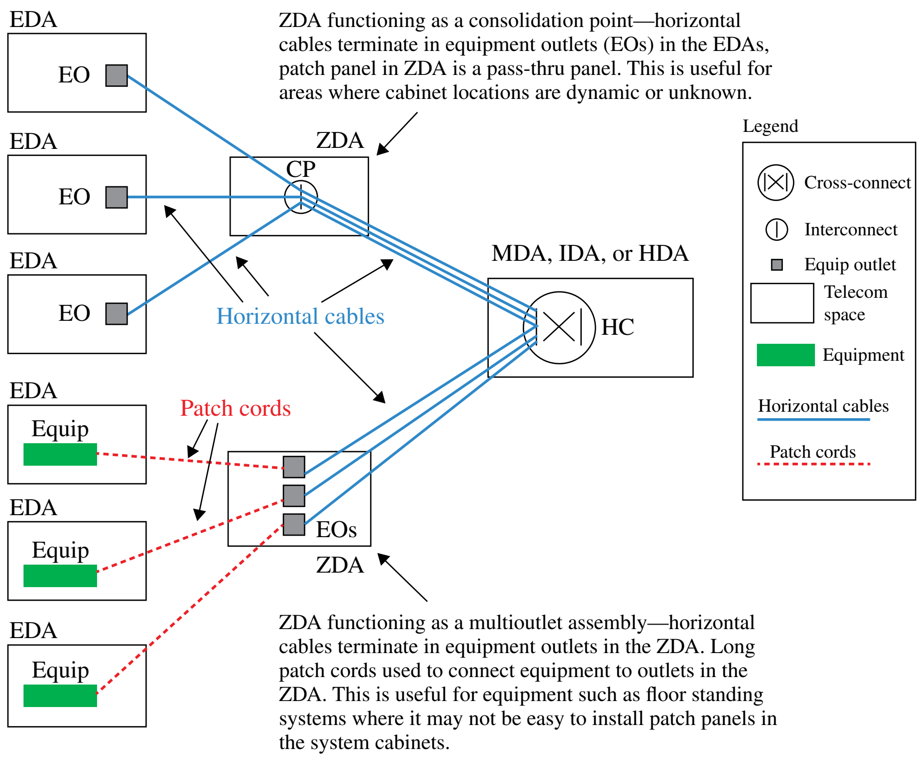


FIGURE 14.6 Two examples of ZDAs.

terminated in EOs. Typically, an EDA has multiple EOs for terminating multiple horizontal cables. These EOs are typically located in patch panels located at the rear of the cabinet or rack (where the connections for the servers are usually located).

Point-to-point cabling (i.e., direct cabling between equipment) may be used between equipment located in EDAs. Point-to-point cabling should be limited to 10 m (33 ft) in length and should be within a row of cabinets or racks. Permanent labels should be used on either end of each cable.

14.4.8 Telecommunications Room

The telecommunications room (TR) is an area that supports cabling to areas outside of the computer room, such as operations staff support offices, security office, operation center, electrical room, mechanical room, or staging area. They are usually located outside of the computer room but may be combined with an MDA, IDA, or HDA.

14.4.9 Support Area Cabling

Cabling for support areas of the data center outside the computer room is typically supported from one or more dedicated TRs to improve security. This allows technicians working on telecommunications cabling, servers, or network hardware for these spaces to remain outside the computer room.

Operation rooms and security rooms typically require more cables than other work areas. Electrical rooms, mechanical rooms, storage rooms, equipment staging rooms, and loading docks should have at least one wall-mounted phone in each room for communication within the facility. Electrical and mechanical rooms need at least one data connection for management system access and may need more connections for equipment monitoring.

14.5 STRUCTURED CABLING TOPOLOGY

The structured cabling system topology described in data center telecommunications cabling standards is a hierarchical star (Fig. 14.7).

The horizontal cabling is the cabling from the HCs to the EDAs and ZDAs. This is the cabling that supports end equipment such as servers.

The backbone cabling is the cabling between the distributors where cross-connects are located—TERs, TRs, MDAs, IDAs, and HDAs.

Cross-connects are patch panels that allow cables to be connected to each other using patch cords. For example, the HC allows backbone cables to be patched to horizontal cables. An interconnect, such as a consolidation point in a ZDA, connects two cables directly through the patch panel. See Figure 14.8 for examples of cross-connects and interconnects used in data centers.

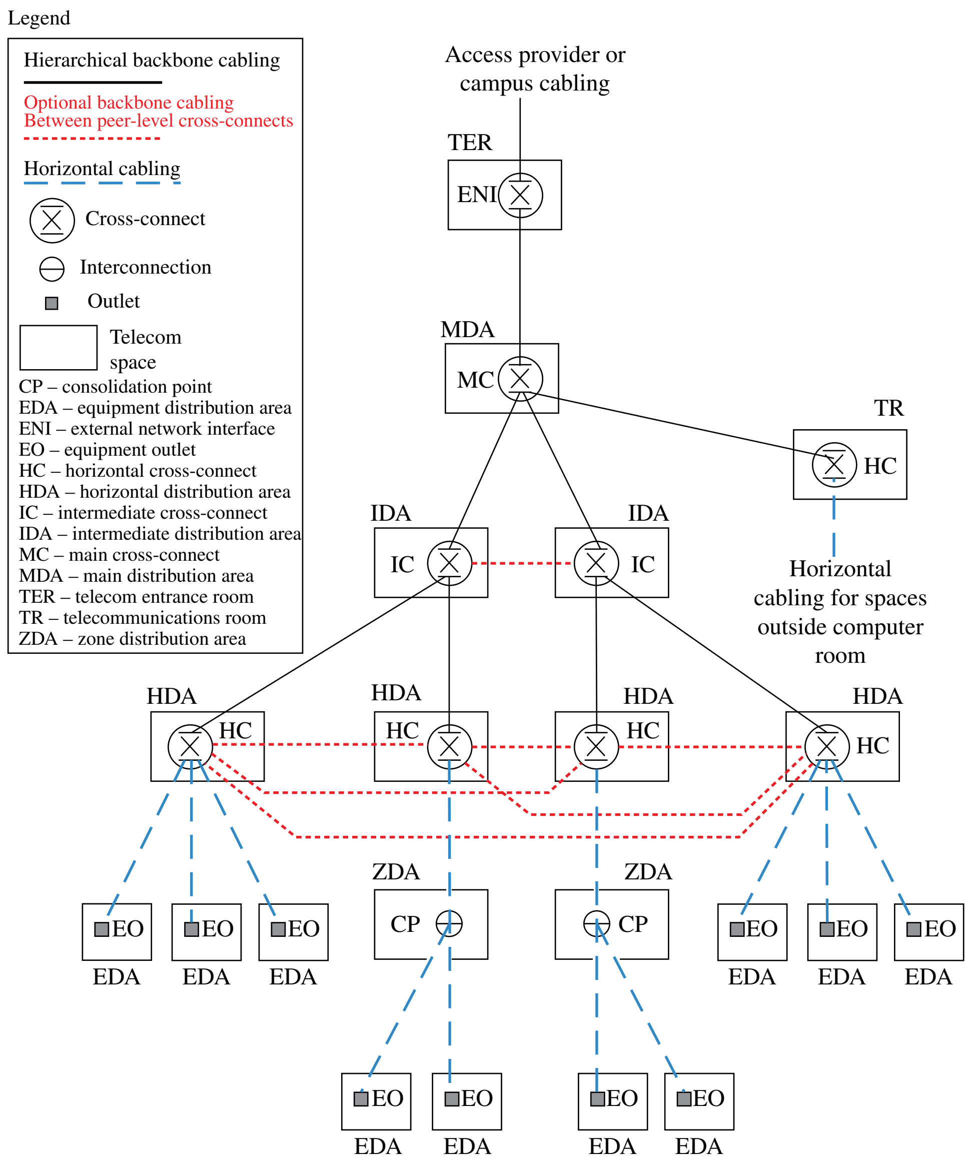


FIGURE 14.7 Hierarchical star topology.

Note that switches can be patched to horizontal cabling using either a cross-connect or interconnect scheme (see the two diagrams on the right side of Fig. 14.8). The interconnect scheme avoids another patch panel; however, the cross-connect scheme may allow more compact cross-connects since the switches don't need to be located in or adjacent to the cabinets containing the HCs.

Most of the components of the hierarchical star topology are optional. However, each cross-connect must have backbone cabling to a higher-level cross-connect:

- ENIs must have backbone cabling to an MC. They may also have backbone cabling to an IC or HC as required to ensure that WAN circuit lengths are not exceeded.
- HCs in TRs located in a data center must have backbone cabling to an MC and may optionally have backbone cabling to other distributors (ICs, HCs).
- ICs must have backbone cabling to an MC and one or more HCs. They may optionally have backbone cabling to an ENI or IC either for redundancy or to ensure that maximum cable lengths are not exceeded.
- HCs in an HDA must have backbone cabling to an MC or IC. They may optionally have backbone cabling to an HC, ENI, or IC either for redundancy or to ensure that maximum cable lengths are not exceeded.
- Because ZDAs only support horizontal cabling, they may only have cabling to an HDA and EDA.

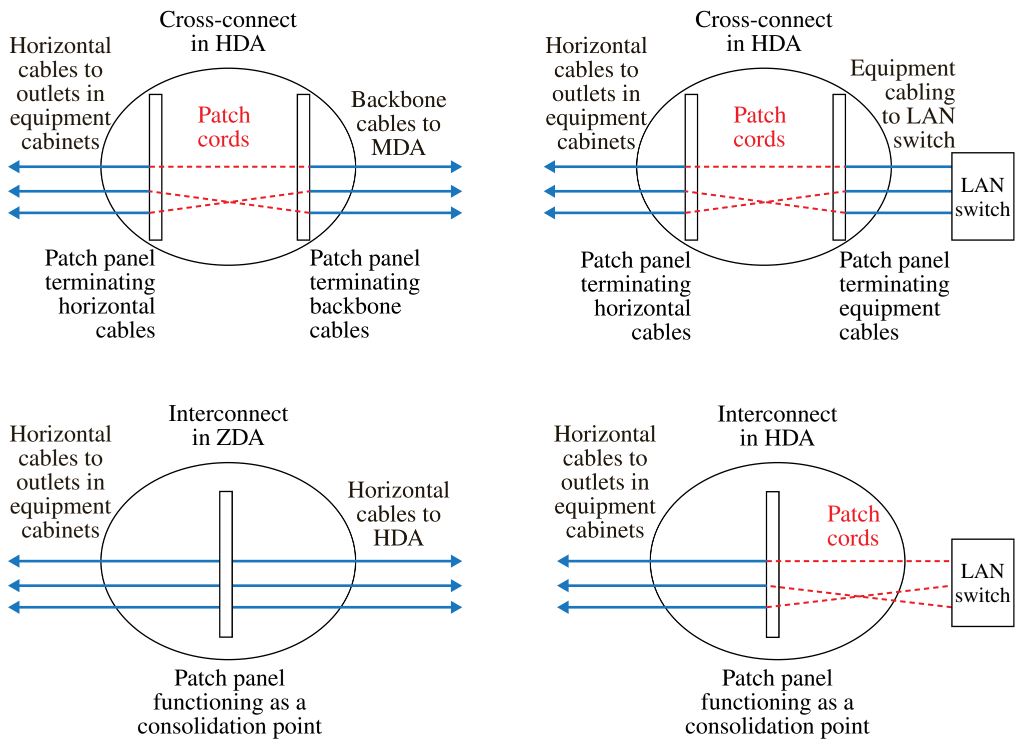


FIGURE 14.8 Cross-connects and interconnect examples.

Cross-connects such as the MC, IC, and HC should not be confused with the telecommunications spaces in which they are located, that is, the MDA, IDA, and HDA. The cross-connects are components of the structured cabling system and typically comprise patch panels. The spaces are dedicated rooms or more commonly dedicated cabinets, racks, or cages within the computer room.

EDAs and ZDAs may have cabling to different HCs to provide redundancy. Similarly, HCs, ICs, and ENIs may have redundant backbone cabling. The redundant backbone cabling may be to different spaces (for maximum redundancy) or between the same to spaces on both ends but follow different routes. See Figure 14.9 for degrees of redundancy in the structured cabling topology at various levels as defined in ANSI/TIA-942-A.

A level 1 cabling infrastructure has no redundancy.

A level 2 cabling infrastructure requires redundant access provider (telecommunications carrier) routes into the data center. The two redundant routes must go to different carrier central offices and be separated from each other along their entire route by at least 20 m (66 ft).

A level 3 cabling infrastructure has redundant TERs. The data center must be served by two different access providers (carriers). The redundant routes that the circuits take from the two different carrier central offices to the data center must be separated by at least 20 m (66 ft).

A level 3 data center also requires redundant backbone cabling. The backbone cabling between any two cross-connects must use at least two separate cables, preferably following different routes within the data center.

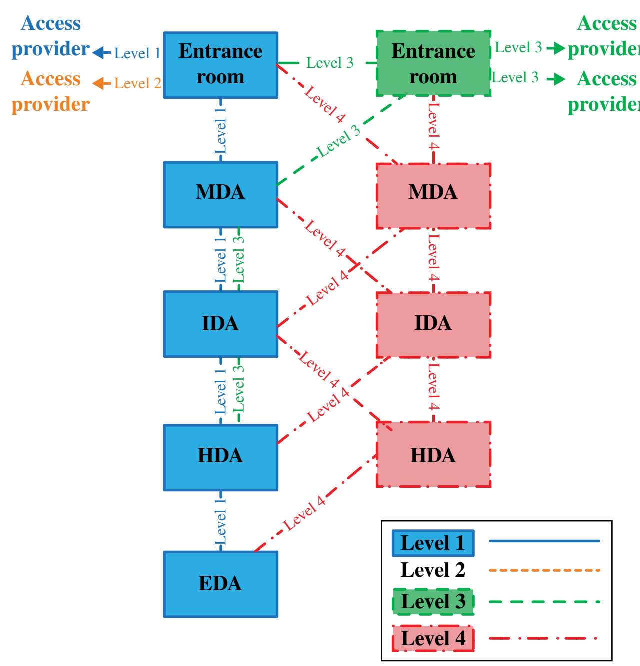


FIGURE 14.9 Structured cabling redundancy at various levels.

A level 4 data center adds redundant MDAs, IDAs, and HDAs. Equipment cabinets and racks (EDAs) must have horizontal cabling to two different HDAs. HDAs must have redundant backbone cabling to two different IDAs (if present) or MDAs. Each entrance room must have backbone cabling to two different MDAs.

14.6 CABLE TYPES AND MAXIMUM CABLE LENGTHS

There are several types of cables one can use for telecommunications cabling in data centers. Each has different characteristics, and they are chosen to suit the various conditions to which they are subject. Some cables are more flexible than others. The size of the cable can affect its flexibility as well as its shield. A specific type of cable may be chosen because of space constraints or required load or because of bandwidth or channel capacity. Equipment vendors may also recommend cable for use with their equipment.

14.6.1 Coaxial Cabling

Coaxial cables are composed of a center conductor, surrounded by an insulator, surrounded by a metallic shield, and covered in a jacket. The most common types of coaxial cable used in data centers are the 75 ohm 734- and 735-type cables used to carry E-1, T-3, and E-3 wide area circuits; see Telcordia Technologies GR-139-CORE regarding specifications for 734-and 735-type cables and ANSI/ATIS-0600404.2002 for specifications regarding 75 ohm coaxial connectors.

Circuit lengths are longer for the thicker, less flexible 734 cable. These maximum cable lengths are decreased by intermediate connectors and DSX panels—see ANSI/TIA-942-A.

Broadband coaxial cable is also sometimes used in data centers to distribute television signals. The specifications of the broadband coaxial cables (Series 6 and Series 11) and connectors (F-type) are specified in ANSI/TIA-568-C.4.

14.6.2 Balanced Twisted-Pair Cabling

The 100 ohm balanced twisted-pair cable is a type of cable that uses multiple pairs of copper conductors. Each pair of conductors is twisted together to protect the cables from electromagnetic interference.

- Unshielded twisted-pair (UTP) cables have no shield.
- The cable may have an overall cable screen made of either foil or braided shield or both.
- Each twisted pair may also have a foil shield.

Balanced twisted-pair cables come in different categories or classes based on the performance specifications of the cables (Table 14.2).

Category 3, 5e, 6, and 6A cables are typically UTP cables but may have an overall screen or shield.

Category 7 and 7A cables have an overall shield and a shield around each of the four twisted pairs.

Balanced twisted-pair cables used for horizontal cabling have four pairs. Balanced twisted-pair cables used for backbone cabling may have four or more pairs. The pair count above 4 pairs is typically a multiple of 25 pairs.

Types of balanced twisted-pair cables required and recommended in standards are as specified in Table 14.3.

Note that TIA-942-A recommends and ISO/IEC 24764 requires a minimum of Category 6A balanced twisted-pair cabling to support 10G Ethernet. Category 6 cabling may support 10G Ethernet for shorter distances (<55 m), but it may require limiting the number of cables that support

TABLE 14.2 Balanced twisted-pair categories

TIA categories	ISO/IEC and CENELEC classes/categories	Max frequency (MHz)	Common application
Category 3	N/A	16	Voice, wide area network circuits, serial console, 10 Mbps Ethernet
Category 5e	Class D/Category 5	100	Same as Category 3 + 100 Mbps and 1 Gbps Ethernet
Category 6	Class E/Category 6	250	Same as Category 5e
Augmented category 6 (cat 6A)	Class E_A/Category 6A	500	Same as Category 5e + 10G Ethernet
N/A	Class F/Category 7	600	Same as Category 6A
N/A	Class F_A/Category 7A	1000	Same as Category 6A

ISO/IEC and CENELEC categories refer to components such as cables and connectors. Classes refer to channels comprising installed cabling including cables and connectors.

Note that the TIA does not currently specify cabling categories above Category 6A. However, higher-performance Category 7/Class F and Category 7A/Class FA are specified in ISO/IEC and CENELEC cabling standards.

Category 3 is no longer supported in ISO/IEC and CENELEC cabling standards.

TABLE 14.3 Balanced twisted-pair requirements in standards

Standard	Type of cabling	Balanced twisted-pair cable categories/classes permitted
TIA-942-A	Horizontal cabling	Category 6 or 6A, Category 6A recommended
TIA-942-A	Backbone cabling	Category 3, 5e, 6 or 6A, Category 6A recommended
ISO/IEC 24764	All cabling except network access cabling	Category 6A/E_A, 7/F, 7A/F_A
ISO/IEC 24764	Network access cabling (to/from telecom entrance room/ENI)	Category 5/Class D, 6/E, 6A/E_A, 7/F, 7A/F_A
CENELEC EN 51073-5	All cabling except network access cabling	Category 6/Class F, 6A/E_A, 7/F, 7A/F_A
CENELEC EN 51073-5	Network access cabling (to/from telecom entrance room/ENI)	Category 5/Class D, 6/E, 6A/E_A, 7/F, 7A/F_A

10G Ethernet and other mitigation measures to function properly; see TIA TSB-155-A, *Guidelines for the Assessment and Mitigation of Installed Category 6 to Support 10GBase-T*.

The TIA is developing specifications for Category 8 and ISO/IEC is developing specifications for Categories 8.1 and 8.2 to support a future 40 Gbps Ethernet specification that will use balanced twisted-pair cabling up to a distance of 30 m.

TABLE 14.4 E-1, T-3, and E-3 circuits' lengths over coaxial cable

Circuit type	734 cable	735 cable
E-1	332 m (1088 ft)	148 m (487 ft)
T-3	146 m (480 ft)	75 m (246 ft)
E-3	160 m (524 ft)	82 m (268 ft)

14.6.3 Optical Fiber Cabling

Optical fiber comprises a thin transparent filament, typically glass, surrounded by a cladding, which is used as a waveguide. Both single-mode and multimode fibers can be used over long distances and have high bandwidth. Single-mode fiber uses a thinner core, which allows only one mode (or path) of light to propagate. Multimode fiber uses a wider core, which allows multiple modes (or paths) of light to propagate. Multimode fiber uses less expensive transmitters and receivers but has less bandwidth than single-mode fiber. The bandwidth of multimode fiber reduces over distance, because light following different modes will arrive at the far end at different times.

There are four classifications of multimode fiber: OM1, OM2, OM3, and OM4. OM1 is a 62.5/125 µm multimode optical fiber. OM2 can be either a 50/125 µm or 62.5/125 µm multimode optical fiber. OM3 and OM4 are both 50/125 µm 850 nm laser-optimized multimode fibers, but OM4 optical fiber has higher bandwidth.

A minimum of OM3 is specified in data center standards. TIA-942-A recommends the use of OM4 multimode optical fiber cable to support longer distances for 100G Ethernet.

There are two classifications of single-mode fiber: OS1 and OS2. OS1 is a standard single-mode fiber. OS2 is a low water peak single-mode fiber that has processed to reduce attenuation at 1400 nm frequencies allowing those frequencies to be used. Either type of single-mode optical fiber may be used in data centers.

14.6.4 Maximum Cable Lengths

The following are the maximum circuit lengths over 734- and 735-type coaxial cables with only two connectors (one at each end) and no DSX panel (Table 14.4).

Generally, the maximum length for LAN applications that are supported by balanced twisted-pair cables is 100 m (328 ft), with 90 m being the maximum length permanent link between patch panels and 10 m allocated for patch cords.

Channel lengths (lengths including permanently installed cabling and patch cords) for common data center LAN applications over multimode optical fiber are shown in Table 14.5. Channel lengths for single-mode optical fiber are several kilometers since single-mode fiber is used for long-haul communications.

IEEE is developing a lower-cost four-lane (eight-fiber) implementation of 100G Ethernet. The channel lengths for four-lane 100G Ethernet are expected to be 70 m over OM3 and 100 m over OM4.

Refer to ANSI/TIA-568-C.0 and ISO 11801 for tables that provide more details regarding maximum cable lengths for other applications.

TABLE 14.5 Ethernet channel lengths over multimode optical fiber

Fiber Type	1G Ethernet	10G Ethernet	40G Ethernet	100G Ethernet
# OF FIBERS	2	2	8	20
OM1	275 m	26 m	Not supported	Not supported
OM2	550 m	82 m	Not supported	Not supported
OM3	[a]800 m	300 m	100 m	100 m
OM4	[a]1040 m	[a]550 m	150 m	150 m

[a]Distances specified by manufacturers but not in IEEE standards.

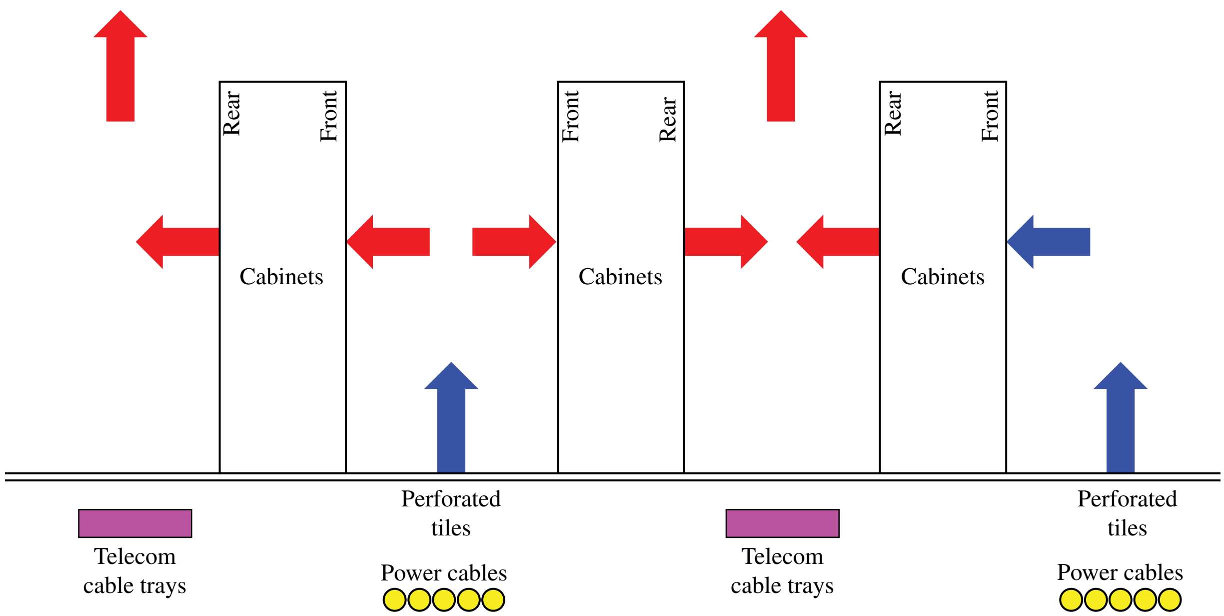


FIGURE 14.10 Hot and cold aisle examples.

14.7 CABINET AND RACK PLACEMENT (HOT AISLES AND COLD AISLES)

It is important to keep computers cool; computers create heat during operation, and heat decreases their functional life and processing speed, which in turn uses more energy and increases cost. The placement of computer cabinets or racks affects the effectiveness of a cooling system. Airflow blockages can prevent cool air from reaching computer parts and can allow heat to build up in poorly cooled areas.

One efficient method of placing cabinets is using hot and cold aisles, which creates convection currents that helps circulate air (Fig. 14.10). This is achieved by placing cabinets in rows with aisles between each row. Cabinets in each row are oriented such that they face one another. The hot aisles are the walkways with the rears of the cabinets on either side, and cold aisles are the walkways with the front of the cabinets on either side.

If telecommunications cables are placed under access floors, telecommunications cables should be placed under the hot aisles so as to not restrict airflow if underfloor cooling ventilation is to be used. If power cabling is distributed under the access floors, the power cables should be placed on the floor in the cold aisles to ensure proper separation of power and telecommunications cabling (Fig. 14.10).

Lighting and telecommunications cabling shall be separated by at least 5 in.

Power and telecommunications cabling shall be separated by the distances specified in ANSI/TIA-569-C or ISO/IEC 14763-2. Generally, it is best to separate large numbers of power cables and telecommunications cabling by at least 600 mm (2 ft). This distance can be halved if the power cables are completely surrounded by a grounded metallic shield or sheath.

The minimum clearance at the front of the cabinets and racks is 1.2 m (4 ft), the equivalent of two full tiles. This ensures that there is proper clearance at the front of the cabinets to install equipment into the cabinets—equipment is typically installed in cabinets from the front. The minimum clearance at the rear of cabinets and equipment at the rear of racks is 900 mm (3 ft). This provides working clearance at the

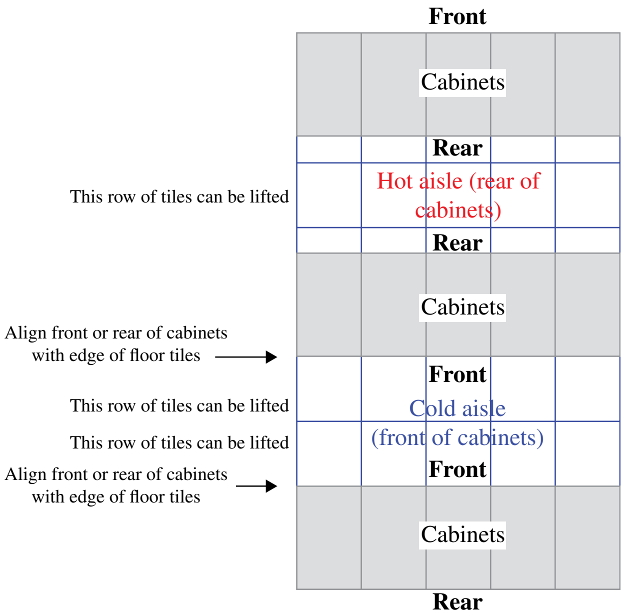


FIGURE 14.11 Cabinet placement example.

rear of the equipment for technicians to work on equipment. If cool air is provided from ventilated tiles at the front of the cabinets, more than 1.2 m (4 ft) of clearance may be specified by the mechanical engineer to provide adequate cool air.

The cabinets should be placed such that either the front or rear edges of the cabinets align with the floor tiles. This ensures that the floor tiles at both rears of the cabinets can be lifted to access systems below the access floor (Fig. 14.11).

If power and telecommunications cabling are under the access floor, the direction of airflow from air-conditioning equipment should be parallel to the rows of cabinets and racks to minimum interference caused by the cabling and cable trays.

Openings in the floor tiles should only be made for cooling vents or for routing cables through the tile. Openings for floor tile for cables should minimize air pressure loss by not cutting excessively large holes and by using a device that restricts airflow around cables, like brushes or flaps. The holes for cable management should not create tripping hazards; ideally, they should be located either under the cabinets or under vertical cable managers between racks.

If there are no access floors or if they are not to be used for cable distribution, cable trays shall be routed above cabinets and racks and not above the aisles.

Sprinklers and lighting should be located above aisles rather than above cabinets, racks, and cable trays, where their efficiency will be significantly reduced.

14.8 CABLING AND ENERGY EFFICIENCY

There should be no windows in the computer room; it allows light and heat into the environmentally controlled area, which creates an additional heat load.

TIA-942-A specifies that the 2011 ASHRAE TC 9.9 guidelines be used for the temperature and humidity in the computer room and telecommunications spaces.

ESD could be a problem at low humidity (dew point below 15°C [59°F], which corresponds approximately to 44% relative humidity at 18°C [64°F] and 25% relative humidity at 27°C [81°F]). Follow the guidelines in TIA TSB-153, *Static Discharge between LAN Cabling and Data Terminal Equipment*, for mitigation of ESD if the data center will operate in low humidity for extended periods. The guidelines include use of grounding patch cords to dissipate ESD built up on cables and use wrist straps per manufacturers' guidelines when working with equipment.

The attenuation of balanced twisted-pair telecommunications cabling will increase as temperatures increase. Since the ASHRAE guidelines permit temperatures measured at inlets to be as high as 35°C (95°F), temperatures in the hot aisles where cabling may be located can be as high as 55°C (131°F). See ISO/IEC 11801, CENELEC EN 50173-1, or ANSI/TIA-568-C.2 for reduction in maximum cable lengths based on the average temperature along the length of the cable. Cable lengths may be further decreased if the cables are used to power equipment, since the cables themselves will also generate heat.

TIA-942-A recommends that energy-efficient lighting such as LED be used in the data center and that the data center follow a three-level lighting protocol depending on human occupancy of each space:

- Level 1—with no occupants, the lighting level should only be bright enough to meet the needs of the security cameras.
- Level 2—detection of motion triggers higher lighting levels to provide safe passage through the space and to permit security cameras to identify persons.
- Level 3—this level is used for areas occupied for work; these areas shall be lit to 500 lux.

Cooling can be affected both positively and negatively by the telecommunications and IT infrastructure. For example, the use of the hot aisle/cold aisle cabinet arrangement described earlier will enhance cooling efficiency. Cable pathways should be designed and located so as to minimize interference with cooling.

Generally, overhead cabling is more energy efficient than underfloor cabling if the space under the access floor is used for cooling since overhead cables will not restrict airflow or clause turbulence.

If overhead cabling is used, the ceilings should be high enough so that air can circulate freely around the hanging devices. Ladders or trays should be stacked in layers in high-capacity areas so that cables are more manageable and do not block the air. If present, optical fiber patch cords should be protected from copper cables.

If underfloor cabling is used, they will be hidden from view, which will give a cleaner appearance. Installation is generally easier. Care should be taken to separate telecommunications cables from the underfloor electrical wiring. Smaller cable diameters should be used. Shallower, wider cable trays are preferred as they don't obstruct underfloor airflow as much. Additionally, if underfloor air-conditioning is used, cables from cabinets should run in the same direction of airflow to minimize air pressure attenuation.

Either overhead or underfloor cable trays should be no deeper than 6 in. (150 mm). Cable trays used for optical fiber patch cords should have solid bottoms to prevent microbends in the optical fibers.

Enclosure or enclosure systems can also assist with air-conditioning efficiency. Consider using systems such as the following:

- Cabinets with isolated air returns (e.g., chimney to plenum ceiling space) or isolated air supply.
- Cabinets with in-cabinet cooling systems (e.g., door cooling systems).
- Hot aisle containment or cold aisle containment systems—note that cold aisle containment systems will generally mean that most of the space including the space occupied by overhead cable trays will be warm.
- Cabinets that minimize air bypass between the equipment rails and the side of the cabinet.

The cable pathways, cabinets, and racks should minimize the mixing of hot and cold air where not intended. Openings in cabinets, access floors, and containment systems should have brushes, grommets, and flaps at cable openings to decrease air loss around cable holes.

The equipment should match the cooling scheme—that is, equipment should generally have air intakes at the front and exhaust hot air out the rear. If the equipment does not match this scheme, the equipment may need to be installed backward (for equipment that circulates air back to front) or the cabinet may need baffles (for equipment that has air intakes and exhausts at the sides).

Data center equipment should be inventoried. Unused equipment should be removed (to avoid powering and cooling unnecessary equipment).

Cabinets and racks should have blanking panels at unused spaces to avoid mixing of hot and cold air.

Unused areas of the computer room should not be cooled. Compartmentalization and modular design should be taken into consideration when designing the floor plans; adjustable room dividers and multiple rooms with dedicated HVACs allow only the used portions of the building to be cooled and unoccupied rooms to be inactive.

Also, consider building the data center in phases. Sections of the data center that are not fully built require less capital and operating expenses. Additionally, since future needs may be difficult to predict, deferring construction of unneeded data center space reduces risk.

14.9 CABLE PATHWAYS

Adequate space must be allocated for cable pathways. In some cases, either the length of the cabling (and cabling pathways) or the available space for cable pathways could limit the layout of the computer room.

Cable pathway lengths must be designed to avoid exceeding maximum cable lengths for WAN circuits, LAN connections, and SAN connections:

- Length restrictions for WAN circuits can be avoided by careful placement of the entrance rooms, demarcation equipment, and wide area networking equipment to which circuits terminate. In some cases, large data centers may require multiple entrance rooms.
- Length restrictions for LAN and SAN connections can be avoided by carefully planning the number and location of MDAs, IDAs, and HDAs where the switches are commonly located.

There must be adequate space between stacked cable trays to provide access for installation and removal of cables. TIA and BICSI standards specify a separation of 12 in. (300 mm) between the top of one tray and the bottom of the tray above it. This separation requirement does not apply to cable trays run at right angles to each other.

Where there are multiple tiers of cable trays, the depth of the access floor or ceiling height could limit the number of cable trays that can be placed.

Standards and the NFPA National Electrical Code limit the maximum depth of cable and cable fill of cable trays:

- Cabling inside cable trays must not exceed a depth of 150 mm (6 in.) regardless of the depth of the tray.
- With cable trays that do not have solid bottom, the maximum fill of the cable trays is 50% by cross-sectional area of the cables.
- With cable trays that have solid bottoms, the maximum fill of the cable trays is 40%.

Cables in underfloor pathways should have a clearance of at least 50 mm (2 in.) from the bottom of the floor tiles to the top of the cable trays to provide adequate space between the cable trays and the floor tiles to route cables and avoid damage to cables when floor tiles are placed.

Optical fiber patch cords should be placed in cable trays with solid bottoms to avoid attenuation of signals caused by microbends.

Optical fiber patch cords should be separated from other cables to prevent the weight of other cables from damaging the fiber patch cords.

When they are located below the access floors, cable trays should be located in the cold aisles. When they are located overhead, they should be located above the cabinets and racks. Lights and sprinklers should be located above the aisles rather than the cable trays and cabinets/racks.

Cabling shall be at least 5 in. (130 mm) from lighting and adequately separated from power cabling as previously specified.

14.10 CABINETS AND RACKS

Racks are frames with side mounting rails on which equipment may be fastened. Cabinets have adjustable mounting rails, panels, and doors and may have locks. Because cabinets are enclosed, they may require additional cooling if natural airflow is inadequate; this may include using fans for forced airflow, minimizing return air flow obstructions, or liquid cooling.

Empty cabinet and rack positions should be avoided. Cabinets that have been removed should be replaced and gaps should be filled with new cabinets/racks with panels to avoid recirculation of hot air.

If doors are installed in cabinets, there should be at least 63% open space on the front and rear doors to allow for adequate airflow. Exceptions may be made for cabinets with fans or other cooling mechanisms (such as dedicated air returns or liquid cooling) that ensure that the equipment is adequately cooled.

In order to avoid difficulties with installation and future growth, consideration should be taken when designing and installing the preliminary equipment. 480 mm (19 in.) racks should be used for patch panels in the MDAs, IDAs, and HDAs, but 585 mm (23 in.) racks may be required by the service provider in the entrance room. Both racks and cabinets should not exceed 2.4 m (8 ft) in height.

Except for cable trays/ladders for patching between racks within the MDA, IDA, or HDA, it is not desirable to secure cable ladders to the top of cabinets and racks as it may limit the ability to replace the cabinets and racks in the future.

To ensure that infrastructure is adequate for unexpected growth, vertical cable management size should be calculated by the maximum projected fill plus a minimum of 50% growth.

The cabinets should be at least 150 mm (6 in.) deeper than the deepest equipment to be installed.

14.11 PATCH PANELS AND CABLE MANAGEMENT

Organization becomes increasingly difficult as more interconnecting cables are added to equipment. Labeling both cables and patch panels can save time, as accidentally switching or removing the wrong cable can cause outages that can take an indefinite amount of time to locate and correct. The simplest and most reliable method of avoiding patching errors is by clearly labeling each patch panel and each end of every cable as specified in ANSI/TIA-606-B.

However, this may be difficult if high-density patch panels are used. It is not generally considered a good practice to use patch panels that have such high density that they cannot be properly labeled.

Horizontal cable management panels should be installed above and below each patch panel; preferably, there should be a one-to-one ratio of horizontal cable management to patch panel unless angled patch panels are used. If angled patch panels are used instead of horizontal cable managers, vertical cable managers should be sized appropriately to store cable slack.

Separate vertical cable managers are typically required with racks unless they are integrated into the rack. These vertical cable managers should provide both front and rear cable managements.

Patch panels should not be installed on the front and back of a rack or cabinet to save space, unless both sides can be easily accessed from the front.

14.12 RELIABILITY LEVELS AND CABLING

Data center infrastructure levels have four categories: telecommunications (T), electrical (E), architectural (A), and mechanical (M). Each category is rated from one to four with one providing the lowest availability and four providing the highest availability. The ratings can be written as $T_N E_N A_N M_N$, with TEAM standing for the four categories and N being the rating of the corresponding category. Higher ratings are more resilient and reliable but more costly. Higher ratings are inclusive of the requirements for lower ratings. So, a data center with level 3 telecommunications, level 2 electrical, level 4 architectural, and level 3 mechanical infrastructure would be classified as TIA-942 level Rating $T_3 E_2 A_4 M_3$. The overall level rating for the data center would be level 2, the rating of the lowest level portion of the infrastructure (electrical level 2).

The TIA-942 level classifications are specified in more detail in ANSI/TIA-942-A. There are also other schemes for assessing the reliability of data centers. In general, systems that require more detailed analysis of the design and operation of a data center provide a better indicator of the expected availability of a data center.

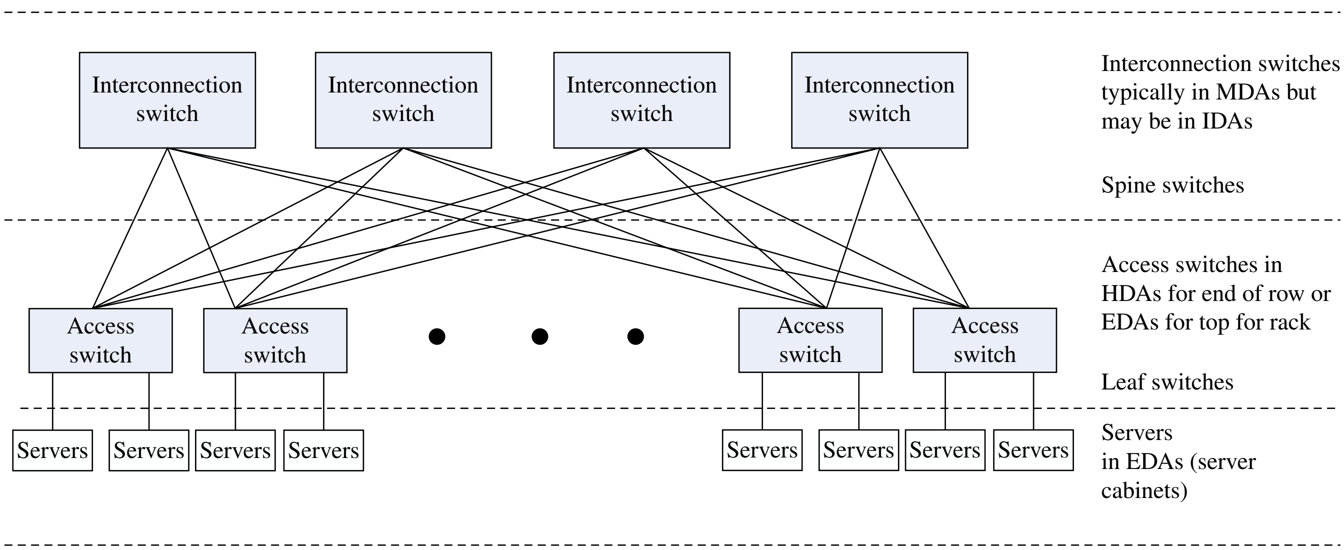


FIGURE 14.12 Data center switch fabric example.

14.13 CONCLUSION AND TRENDS

The requirements of telecommunications cabling, including maximum cable lengths, size, and location of telecommunications distributors, and requirements for cable pathways influence the configuration and layout of the data center.

The telecommunications cabling infrastructure of the data center should be planned to handle the expected near-term requirements and preferably at least one generation of system and network upgrades to avoid the disruption of removing and replacing the cabling.

For current data centers, this means the following:

- Balanced twisted-pair cabling should be Category 6A or higher.
- Multimode optical fiber should be OM4.
- Either install or plan capacity for single-mode optical fiber backbone cabling within the data center.

It is likely that LAN and SAN connections for servers will be consolidated. The advantages of consolidating LAN and SAN networks include the following:

- Fewer connections permit the use of smaller form factor servers that cannot support a large number of network adapters.
- Reduces the cost and administration of the network because it has fewer network connections and switches.
- It simplifies support because it avoids the need for a separate Fiber Channel network to support SANs.

Converging LAN and SAN connections requires high-speed and low-latency networks. The common server connection for converged networks will likely be 10 Gbps Ethernet. Backbone connections will likely be 100 Gbps Ethernet or higher.

The networks required for converged networks will require low latency. Additionally, cloud computing architectures typically require high-speed, device-to-device communication within the data center (e.g., server-to-storage array and server-to-server). New data center switch fabric architectures are being developed to support these new data center networks.

There are a wide variety of implementations of data center switch fabrics. See Figure 14.12 for an example of the fat tree or leaf-and-spine configuration, which is one common implementation.

The various implementations and the cabling to support them are described in ANSI/TIA-942-A-1. Common attributes of data center switch fabrics are the need for much more bandwidth than the traditional switch architecture and many more connections between switches than the traditional switch architecture.

When planning data center cabling, consider the likely future need for data center switch fabrics.

FURTHER READING

For further reading, see the following telecommunications cabling standards:

- ANSI/BICSI-002. Data Center Design and Implementation Best Practices Standard
- ANSI/NECA/BICSI-607. Standard for Telecommunications Bonding and Grounding Planning and Installation Methods for Commercial Buildings

- ANSI/TIA-942-A. Telecommunications Infrastructure Standard for Data Centers
- ANSI/TIA-942-A-1. Cabling Guidelines for Data Center Fabrics
- ANSI/TIA-568-C.0. Generic Telecommunications Cabling for Customer Premises
- ANSI/TIA-569-C. Telecommunications Pathways and Spaces
- ANSI/TIA-606-B. Administration Standard for Telecommunications Infrastructure
- ANSI/TIA-607-B. Telecommunications Bonding and Grounding (Earthing) for Customer Premises
- ANSI/TIA-758-B. Customer-Owned Outside Plant Telecommunications Infrastructure Standard

In Europe, the TIA standards may be replaced by the equivalent CENELEC standard:

- CENELEC EN 50173-5. Information Technology: Generic Cabling—Data Centers
- CENELEC EN 50173-1. Information Technology: Generic Cabling—General Requirements
- CENELEC EN 50174-1. Information Technology: Cabling Installation—Specification and Quality Assurance
- CENELEC EN 50174-2. Information Technology: Cabling Installation—Installation Planning and Practices Inside Buildings
- CENELEC EN 50310. Application of Equipotential Bonding and Earthing in Buildings With Information Technology Equipment

In locations outside the United States and Europe, the TIA standards may be replaced by the equivalent ISO/IEC standard:

- ISO/IEC 24764. Information Technology: Generic Cabling Systems for Data Centers
- ISO/IEC 11801. Information Technology: Generic Cabling for Customer Premises
- ISO/IEC 14763-2. Information Technology: Implementation and Operation of Customer Premises Cabling—Planning and Installation

Note that there is no CENELEC or ISO/IEC equivalent to ANSI/TIA-942-A-1, *Cabling Guidelines for Data Center Fabrics*. The ISO/IEC standard for telecommunications bonding and earthing is being developed.

Also note that standards are being continually updated; please refer to the most recent edition and all addenda to the listed standards.

15

DEPENDABILITY ENGINEERING FOR DATA CENTER INFRASTRUCTURES

Malik Megdiche

Schneider Electric, Grenoble, France

15.1 INTRODUCTION

Reliability engineering is defined as the science of failure. The first issues and reliability concepts appeared at the beginning of the twentieth century. Today, reliability engineering is widely used in many areas. A short history of reliability engineering is presented in Figure 15.1.

Reliability engineering uses equipment reliability statistics, probability theories, system functional analysis, and dysfunctional analysis to set requirements, measure or predict reliability, identify system weakness points, and propose improvements of the system. Various reliability engineering techniques are used in reliability engineering:

- Equipment reliability analysis
  - Field experience reliability statistics
  - Reliability testing
  - Accelerated life testing
- System reliability and availability analysis
  - Qualitative analysis
    - Hazard risk analysis
    - Failure mode and effects analysis (FMEA)
  - Reliability prediction
    - Electronic FMEA
    - Fault tree analysis
    - Statistical simulations
  - Maintainability analysis
- Integrated logic support

Reliability engineering techniques can be used for Reliability, Availability, Maintainability, and Safety purposes. As this chapter is dedicated to reliability and availability engineering of data center infrastructure, safety and security are not considered. However, the same concepts could be applied to study safety and security issues.

The dependability of a system deals with the following attributes:

- Availability: readiness for correct service
- Reliability: continuity of correct service
- Maintainability: to undergo modifications and repairs

Dependability will be used to name the reliability and availability performance of a data center infrastructure.

The next parts will provide:

- The understanding of basic concepts of system reliability and availability analysis including equipment dependability data and dependability methods
- The main features to be assessed to perform a relevant analysis
- An application guide to use to a dependability analysis for data infrastructure at design phase and operational phase

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

1900 ☐ *Railway field and development of statistical standards for mechanical parts*

☐ *Then appearance of electricity and use of transformers and redundant lines to increase power availability*

1939 ☐ *Introduction of quantitative objectives of accident occurrence per hour of operation (aircraft)*

1940 ☐ *The V1 rocket, Robert Lusser, and Probability of failure of elements in series*

1950 ☐ *Electronics with tubes, 30% of availability, and MTBF appearance*

1960 ☐ *Space conquest and formalization by the United States of Dependability methods (MIL standards)*

☐ *Nuclear engineering required a large number of reliability and safety analysis*

☐ *Diffusion of the operational and predictive evaluation techniques; acceptance of the concept of reliability (MIL, CNET, etc.)*

☐ *Maintenance consideration (system approach)*

1980 ☐ *Industrial data processing and human factor consideration*

2000 ☐ *There is no industrial field in which reliability has not been taken into account in all the life cycles of the products*

FIGURE 15.1 Historical development of dependability engineering during the twentieth century.

15.2 DEPENDABILITY THEORY

Dependability engineering is based on probability theories to be able to perform statical estimations. A few basic concepts will be presented in this section including:

- The definition of dependability vocabulary
- The definition of dependability data and indexes
- Some basic probabilistic calculations that could be used for dependability assessment

15.2.1 System Dependability Analysis Definition

System dependability analysis basics are to study the consequences of component failures on the system based on:

- Equipment failures and maintenance data
- System behaviors in case of failures

15.2.2 System Dependability Indexes

To analyze and estimate the dependability of system, the main system dependability indexes are the following.

15.2.2.1 Reliability This is the ability of a product to perform a given function, in specified conditions, for a given period of time.

Mathematical indexes to be used are:

The reliability, $R(t)$ = "probability to experience no failure on [0;t]"

The unreliability, $\bar{R}(t)$ = "probability to experience at least one failure on [0;t]"

$$\bar{R}(t) = 1 - R(t)$$

The mean failure frequency, F = "estimated number of failures per year (or hour)."

15.2.2.2 Availability It is the ability of an item to be in a state to perform a required function.

Mathematical indexes to be used are:

The availability, $A(t)$ = "probability to be capable to perform a required function at t"

The unavailability, $\bar{A}(t)$ = "probability not to be capable to perform a required function at t"

Asymptotic availability, $A = \lim_{t=\infty} A(t)$

Asymptotic unavailability, $\bar{A} = \lim_{t=\infty} \bar{A}(t)$

$$\bar{A}(t) = 1 - A(t) \quad \text{and} \quad \bar{A} = 1 - A$$

15.2.2.3 Maintainability It is the ability of an item to be repaired.

The maintainability, $M(t)$ = "probability to be repaired at t"

15.2.2.4 System Functions System dependability indexes are associated to a function or a group of functions of the system as mentioned in Figure 15.2.

15.2.3 Equipment Dependability Data

15.2.3.1 MTTF, MTBF, MDT, MTTR, and MUT Figure 15.3 shows the equipment state as a function of time.

Time to failure and downtime are random variables.

The mean values are defined as follows:

MTTF = mean time to (first) failure

MTBF = mean time between failure

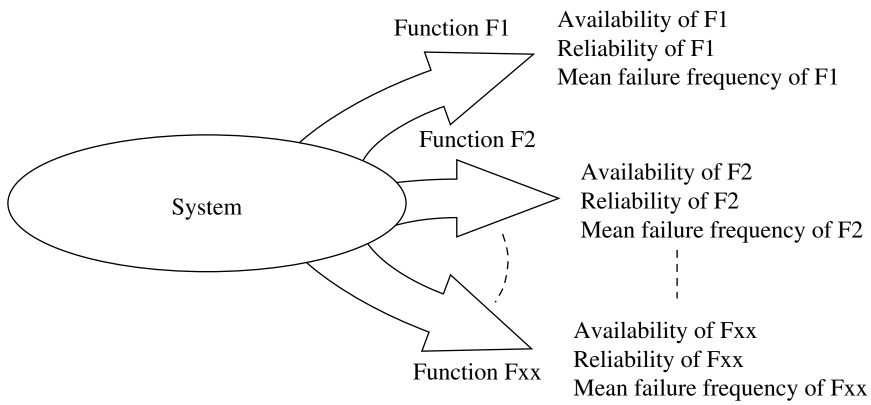


FIGURE 15.2 System external functions.

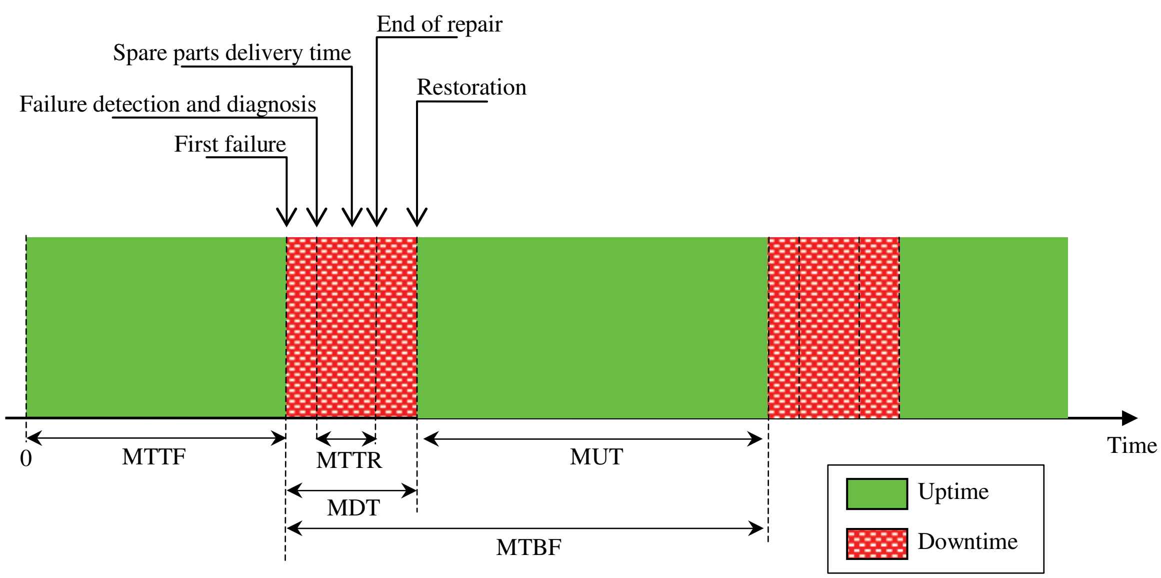


FIGURE 15.3 Equipment state during its service period.

MDT = mean downtime (= fault detection time + spare delivery time + MTTR)

MUT = mean uptime

MTTR = mean time to repair

It is important to note that:

- MTTF and MTBF are statistical values and are meant to be the mean over a long period of time and a large number of units.
- As MTBF = MTTF + MDT, MTBF value depends on the time to repair and hence is dependent on site service maintenance, whereas MTTF is an equipment data that is not dependent on the time to repair.
- Technically, MTBF should be used only in reference to a repairable item, while MTTF should be used for nonrepairable items. However, MTBF and MTTF are commonly used for both repairable and nonrepairable items.

15.2.3.2 Failure Rate The failure rate $\Lambda(t)$ is defined as the probability to fail between $[t;t+\Delta t]$:

$$\Lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \cdot P\left[\text{failure during } (t;t+\Delta t) \text{ knowing that no failure happened before}\right].$$

$$\Lambda(t) = -\frac{1}{R(t)} \cdot \frac{dR(t)}{dt}$$

The failure rate is not always constant, and its evolution over time can be described by the well-known "bathtub" curve as represented in Figure 15.4.

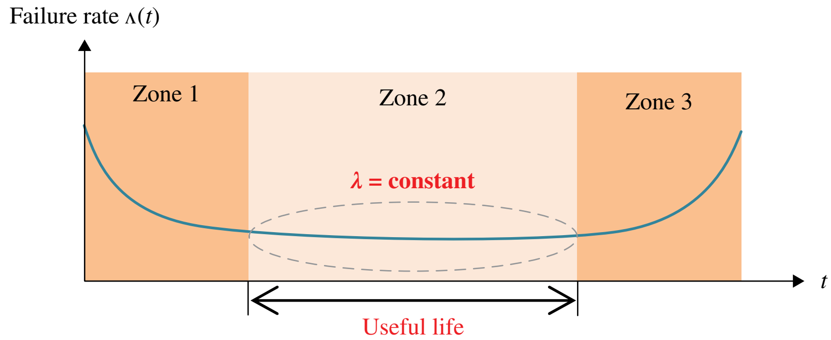


FIGURE 15.4 Failure rate curve.

There are three distinct zones on this curve:

Zone 1: Early-life or infant mortality phase
During this period, the failure rate value decreases. Failures are due to latent faults. Burn-in is intended to eliminate latent failures in the factory and prevent defective parts from being shipped to customers. Commissioning tests are also intended to detect equipment failure before its operational phase. It should minimize the size of this zone.

Zone 2: Useful life period
Failures occur randomly, and the failure rate is constant. In dependability studies, it is common to use constant failure rates corresponding to the useful period.

Zone 3: Wear-out phase
At the end of the useful life period, aging affects components and generates drastic and uncontrolled increase of their failure rate. It is the beginning of the wear-out phase.

If the failure rate is assumed to be constant (zone 2):

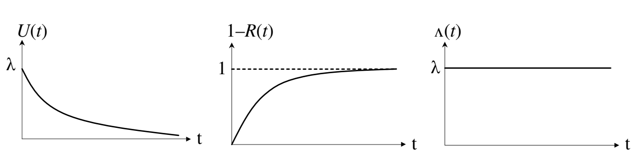
$$\Lambda(t) = \lambda$$

$$R(t) = 1 - e^{-\lambda t}$$

The random variable "time to failure" follows an exponential distribution $e^{-\lambda t}$.

The mean of the random variable "time to failure" is MTTF = (1/λ).

Probability density function (failure density)	Cumulative distribution function (unreliability)	Mean value (MTTF)
$U(t) = \lambda \cdot e^{-\lambda \cdot t}$	$1 - R(t) = 1 - e^{-\lambda \cdot t}$	$\mathrm{MTTF} = \frac{1}{\lambda}$



Commonly, the failure rate is assumed to be constant and written as λ. This assumption permits:

- Simple field experience failure estimations
  An estimation of the failure rate is defined as the ratio between the number of observed failures and the cumulative operating time of the products. Therefore, its unit is consistent with the inverse of a time.
- Simple system predictive reliability and availability calculations

Warning

A common misunderstanding is to mix the concepts of the lifetime and the MTTF (or MTBF) of a component. These two parameters are different as mentioned in Figure 15.5.

The lifetime is set according to the wear-out phase to replace the component before its failure rate increases due to aging effects, whereas the MTTF is related to the random failure frequency during the lifetime.

Particular Case

For some component, the failure rate γ is expressed per number of operation instead of hour of operation (opening/closing cycle or start sequence).

In this case, the following equation is used to convert the failure rate in failure per hour:

$$\lambda = \gamma \cdot N_{\text{operations}}$$

Failure per hour
Failure per operation
Operations/our

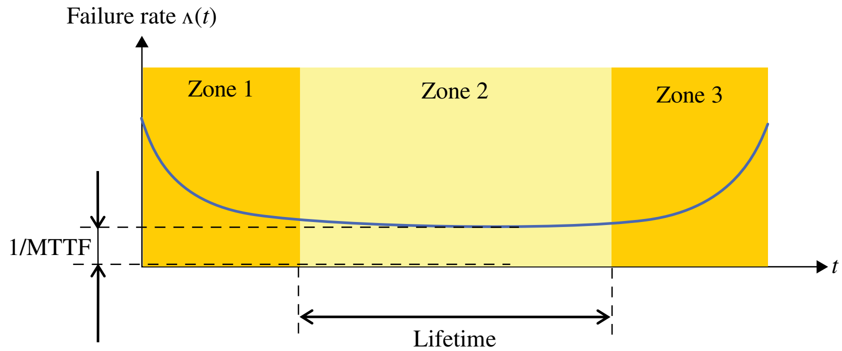


FIGURE 15.5 Lifetime versus MTTF.

15.2.3.3 Failure Modes To perform the analysis of the consequences of each possible failure, it is important to define clearly the failures for each component of the system. As a component can fail in multiple ways and due to multiple causes, defining all the possible failures for a component can be very long and will lead to a huge number of possible failures when studying a system composed of several components.

To simplify the dependability data of a component, an interesting way is to group the failures leading to the same consequences: the degradation of a function. This group of failures is named "failure mode."

The failure mode is defined as the degradation of a function of the component.

Table 15.1 shows an example of the definition of the failure mode of a circuit breaker.

For each failure mode, the associated failure rate and mean downtime have to be determined.

15.2.3.4 Curative Maintenance Data As mentioned previously, the global downtime after a failure can be decomposed in a series of events, which is represented in Figure 15.6.

The global downtime of a component is the sum of:

- The time to detect the failure

A failure can be detected by:

- Its direct consequence on the system functions (e.g., a short circuit will trip the upstream circuit breaker and the supervisory control and data acquisition (SCADA) or the users will detect immediately the failure)
- A protection system
- A watchdog trip
- A periodical test
- A preventive maintenance operation

Note that the monitoring system can minimize the failure detection

TABLE 15.1 Circuit breaker failure modes

Failure modes of a circuit breaker	Contributions (% of total failures)	Failure rate
Fail to trip on fault	xx	$\lambda_1 = xx\% \cdot \lambda$
Spurious opening	xx	$\lambda_2 = xx\% \cdot \lambda$
Unintended closing	xx	$\lambda_3 = xx\% \cdot \lambda$
Fail to open on demand	xx	$\lambda_4 = xx\% \cdot \lambda$
Fail to close on demand	xx	$\lambda_5 = xx\% \cdot \lambda$
Insulation breakdown	xx	$\lambda_6 = xx\% \cdot \lambda$

- The time to diagnostic the failure

The time to diagnostic includes:

- The customer monitoring system functions
- The customer on-site maintenance shift
- The customer on-site maintenance competencies
- The manufacturer service maintenance contract

Note that the monitoring system and disturbance analysis functions can minimize the failure diagnostic

- The spare parts delivery time
- The time to lock out the equipment
- The mean time to repair or to replace the equipment
- The time to unlock and restore the equipment

It is important to remember that the maintenance parameters for a component depend highly on:

- The manufacturer service contracts
- The spare parts delivery time
- The manufacturer service contracts
- The customer on-site maintenance
- The equipment MTTR for each failure type

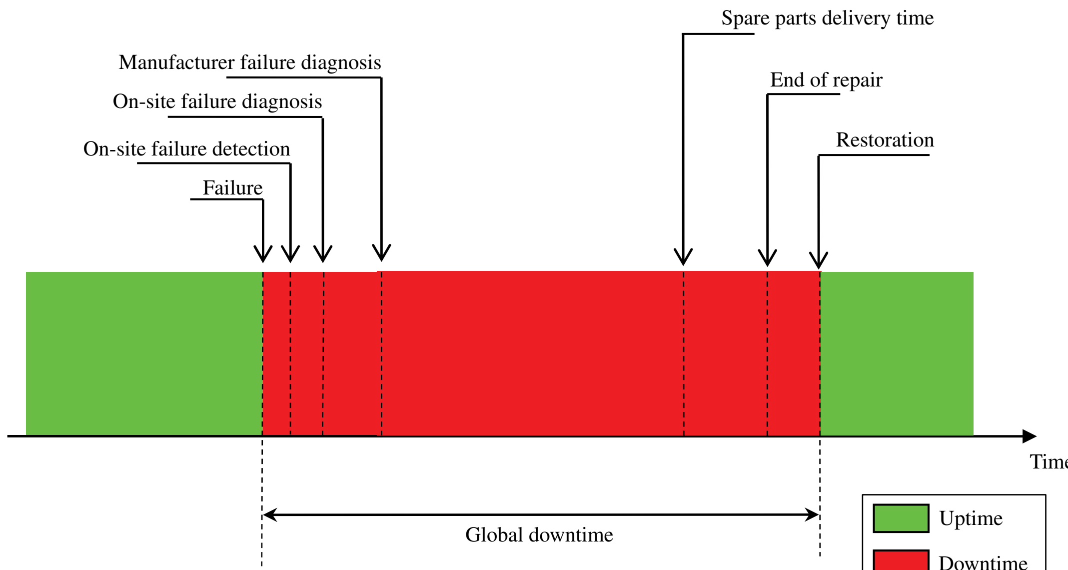


FIGURE 15.6 Curative maintenance times.

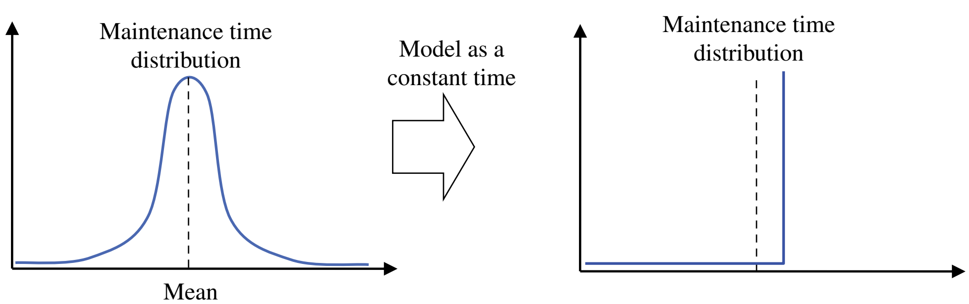


FIGURE 15.7 Maintenance time distribution model.

All maintenance times are also random variables. It is common to approximate these random times by assuming that the maintenance times are constant as shown in Figure 15.7.

However, this approximation has to be carefully assumed according to:

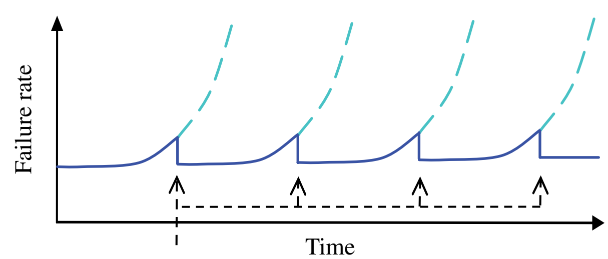
- The relation between some equipment MDT and time-limited redundancies
- The relation between some equipment MDT and time criticality of process interruptions

15.2.3.5 *Preventive Maintenance Data*

Preventive maintenance is a planned maintenance that is designed to:

- Improve equipment life
- Guarantee no aging phenomena
- Detect latent failures

Preventive maintenance ensures constant failure rates as shown in Figure 15.8.



Periodical maintenance permits to guarantee constant equipment failure rate.

FIGURE 15.8 Failure rate evolution according to periodical maintenance.

Preventive maintenance operation can include:

- Inspections
- Cleaning
- Tests and measurements
- Adjustments
- Parts replacements

TABLE 15.2 Preventive maintenance operations

Scheduled maintenance data				
Equipment maintenance operation	Maintenance operation features	What are the equipment or functions unavailable during the operation	Frequency/year	Duration (h)
Equipment xx—maintenance operation *n* xx				
…				

However, maintenance operation may require to lock out the equipment or to lead to some equipment function unavailability during the operation.

For each equipment preventive maintenance operation, the required data to perform a dependability analysis can be summed up as mentioned in Table 15.2.

15.2.3.6 Particular Case of the Utility Dependability Data IEEE 1366 standard defines indexes to measure electrical utility distribution system dependability. Here are the definitions of these indexes:

System Average Interruption Frequency Index (SAIFI)

$$\mathrm{SAIFI} = \frac{\text{number of customer interruptions}}{\text{number of customers served}}$$

System Average Interruption Duration Index (SAIDI)

$$\mathrm{SAIDI} = \frac{\sum \text{customer minutes interrupted}}{\text{number of customers served}}$$

Average Service Availability Index (ASAI)

$$\mathrm{ASAI} = \left(1 - \frac{\mathrm{SAIDI}}{\text{minutes per year}}\right) \times 100$$

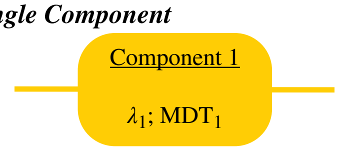
Momentary Average Interruption Frequency Index (MAIFI)

$$\mathrm{MAIFI} = \frac{\text{number of customer momentary interruptions}}{\text{number of customers served}}$$

Similar dependability data can be calculated for other utilities like water or gas utilities.

15.2.4 Basic System Dependability Modeling

15.2.4.1 Single Component

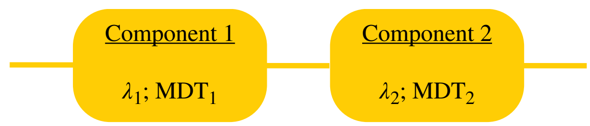


The availability and reliability of a single component are defined by the following formula:

$$\lambda_{\text{equivalent}} = \lambda_1$$

$$\text{Unavailability} = \frac{\lambda_1}{\lambda_1 + \dfrac{1}{\mathrm{MDT}_1}} \approx \lambda_1 \cdot \mathrm{MDT}_1$$

15.2.4.2 Nonredundant Components

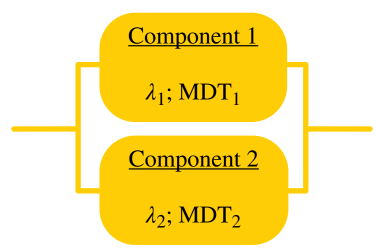


The availability and reliability of two nonredundant components are defined by the following formula:

$$\lambda_{\text{equivalent}} = \lambda_1 + \lambda_2$$

$$\text{Unavailability} = \frac{\lambda_1}{\lambda_1 + \dfrac{1}{\mathrm{MDT}_1}} + \frac{\lambda_2}{\lambda_2 + \dfrac{1}{\mathrm{MDT}_2}} \approx \lambda_1 \cdot \mathrm{MDT}_1 + \lambda_2 \cdot \mathrm{MDT}_2$$

15.2.4.3 Two Components with Active Redundancy "Two components with active redundancy" means that the two components are running together. If one fails, the other one is able to keep the system running.

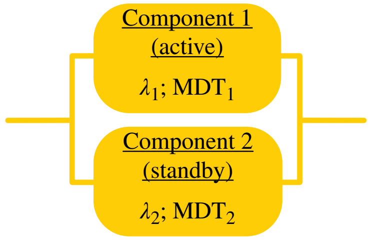


The availability and reliability of two redundant components are defined by the following formula:

$$\lambda_{\text{equivalent}} \approx \lambda_1 \cdot \lambda_2 \cdot (\mathrm{MDT}_1 + \mathrm{MDT}_2)$$

$$\text{Unavailability} = \frac{\lambda_1}{\lambda_1 + \dfrac{1}{\mathrm{MDT}_1}} \cdot \frac{\lambda_2}{\lambda_2 + \dfrac{1}{\mathrm{MDT}_2}} \approx \lambda_1 \cdot \mathrm{MDT}_1 \cdot \lambda_2 \cdot \mathrm{MDT}_2$$

15.2.4.4 Two Components with Passive Redundancy Two components with active redundancy means that one component is in standby mode (partially activated or switched off). If the first component fails, the second one is activated to keep the system running.



The availability of two redundant components is defined by the following formula:

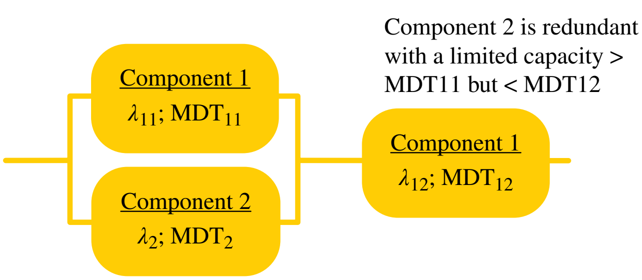
$$\lambda_{\text{equivalent}} = \lambda_1 \cdot \lambda_2 \cdot \text{MDT}_2$$
$$\text{Unavailability} \approx \lambda_1 \cdot \lambda_2 \cdot \text{MDT}_2 \cdot \min(\text{MDT}_1;\text{MDT}_2)$$

15.2.4.5 Partial Redundancy Redundancy can be achieved on two components only partially due to capacity limitation.

This capacity limitation can be:

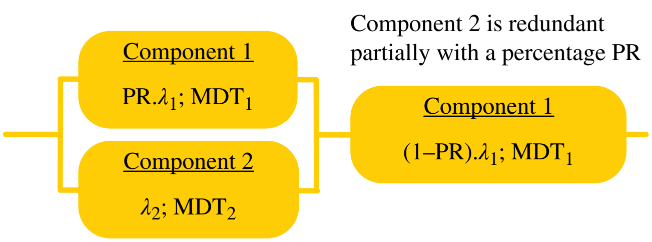
- Time limited like components using batteries, fuel storage, or water storage

  In this case, the component is redundant for any failure that does not exceed the time-limited capacity of the component.



- Dependent on a variable condition
  - Environment condition (temperature, humidity, etc.)
  - Load probability that can exceed the component capacity

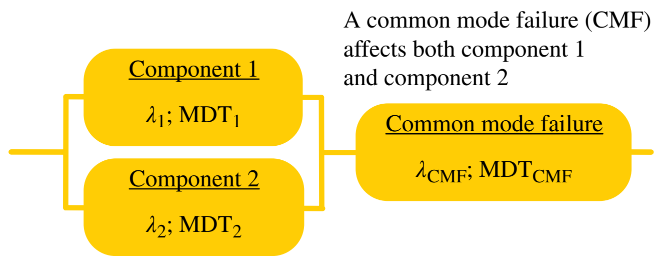
  To model that type of redundancy, the ratio PR defined as "the percentage of time for which the redundancy is achieved" has to be determined. Then the dependability diagram canted as follows:



15.2.4.6 Common Cause Failures A common cause failure is one in which a single failure or condition affects the operation of multiple devices that would otherwise be considered independent. Common cause failures can be classified as follows:

- Human error
  - Error during design, manufacturing, and installation phases
  - Unintended action
  - Inadequate or incorrect procedure
  - Inadequate training
  - Inadequate maintenance
- Environment
  - Fire and smoke
  - Temperature, humidity, and moisture
  - Electromagnetic field
  - Animals and bio-organisms
  - Contamination, dust, and dirt
  - Wind, flood, lightning, snow, ice, and earthquake

The common mode failure (CMF) can be modeled as follows:

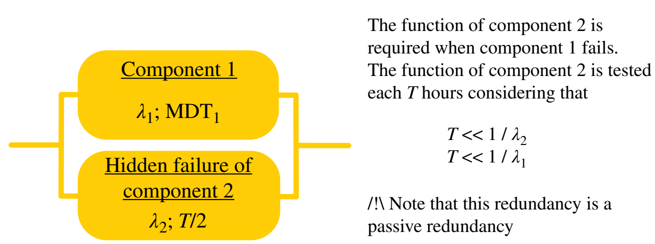


15.2.4.7 Hidden Failure A failure of a function that is normally required to keep the system running will have no direct consequence on the system. This type of failure is named hidden failure or latent failure.

A hidden failure is detected:

- When the failed function is required
- During periodical test of the function

A simple way to represent a hidden failure is the following:



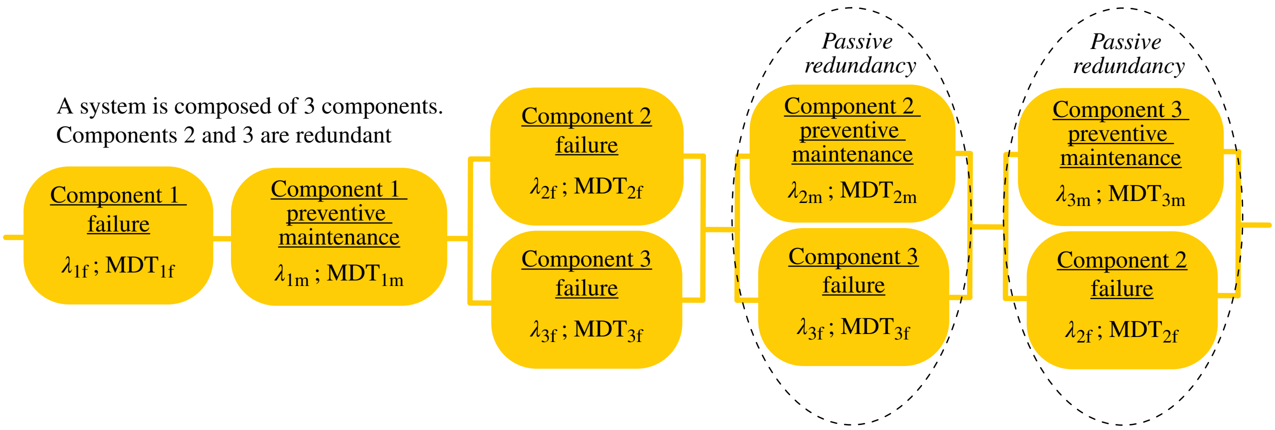


FIGURE 15.9 Reliability calculation of 2 redundant components taking into account preventative maintenance.

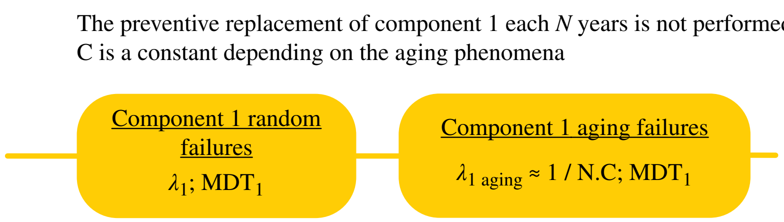


FIGURE 15.10 Reliability calculation taking into account ageing failures.

15.2.4.8 Preventive Maintenance If the process needs to run during a preventive maintenance operation that require to lock out some components, the maintenance operation needs to be taken into account for the dependability analysis as mentioned in Figure 15.9.

15.2.4.9 Lifetime and Preventive Replacement As mentioned previously, the preventive replacement of aging part of a component permits to guarantee constant failure rate.

In some case, the customer prefers to replace the aging parts after a failure; a simple and pessimistic way to model the increased failure rate is mentioned in Figure 15.10.

15.3 SYSTEM DYSFUNCTIONAL ANALYSIS

This section is dedicated to the system dependability analysis methods. After the presentation of the main steps of system dependability analysis, the main system failure analysis methods are described.

15.3.1 Dependability Analysis Methodology

The general dependability methodology as used in dependability engineering field is simply represented in Figure 15.11.

15.3.1.1 Preliminary Risk Analysis This preliminary step is essential to define precisely what are the critical processes to be studied and to perform the right study. The method is the following:

- Identify the external functions of the system and define each Unexpected Event (UE) of the system as the degradation or the unavailability of one or several functions of the system (Fig. 15.12).

For example, UE definitions can be:

- UE1: “Loss of F1 during more than …”
- UE2: “Loss of F1 during more than …”
- UE3: “Loss of F3 and F4 during more than …”

Note that sometimes the consequences of an expected event are very different depending on the duration of the UE. In this case, it is important to define different UE with different durations:

- Define reliability and/or availability targets for each UE as function of risk acceptance limit and the UE gravity as shown in Figure 15.13.

15.3.1.2 System Functional Analysis The aim of this task is understand how the system works. It has to characterize:

- The operational modes of the system including eventual upgrade and evolution phases
- The process automation system

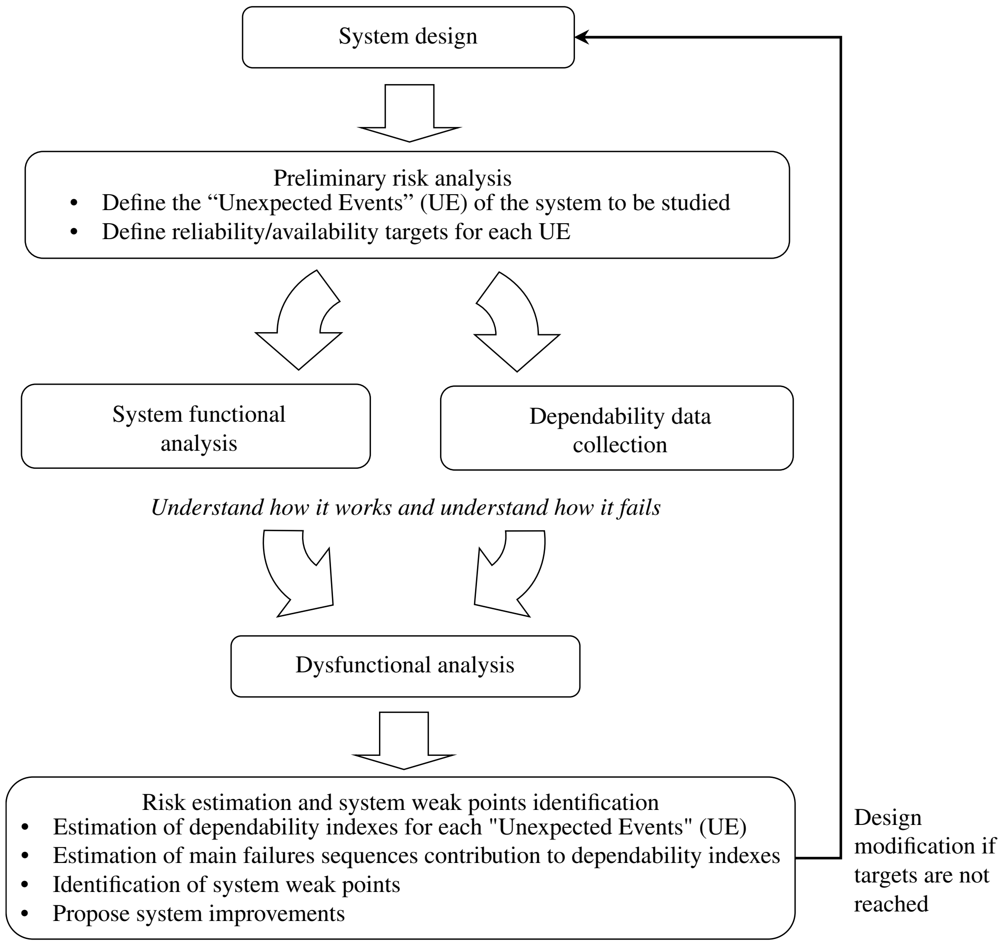


FIGURE 15.11 Dependability methodology.

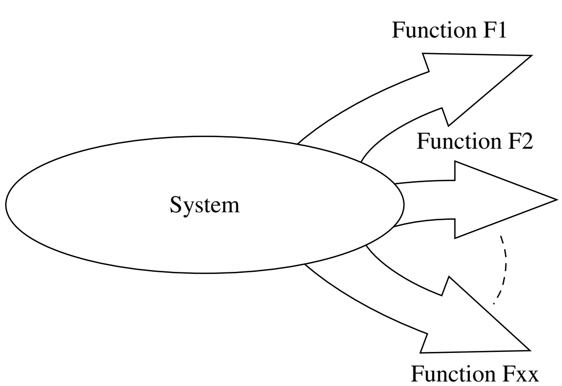


FIGURE 15.12 System external functional analysis.

- The protection and automation systems
- The monitoring system
- The emergency maintenance actions to reconfigure the system (Fig. 15.14)

Note that some assumptions are needed to determine the consequences of failure sequence:

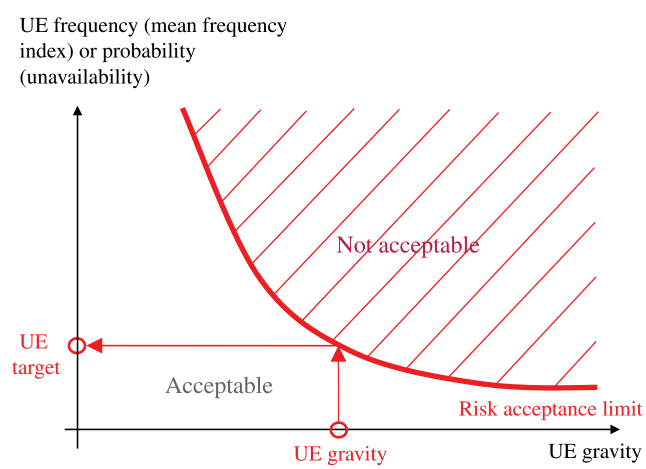


FIGURE 15.13 Risk acceptance graph.

- Equipment tolerance to supply interruptions
- Protection and automation system behaviors
- Emergency maintenance behaviors
- Starting time of equipment or functions after a blackout

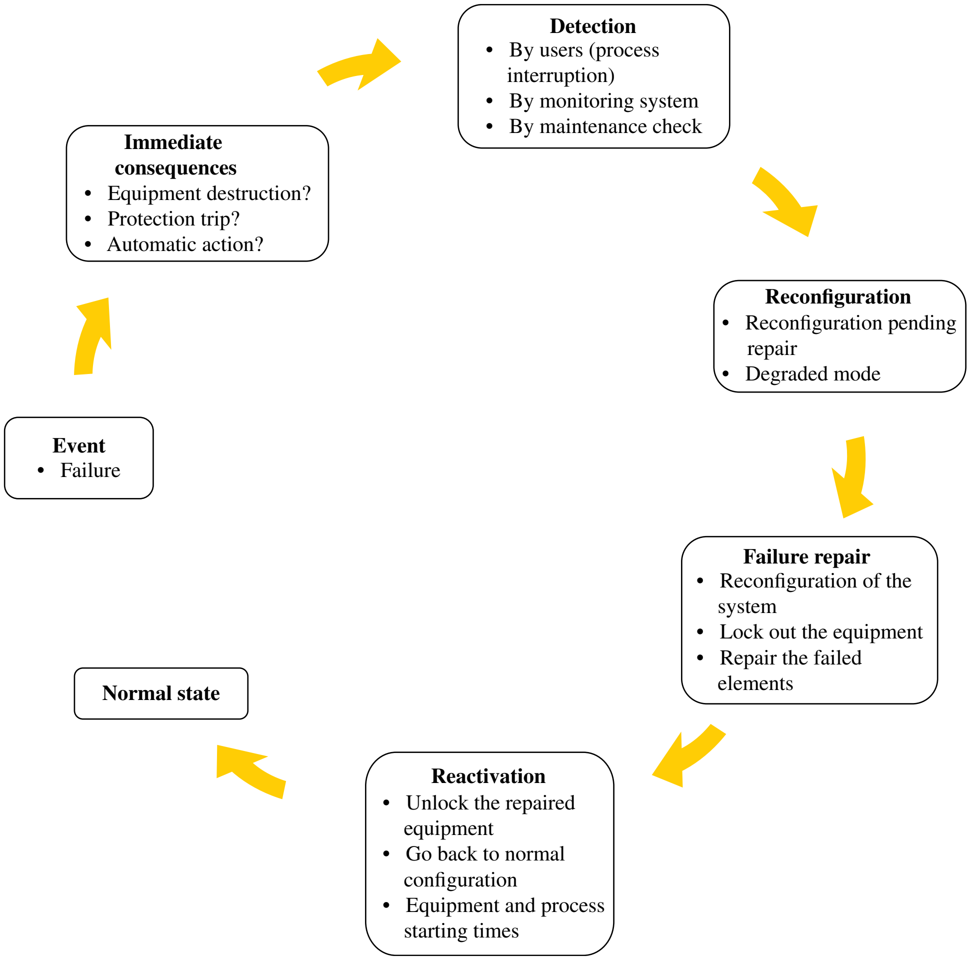


FIGURE 15.14 System behavior after a failure.

These assumptions have to be done according to equipment datasheet, design studies, and on-site maintenance.

15.3.1.3 Dependability Data Collection As mentioned previously, the dependability data collection needs to include both component failure data and maintenance data as shown in Table 15.3.

Some specific additional data can be added:

- The time to perform manual operations like switching operation to change the system configuration or a manual restart
- Onsite maintenance team level, etc.

15.3.1.4 Dysfunctional Analysis and System Weak-Point Identification The dysfunctional analysis is the study of the consequences of each possible failure on the system.

As shown in Figure 15.15, based on the possible failures of each component and the behavior of the system, the dysfunctional analysis:

- Generates the failure sequence
- Determines all the actions of system until equipment restoration
- Determines if the failure sequence affects an Unwanted Event
- Computes dependability indexes for each UE
- Computes failure sequence contribution to each UE

Note that a dependability assessment has to make the assessment for single-failure sequences and also multiple-failure sequences. However, some assumptions can be done to simplify the analysis. This will be detailed further in the chapter.

The dysfunctional analysis results can be summarized for each UE as represented in Table 15.4.

If the target is not reached, the system can be improved by:

- Identifying the main failure sequence contributions to UE
- Proposing improvements that clear or minimize the critical failure consequences

TABLE 15.3 Dependability data table

	Reliability data				Curative maintenance data				Preventive maintenance data			
Equipment type	Reliability source	Failure rate (1/h)	Failure modes	Contributions to failure rate (%)	Detection time (h)	Diagnostic time (h)	Spare parts delivery time (h)	Time to repair and restore (h)	Maintenance operation features	What are the equipment or functions unavailable during the operation	Frequency (/year)	Duration (h)
Component 1												
Component 2												
Etc.												

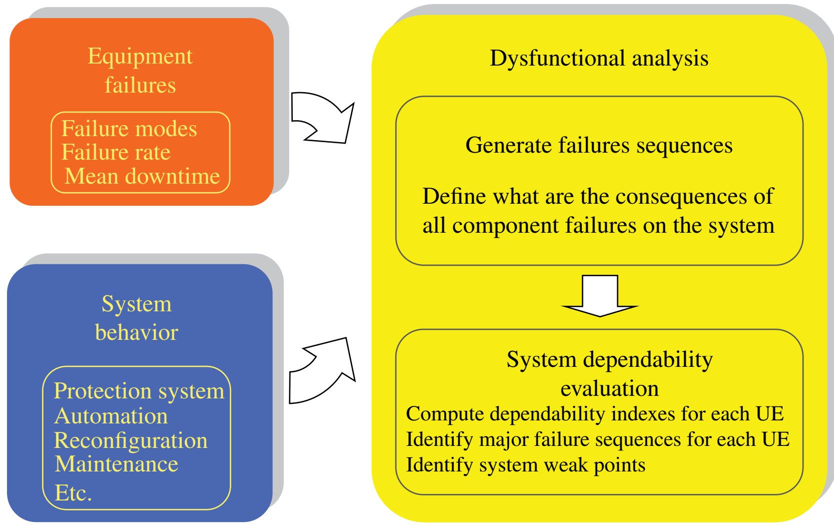


FIGURE 15.15 Dysfunctional analysis principles.

TABLE 15.4 Dependability assessment results

Unexpected event $n°$xx		
Mean frequency index	xxx/year	
Main failure sequences		Contributions to mean frequency index (%)
xx xx xx		xx xx xx
Mean unavailability index	xxx h/year	
Main failure sequences		Contributions to mean unavailability index (%)
xx xx xx		xx xx xx

This can be done in several ways by:

- Designing adequate redundancy
- Setting adequate maintenance

However, the proposed solutions has to take into consideration the following points:

- Be sure that the proposed solution is technically possible and cost-effective.
- Be sure that the proposed solution is more dependable than before.
- Keep as possible the system simple to operate.
- Respect as possible the customer habits in terms of system architecture design and operation.

15.3.1.5 Example of a Simple System The system consists in an MV/LV power system supplying LV critical loads as represented in Figure 15.16.

Primary Risk Analysis To determine the main functions, an external functional analysis is performed in Figure 15.17.

Function 1: Supply critical loads.

Function 2: Permit maintenance operation and perform installation monitoring.

Function 3: Ensure safety.

Function 4: Prevent environment pollution (EMC, chemical).

The goal of the study is to improve the system architecture to optimize the critical load power supply reliability. Therefore, the UE to be studied is defined as UE1, “loss of critical load power supply.” As an interruption of the process UE1 is critical whenever the UE duration, the target is to minimize UE1 frequency.

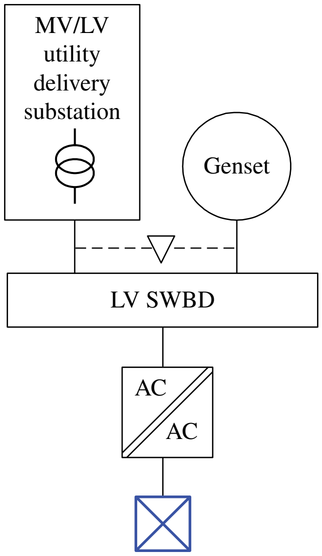


FIGURE 15.16 Single-line diagram of the MV/LV power system.

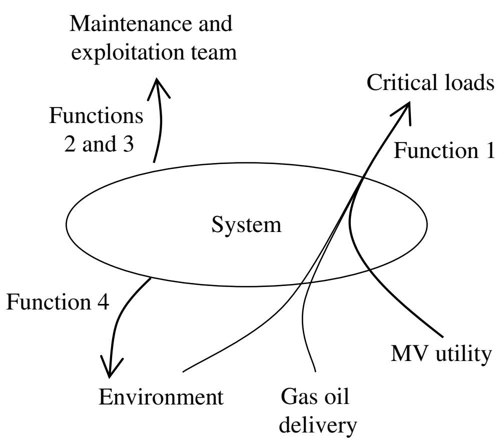


FIGURE 15.17 External functional analysis.

Functional Analysis

- The MV utility delivery is backed up by an LV Genset.
- The LV Genset is started when the switchboard is de-energized and is designed for continuous operation.
- The LV Genset fuel storage permits a 72 h autonomy at full load. A specific emergency fuel delivery service permits to refill the fuel storage during Genset long period operation.
- The LV Genset is tested each month.
- An automatic transfer switch (ATS) permits automatic changeover from one source to another.
- A UPS system with 5 min batteries permits to supply the loads during the Genset starting sequence.
- Critical loads are tolerant to power supply interruptions with voltage drop of more than 40% of rated voltage during more than 100 ms.
- Maintenance team 24/24 h permits a mean time to intervention of 15 min and is able to perform manual reconfiguration in 2 h.

Dependability Data Collection

Equipment failure mode	Failure rate (1/h)	Mean downtime (h)
Major blackout on HV grid	1.00E−06	4
MV electrical utility failure	1.00E−04	1
MV electrical utility short interruption (<3 min)	1.00E−03	0.033
Genset—fail while in standby mode	1.00E−04	365
Changeover—fail to switch	1.00E−06	365
Changeover—unexpected opening of both switches	1.00E−08	2
LV switchboard failure	1.00E−07	2190
UPS—short circuit on output	3.00E−07	168
UPS—loss of UPS path (switch on static bypass)	1.00E−05	50

Dysfunctional Analysis Results A fault tree is performed to estimate UE1 frequency and identify main failure sequences. The results are mentioned in Table 15.5.

Weak-Point Identifications and Improvement Proposal The dysfunctional assessment shows that the main contribution to critical load interruptions is the failure sequence "Genset failure during standby" and "utility failure." This means that the redundancy of MV utility with a single Genset is not sufficient. Several ways can be investigated:

- Improve Genset availability by more frequent tests and periodical maintenance and also reduce the Genset MTTR.
- Improve Genset availability by designing additional redundancy on the Genset starting system.
- Provide "$N+1$" redundant Genset power plant.

15.3.2 Main System Dysfunctional Analysis Methods

Based on the system functional analysis and the component dependability data, the system dysfunctional analysis consists in the analysis of the consequences of each possible failure sequence. Several methods can be used depending on the accuracy and the time allowed to perform the study. This section will detail some of the main methods and gives the advantages and drawbacks for each of them.

TABLE 15.5 Dysfunctional analysis results of the example

UE1 "Loss of critical load"–Estimated Mean Frequency: 0.041/year	
Main failure sequences	Contribution to UE1 frequency (%)
"Genset failure during standby" and "utility failure"	80
"Changeover—fail to switch" and "utility failure"	1
"Changeover—unexpected opening of both switches"	0
"LV switchboard failure"	2
"UPS—short circuit on output"	6
"UPS—loss of UPS path" and "utility short interruption"	11

TABLE 15.6 Example of FMECA table according to IEC 60812

Function	Failure modes	Local effect	Final effect	F	D	G	Criticality	Actions	Comments
							$= F \times D \times G$		

TABLE 15.7 Frequency index table

Ranking	Frequency	Criteria failure mode occurrence (failure rate)
1	Improbable	<1E–9/h
2	Remote	<1E–8/h
3	Occasional	<1E–7/h
4	Probable	<1E–6/h
5	Frequent	<1E–5/h

15.3.2.1 Failure Mode, Effects, and Criticality Analysis

Definition According to IEC 60812 Failure Mode, Effects, and Criticality Analysis (FMECA) is a method defined by the IEC 60812 standard:

> Failure Modes and Effect Analysis (FMEA) is a systematic procedure for the analysis of a system to identify the potential failure modes, their causes and effects on system performance (performance of the immediate assembly and the entire system or a process).
>
> FMECA is an extension to the FMEA to include a means of ranking the severity of the failure modes to allow prioritization of countermeasures. This is done by combining the severity measure and frequency of occurrence to produce a metric called criticality.
>
> It is important to note that the FMECA makes only the assessment of single failure and does not consider multiple failures.

The classical system FMECA features are the following:

- The system is divided into components (choice of detail level). The FMECA is presented in Table 15.6.
- Local and final effects are identified for each component failure modes.
- Indicators frequency (F), detection (D), and gravity (G) levels are determined using reference tables. Reference tables are defined according to the system. Examples of reference tables are given in Tables 15.7, 15.8, and 15.9.

  Various reference tables can be defined:
  - With several ranks
  - Using failure rate instead of frequency levels
  - Using or not using detection indicator
- Criticality level (frequency $F \times$ detection $D \times$ gravity G) is estimated for each failure mode. Criticality level is used as global risk assessment indicator. The risk acceptability is defined with the customer and can be presented in Table 15.10.
- The criticality level of each failure mode gives the identification of main system weak points and the main critical failures for which an action has to be taken to decrease the risk.

The FMECA is widely used for risk analysis of system for design or operational purpose. Its main advantage is its simplicity that:

- Makes the FMECA affordable for a lot of people to understand the method and its results

TABLE 15.8 Detection index table

Ranking	Detection	Criteria: likelihood of detection by design control ranking
1	Almost certain	Design control will almost certainly detect a potential cause/mechanism and subsequent failure mode
4	Moderately high	There is moderately high chance that the design control will detect a potential cause/mechanism and subsequent failure mode
7	Very low	There is very low chance that the design control will detect a potential cause/mechanism and subsequent failure mode
10	Absolutely uncertain	Design control will not and/or cannot detect a potential cause/mechanism and subsequent failure mode, or there is no design control

Note that the "failure detection" indicator means that the failure is detected and hazard risk is avoided.

TABLE 15.9 Gravity index table

Ranking	Gravity	Criteria
1	None	No discernible effect
4	Very minor	Fit and finish/squeak and rattle item does not conform. Defect noticed by most customers (>75%)
7	Very low	Vehicle/item operable but at a reduced level of performance. Customer very dissatisfied
8	Very high	Vehicle/item inoperable (loss of primary function)
9	Hazardous with warning	Very high severity ranking when a potential failure mode affects safe vehicle operation and/or involves noncompliance with government regulation with warning
10	Hazardous without warning	Very high severity ranking when a potential failure mode affects safe vehicle operation and/or involves noncompliance with government regulation without warning

TABLE 15.10 Risk acceptance matrix

Risk acceptability matrix				
	Severity = detection × gravity			
Frequency of occurrence of failure effect	Insignificant	Marginal	Critical	Catastrophic
Frequent	Undesirable	Intolerable	Intolerable	Intolerable
Probable	Tolerable	Undesirable	Intolerable	Intolerable
Occasional	Tolerable	Undesirable	Undesirable	Intolerable
Remote	Negligible	Tolerable	Undesirable	Undesirable
Improbable	Negligible	Negligible	Tolerable	Tolerable

- Requires no specific tool
- Makes it easy to perform an exhaustive analysis
- Brings a synthesis of all risks at the same time

However, its main drawback is the fact that the assessment depends highly on:

- The reference table definition
- The people that make the assessment of indicators

FMECA Customized for System Dependability Analysis FMECA can be customized to avoid the use of qualitative indicators. The principle is to make the assessment of UE of the system.

Basically, the FMECA consists in the study of each single-failure consequence on the system as described in Figure 15.18.

The customized FMECA table is presented in Table 15.11.

In case of an undetected failure (hidden failure), a pessimistic way is to consider that this failure is detected on a contingency.

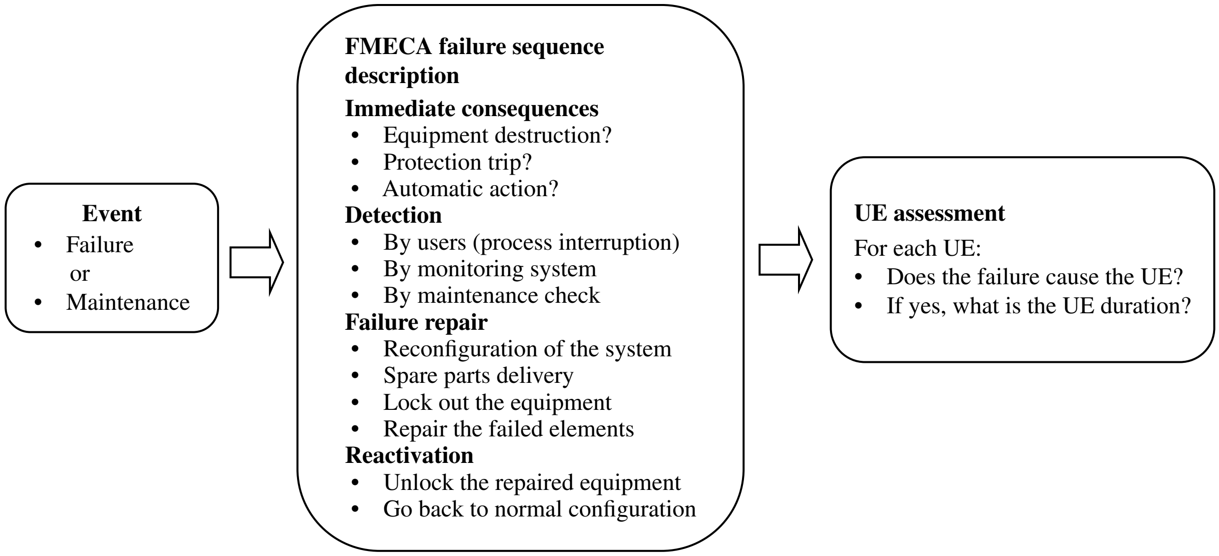


FIGURE 15.18 Assessment principle of an equipment failure mode during a FMECA approach.

The UE estimation is done by.

$$\text{Mean frequency of UE} = \sum \text{failure rate} \times \text{UE happens?}$$

$$\text{Mean unavailability of UE} = \sum \text{unavailability} \times \text{UE happens?}$$

The contribution of main failures to each index is obtained using a sorting of FMECA lines with the column "failure rate × UE happens?"

> FMECA does not make the assessment of multiple failures; it only makes the assessment of single failures for each UE.
>
> The statistical estimations of UE indexes are assuming that failures combinations are negligible compared to single failures. This assumption has to be demonstrated by verifying that the main failure combinations of the considered UE are negligible.

The FMECA results can be presented for each UE as mentioned in Table 15.4.

Simplified FMECA for System Dependability Assessment Sometimes, it is required to perform a failure analysis within a short time. In this case, it is interesting to use a simplified FMECA to identify single points of failure without any statistical estimation with an FMECA table as mentioned in Table 15.12.

The equipment-level definition is not detailed, and the failure modes are very simplified, and only the worst failure modes are selected. For example, an LV switchboard is considered as a component with a unique failure mode that results in a loss of insulation and the unavailability of the whole switchboard until repair.

15.3.2.2 Failure Combination Analysis As mentioned before, the FMECA approach permits the assessment of a UE by performing a single-contingency analysis. A single-contingency analysis is acceptable as long as multiple-failure sequences are negligible compared to single-failure sequences. In some cases, when the system is highly reliable or available, this assumption is no more valid and the dependability assessment needs an analysis of single and multiple contingencies to provide accurate results.

Principle Failure sequences can be represented as shown in Figure 15.19.

The number of failure sequences becomes very important if multiple-failure sequences are taken into account. Indeed, if a system is composed of n component with p failure mode for each element, possible failure sequences are:

- $(n \times p)$ sequences with a single failure
- 2^{n+p} sequences with two failures
- 3^{n+p} triple failure with three failures

That is why, in general, the use of a specific tool dedicated to system dependability modeling is essential. Several methods have been developed such as reliability

TABLE 15.11 Customized FMECA table

								UE1: "loss of xxxx"			UE2: "loss of xxxx"			UE3: "loss of xxxx"		
Line n°	Localization	Equipment reference	Equipment functions	Failure mode	Direct effects	Detection and consequences until normal state	Failure rate	UE happens?	Duration (h)	Unavailability (h/year)	UE happens?	Duration (h)	Unavailability (h/year)	UE happens?	Duration (h)	Unavailability (h/year)

TABLE 15.12 Simplified FMECA

							UE1: "loss of xxxx"	UE2: "Loss of xxxx"	UE3: "loss of xxxx"
Localization	Equipment reference	Equipment functions	Failure mode	Direct effects	Consequence until repair and back to normal state	Frequency index estimation	UE happens?	UE happens?	UE happens?

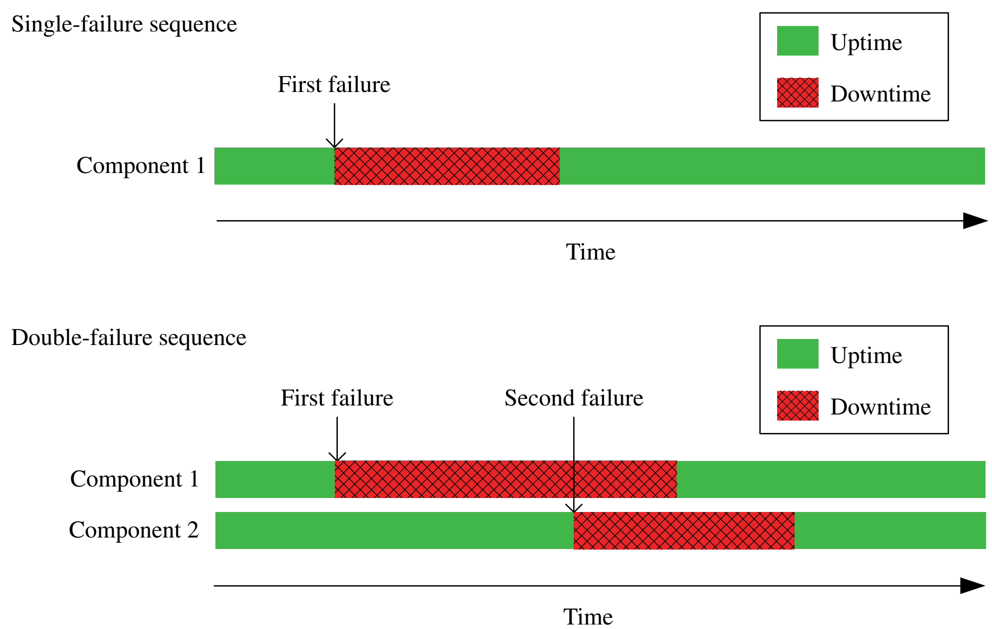


FIGURE 15.19 Single- and double-failure time sequence graphs.

block diagram, fault tree, event tree, Markov graph, and stochastic simulation.

Reliability Block Diagram The principle of a reliability block diagram is described below:

- Each component failure mode is modeled by a block with the associated dependability parameters MTTF and MTTR.
- Each UE is modeled by a reliability block diagram using blocks connected in series or parallel of redundancy as mentioned in Figure 15.20.
- According to the model, the tool automatically computes the statistical estimation of UE probability and frequency and also the minimal cut sets that are the unique combinations of component failures that lead to the UE.

Minimal cut sets can be used to determine main failure sequence contribution to UE probability or frequency.

Fault Tree The fault concept is a top-down, deductive failure analysis in which a UE of a system is analyzed using Boolean logic to combine a series of events as mentioned in Figure 15.21. Opposite with FMEA concept that consists in an analysis of failure consequences and the identification of failures that leads to the UE, the fault tree process consists in representing the causes of the UE occurrence.

The principles of a fault tree are described below:

- Each possible failure mode is modeled by a base event with the associated dependability parameters MTTF and MTTR.
- Each UE is modeled by a fault tree using the logic gates "AND," "OR," "Voting OR (k/n)," "INHIBIT," etc. to represent the possible causes of the UE occurrence.
- According to the fault tree, the tool automatically generates the associate binary decision diagram of the UE and computes the minimal cut sets, which are the minimal failure combinations that lead to the UE.
- UE probability, UE equivalent failure frequency, and minimal cut set probabilities and equivalent failure rates are computed by the tools.
- Minimal cut sets can be used to determine main failure sequence contribution to UE probability or frequency.

Event Tree As FMECA approach, the event tree is an inductive analytical diagram in which a failure (or an event) is analyzed by describing the chronological following events. The difference with an FMECA approach is that multiple failures are taken into account by using Boolean logic to determine the possible consequences as shown in Figure 15.22.

An event tree displays sequence progression, sequence end states, and sequence-specific dependencies across time.

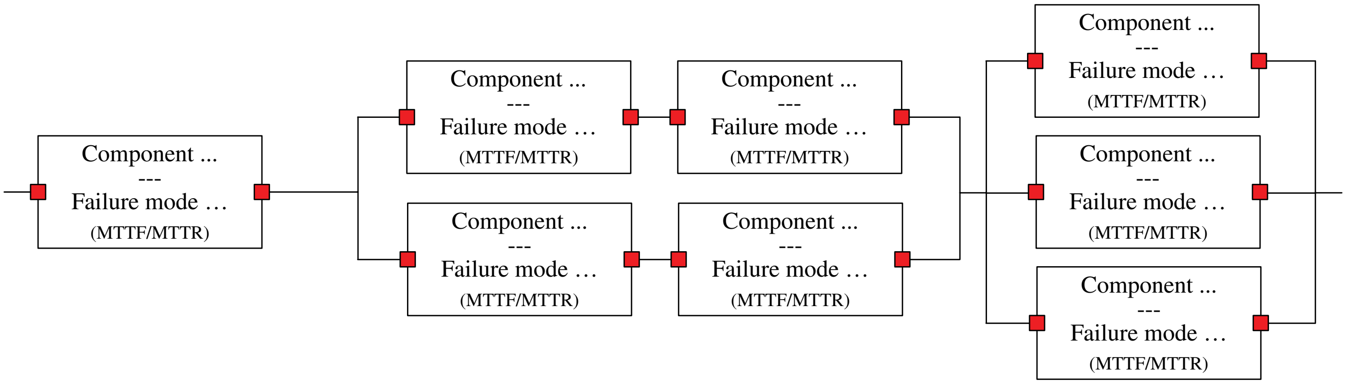


FIGURE 15.20 Reliability block diagram principle.

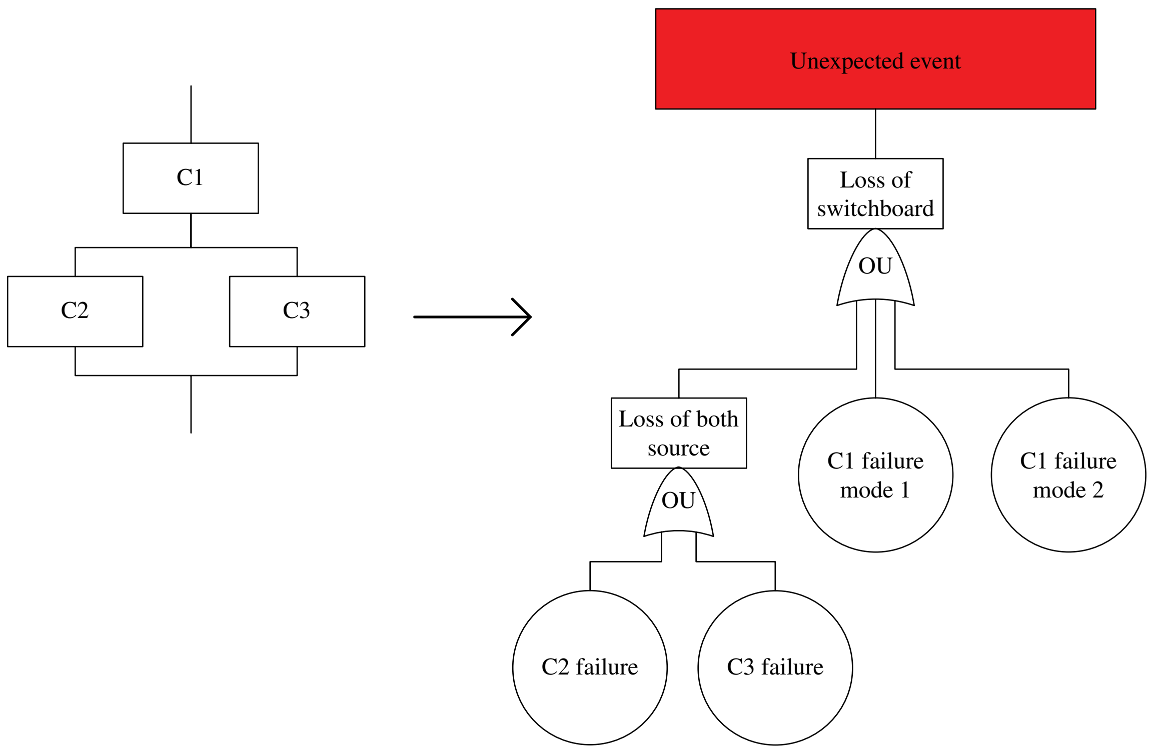


FIGURE 15.21 Fault tree example.

For each initiating event (1st event), a list of possible sequences is identified in terms of success or failure (UE is reached) and in terms of probability (unavailability and mean frequency).

After the analysis of all initiating events and all possible failures, UE probability and equivalent failure frequency are computed by the tool.

Discrete-Time Markov Chain Discrete-time Markov chain is a mathematical model in which the system is modeled by its different states and the transitions from a state to another. Figure 15.23 shows a Markov chain of a system composed of two active and redundant components.

By assuming that transitions are random exponential laws, the system can be mathematically modeled as follows:

$$\left[\frac{dP_1(t+dt)}{dt}\ \frac{dP_2(t+dt)}{dt} \dots \frac{dP_p(t+dt)}{dt}\right] = \left[P_1(t)\ P_2(t)\ \dots P_p(t)\right]\cdot A$$

where $A = \begin{bmatrix} a_{11} & \dots & a_{1p} \\ \dots & \dots & \dots \\ a_{p1} & \dots & a_{pp} \end{bmatrix}$ is the matrix of the transition rates.

The system can be computed to determine the state probabilities and then the transition mean frequency.

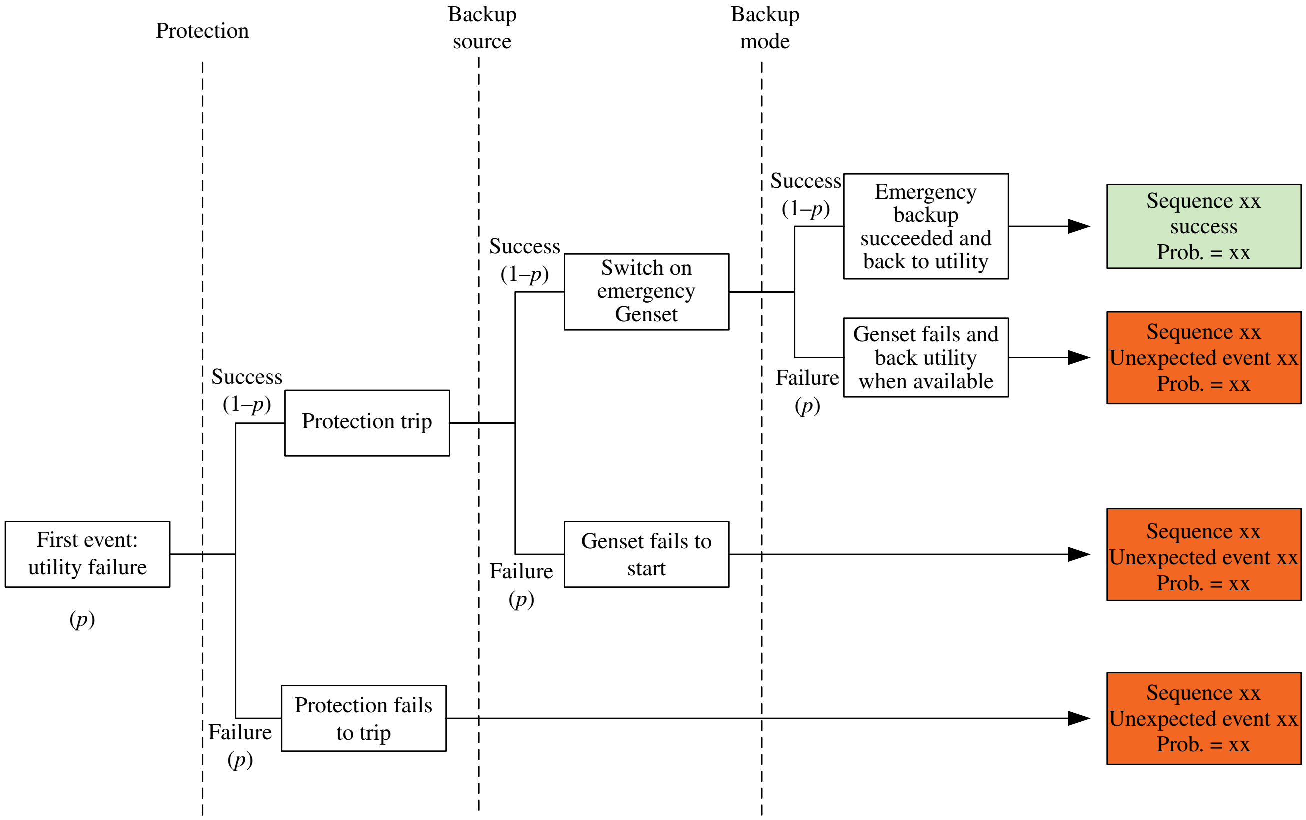


FIGURE 15.22 Event tree example.

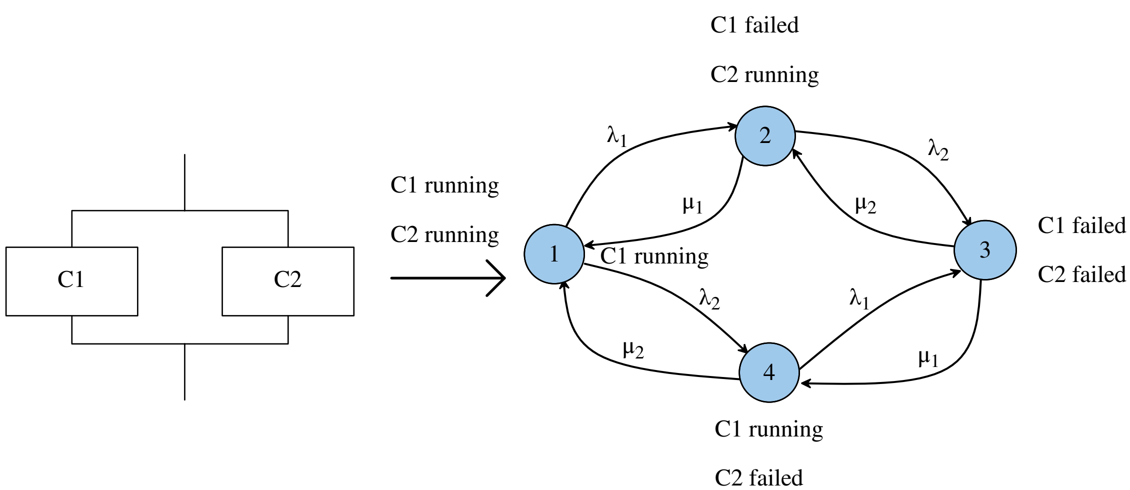


FIGURE 15.23 Markov graph example.

Time Sequential Stochastic Simulation As a Markov chain is a stochastic model, another possibility is the stochastic simulation of a system. The principle is to simulate the possible events according to their probability distribution and the system reactions to these events. After a sufficient number of simulations, statistical estimations of UEs can be computed (Fig. 15.24).

The simulation needs a model of the behavior of the system for each possible failure. This model can be based on various methods:

- A computer program
- Models based on Petri nets

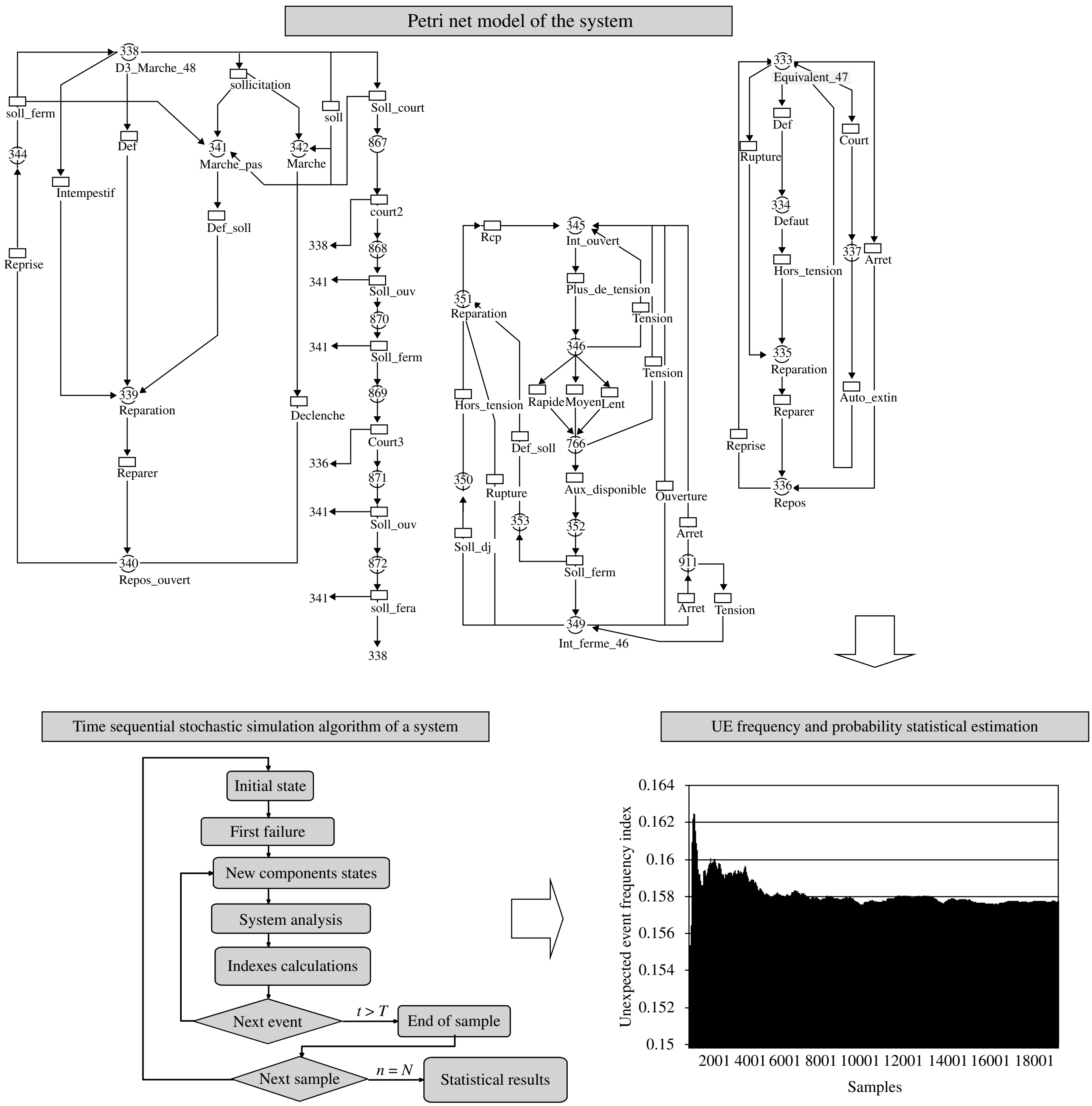


FIGURE 15.24 Time sequential stochastic simulation algorithm of a system.

15.3.3 Advantages and Drawback of Dysfunctional Tools

	Simplified FMECA	FMECA	Fault tree	Event tree	Markov graph	Stochastic simulation
Time required to build the model	++	+	–	––	–	–
Large system	++	++	++	+	–	–
Complex behavior	–	+	+	++	++	++
Results accuracy	–	+	++	++	++	++
System week points identification	++	++	++	++	–	–
Model verification	++	+	–	––	–	–
Easy to understand	++	++	+	++	–	–
Easy to use	++	++	+	++	–	–
Existing software tools	Implementable on a spreadsheet	Implementable on a spreadsheet	Many available tools for fault tree analysis	Many tools available for event tree analysis	Many available tools Markov graph analysis	Existing tools for stochastic simulation of Petri nets model
Summary	Suitable for quick analysis during architecture predesign/basic design phase Permit to identify main single-failure points	Suitable for a quick dependability analysis during detailed design phase Permit to identify ALL single-failure points and point out main contribution to failure frequency and unavailability	Best tool for a complete dependability analysis during detailed design phase Permit to identify main single-failure points Permit to identify inadequate redundancies	Not suitable for exhaustive analysis because of a too large number of initiating events to be studied	Not suitable for large systems Not adapted for quick identification of system weak points	Not suitable for large systems Not adapted for quick identification of system weak points
Remarks	—	Component failure mode level of details has to be fixed in adequacy with the expected precision of the dependability analysis and the also time and cost constraints	Some commercial tools provide contribution to unavailability but may not provide contribution to failure frequency. This can be an issue if the UE target is a failure frequency and not an unavailability Some commercial tools propose an automatic dysfunctional model generation based on a system functional model provided by the user A particular attention has to be paid to the automatic generation: in many tools, the dysfunctional model is automatically generated using many assumptions that can be not acceptable for the study			

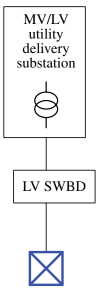
15.4 APPLICATION TO DATA CENTER DEPENDABILITY

This part is dedicated to the data center dependability assessment. Some key points are presented to perform efficient and accurate dependability assessments as well as the benefits of dependability assessment during design phase and finally some key points to manage dependability assessment and tier standard architectures [1].

15.4.1 Benefits of System Dependability Assessment

The main difficulties to achieve a reliable and cost-effective design of a data center infrastructure that includes several critical systems with interdependencies are the following:

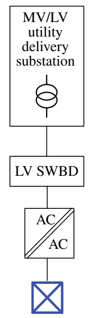
- Some equipment redundancies may not provide any significant improvement of data center reliability performances.



Main weak point is "MV utility short interruption":
⇨ Add a UPS with more than 3 min autonomy

Unexpected event : "load loss"	Failure frequency Failures/year	Unavailability Hours/year
	9.65	**3.12**

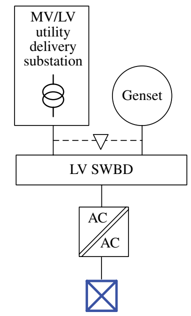
CASE 1 - Base case	Frequency	Unavailability
HV utility blackout	0%	1%
MV utility long interruption	9%	28%
MV utility short interruption	90%	9%
LV switchboard	0%	61%



Main weak point is "MV utility long interruption":
⇨ Add a Genset

Unexpected event : "load loss"	Failure frequency Failures/year	Unavailability Hours/year
	0.89	**3.27**

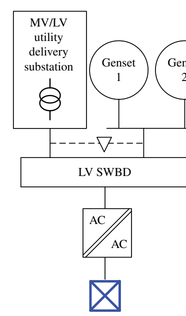
CASE 2 - UPS with 5 mn autonomy	Frequency	Unavailability
Utility blackout	1%	1%
MV utility long interruption	98%	27%
LV switchboard	0%	59%
"UPS—short-circuit on output"	0%	13%
"UPS—loss of UPS path" and "Utility short interruption"	0%	0%



Main weak point is "Genset failure and utility failure"
⇨ Add a redundant Genset

Unexpected event : "load loss"	Failure frequency Failures/year	Unavailability Hours/year
	0.041	**2.39**

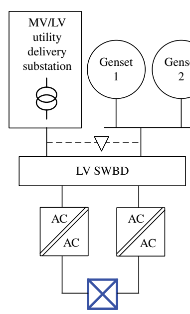
CASE 3 - With UPS and Genset	Frequency	Unavailability
"Genset failure during standby" and "Utility failure"	80%	1%
"Changeover—fail to switch" and "Utility failure"	1%	0%
"Changeover—unexpected opening of both switches"	0%	0%
"LV switchboard failure"	2%	80%
"UPS—short-circuit on output"	6%	18%
"UPS—loss of UPS path" and "Utility short interruption"	11%	0%



Main weak point is "UPS failure and utility short interruption"
⇨ Add a redundant UPS

Unexpected event : "load loss"	Failure frequency Failures/year	Unavailability Hours/year
	0.009	**2.36**

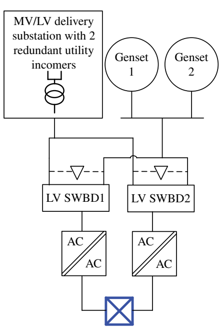
CASE 4 - UPS and 2 redundant Genset	Frequency	Unavailability
"2 Genset failures during standby" and "Utility failure"	12%	0%
"Changeover—fail to switch" and "Utility failure"	3%	0%
"Changeover—unexpected opening of both switches"	1%	0%
"LV switchboard failure"	9%	81%
"UPS—short-circuit on output"	28%	19%
"UPS—loss of UPS path" and "Utility short interruption"	46%	0%



Main weak points are "2 Genset failures and utility failure" and "LV swbd failure":
⇨ 2 redundant LV switchboards
⇨ Redundant utility incomers

Unexpected event : "load loss"	Failure frequency Failures/year	Unavailability Hours/year
	0.002	**1.92**

CASE 5 - redundant UPS distributions and Gensets	Frequency	Unavailability
"2 Genset failures during standby" and "Utility failure"	48%	0%
"Changeover—fail to switch" and "Utility failure"	13%	0%
"Changeover—unexpected opening of both switches"	4%	0%
"LV switchboard failure"	36%	100%
"Loss UPS1 and UPS2"	0%	0%



The target is reached

Unexpected event : "load loss"	Failure frequency Failures/year	Unavailability Hours/year
	0.00013	**0.003**

CASE 6 - Redundant distribution and utility incomers	Frequency	Unavailability
"2 Genset failures during standby" and "Substation failure"	98%	82%
"Loss path1 and path2"	2%	18%

- Some single-failure points are not identified.
- Reliability levels could not be harmonized for different systems (electrical power system, fuel storage, cooling system, water storage, auxiliary systems, monitoring systems).

Dependability assessment can be used at design phase, during the data operation, or during data center infrastructure refurbishment. It permits to:

- Estimate system dependability performances to prove that dependability targets are reached
- Identify and prioritize system weak points to design the right dependability improvements

During design phase, it is common to make iterations with the design team as mentioned in Figure 15.11.

Here is a simple example of the use of a reliability study during a basic design phase. The goal is to provide sufficient redundancy to reach a failure frequency <0.001/year.

15.4.2 Key Points for Data Center Dependability Assessment

15.4.2.1 Data Center Infrastructure Primary Risk Analysis

UE Definition of a General Data Center To ensure its main function that is the "IT process operation," the data center infrastructure consists in several systems that are briefly described in Figure 15.25.

A simple functional analysis is performed below:

F1: Provide IT process to customer (servers and communication with access provider).
- F1.1: Data center operations (IT, power system, and mechanical systems).
- F1.2: Provide electrical power for data center loads.
- F1.3: Provide water supply for cooling systems.
- F1.4: Provide reserve energy to emergency power plant.
- F1.5: Ensure data center process during harsh environment.
- F1.6: Ensure data center security.
- F1.7: Ensure maintenance operation

F2: Ensure people safety.

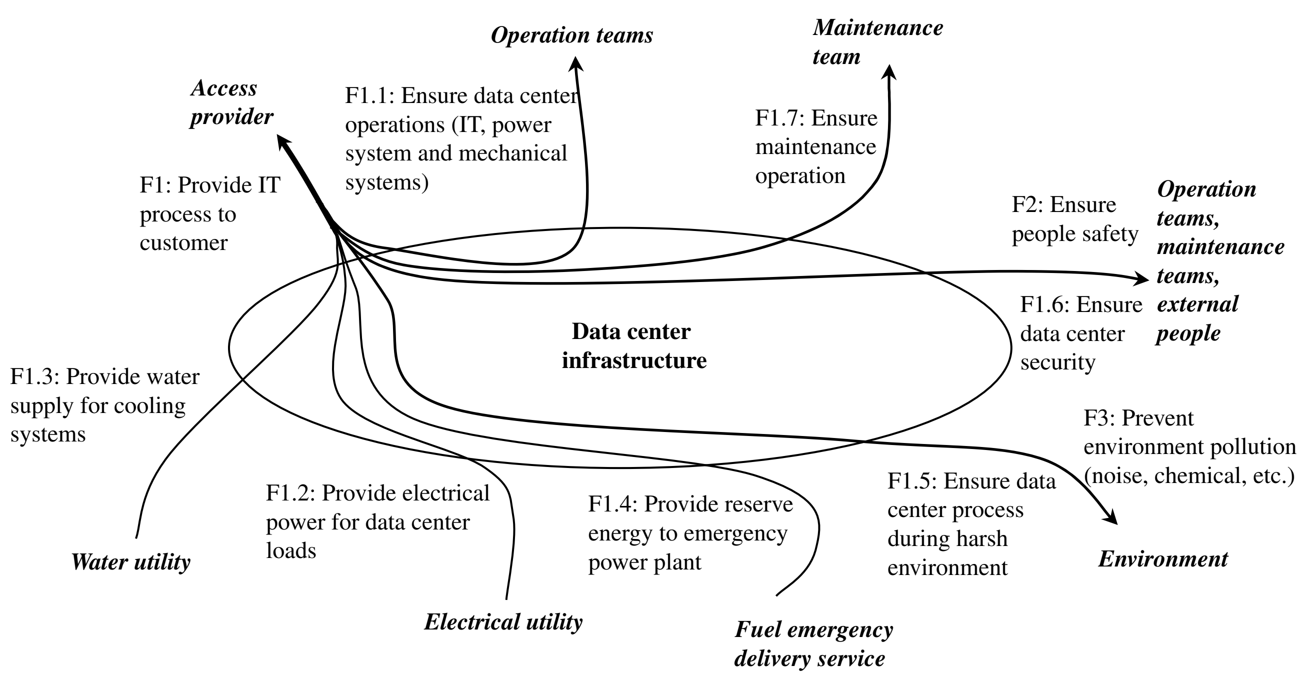
F3: Prevent environment pollution (noise, chemical).

This analysis is not exhaustive but permits to keep in mind that:

- The main functions are the "IT process operation" and to "ensure people safety" and "prevent environment pollution."
- To ensure IT process, many systems are to be taken into account.

The UEs of a "classical" data center infrastructure can be deduced from the main function degradation as mentioned in Section 15.7.1:

UE1: Loss of IT process
UE2: Safety risk
UE3: Environment pollution



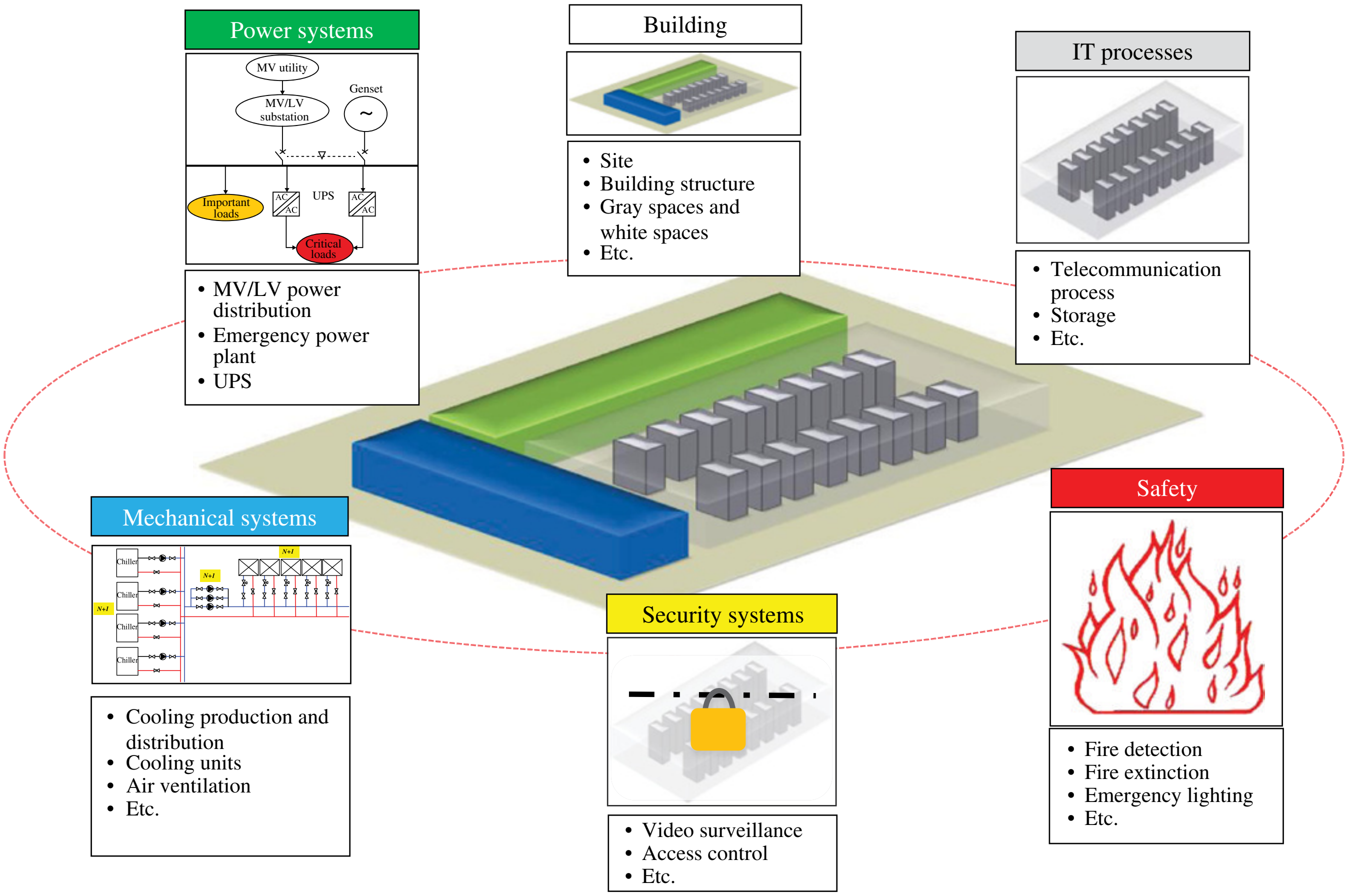


FIGURE 15.25 Data center infrastructure overview.

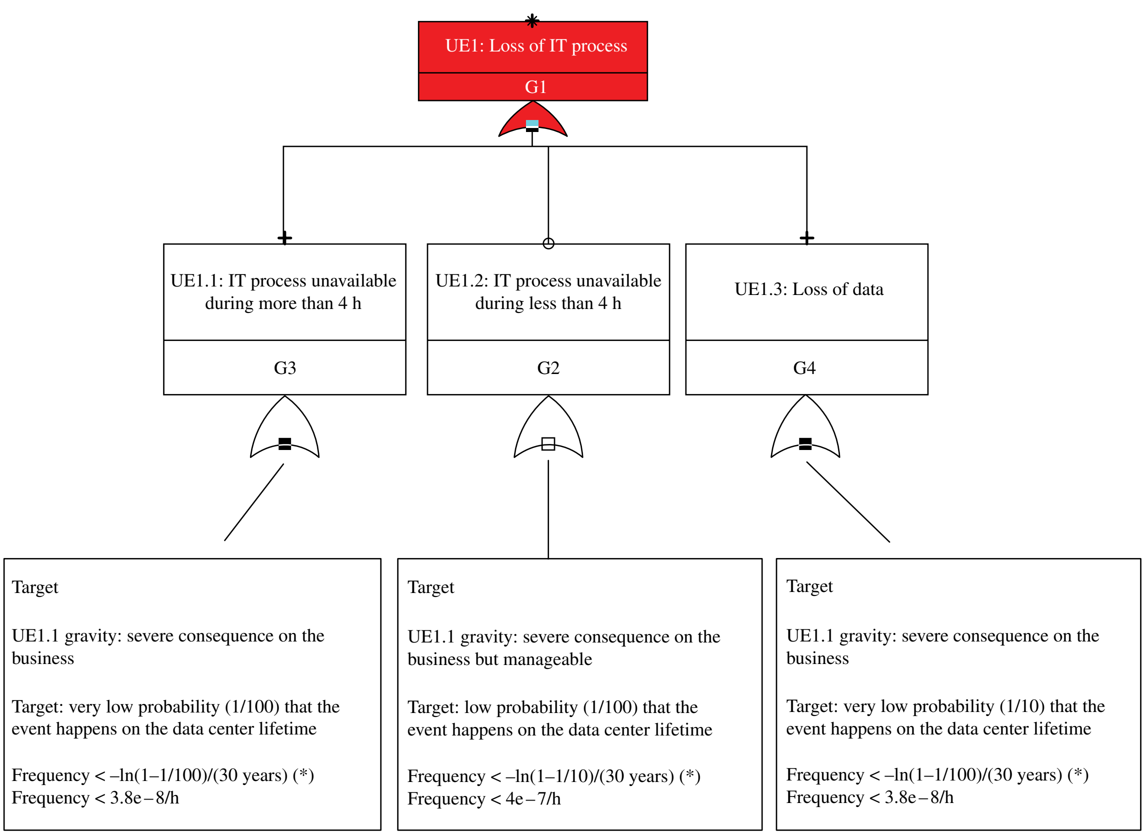


FIGURE 15.26 UE1 "loss of IT process" decomposition
(*) The frequency target is determined using the reliability indicator: $R(t) = e^{-\text{Frequency.lifetime}} >$ *probability of "experience no failure during lifetime"*

Dependability and Safety Targets Concerning the unwanted events related to safety and/or pollution (UE2 and UE3), classical targets can be determined according to standards or with qualitative target as follows:

- No single failure that affects the UE
- No failure combination (including an undetected failure and a 2nd failure) that affects the UE

IT process dependability targets are set according to UE1 gravity; as mentioned in section 0, UE1 can be decomposed in several sub-"Unwanted Events" if the consequences of UE1 can be very different. Moreover, depending on the UE definition, the target can be determined by the unavailability and/or the failure frequency.

An example of the decomposition of UE1 is given in Figure 15.26.

15.4.2.2 System Data Collection

Technical Scope When performing a dependability assessment, it is essential to take into account the whole system to ensure no useless redundancies. Figure 15.27 shows the global scope of a data center infrastructure to be studied in a dependability analysis.

Technical Data Collection Data to be collected to perform a dependability assessment are summarized in Figure 15.28.

Particular attention has to be paid to the following aspects.

Automation

Automation behaviors have to be characterized by:

- The equipment involved in the functions (sensors, logic, and actuators)
- Automation functions global overview (to be able to determine the consequences of different possible failures)

Equipment Operating Modes and Degraded Modes

The behaviors of the system depend also highly on:

- Equipment tolerance to supply interruptions (electrical supply, water supply, air conditioning/ventilation)
- Starting time of equipment or functions after a blackout

Redundancies

Equipment redundancies have to be determined according to the system architectures (electrical, cooling) and the equipment specification checking:

- Equipment limitation that could lead to partial redundancy
- Possible common cause failures linked to:
  - Interdependencies of events (design/production/installation errors, environment, human factors)
  - Auxiliary systems (power supply, water supply, SCADA systems)

Failure Detection

To determine the detection time of a failure, failure detection means have to be described including the following:

- Monitoring system
- Periodical tests (frequency, diagnostic coverage)
- Maintenance operator intervention time

Reliability Data

Equipment reliability data can be determined by several sources:

Field failure rate sources

- Manufacturer databases
- Reliability handbooks
  - IEEE Gold Book Std 493
  - EIReDA 1998 (French field experience of nuclear power plants on electrical and mechanical equipment)
  - NPRD from the Reliability Information Analysis Center (mechanical and electrical equipment)
  - EXIDA—*Safety Equipment Reliability Handbook* (Ed. 2)
- Experts with great experience of field failures that could also provide valuable information on equipment reliability

Several warnings on failure rate assumptions are highlighted below:

Field failure rate/predictive electronic studies: Theoretical electronic reliability studies performed by manufacturer are sometimes available. These failure rates are determined according to standards like IEC 62380 or MIL-HDBK-217 F. These electronic reliability studies are performed to optimize the electronic design, but the failure rate value could be pessimistic. When no field experience failure rates are available, it is commonly accepted in system dependability calculation to use theoretical values.

Failure rate validity: Failure rates are valid under certain conditions (mission profile, lifetime), which need to be highlighted and checked if in adequacy with real conditions.

Failure modes: Failure mode contributions to global failure rate are sometimes mentioned in field failure sources. When not available, a short simplified FMEA analysis of the

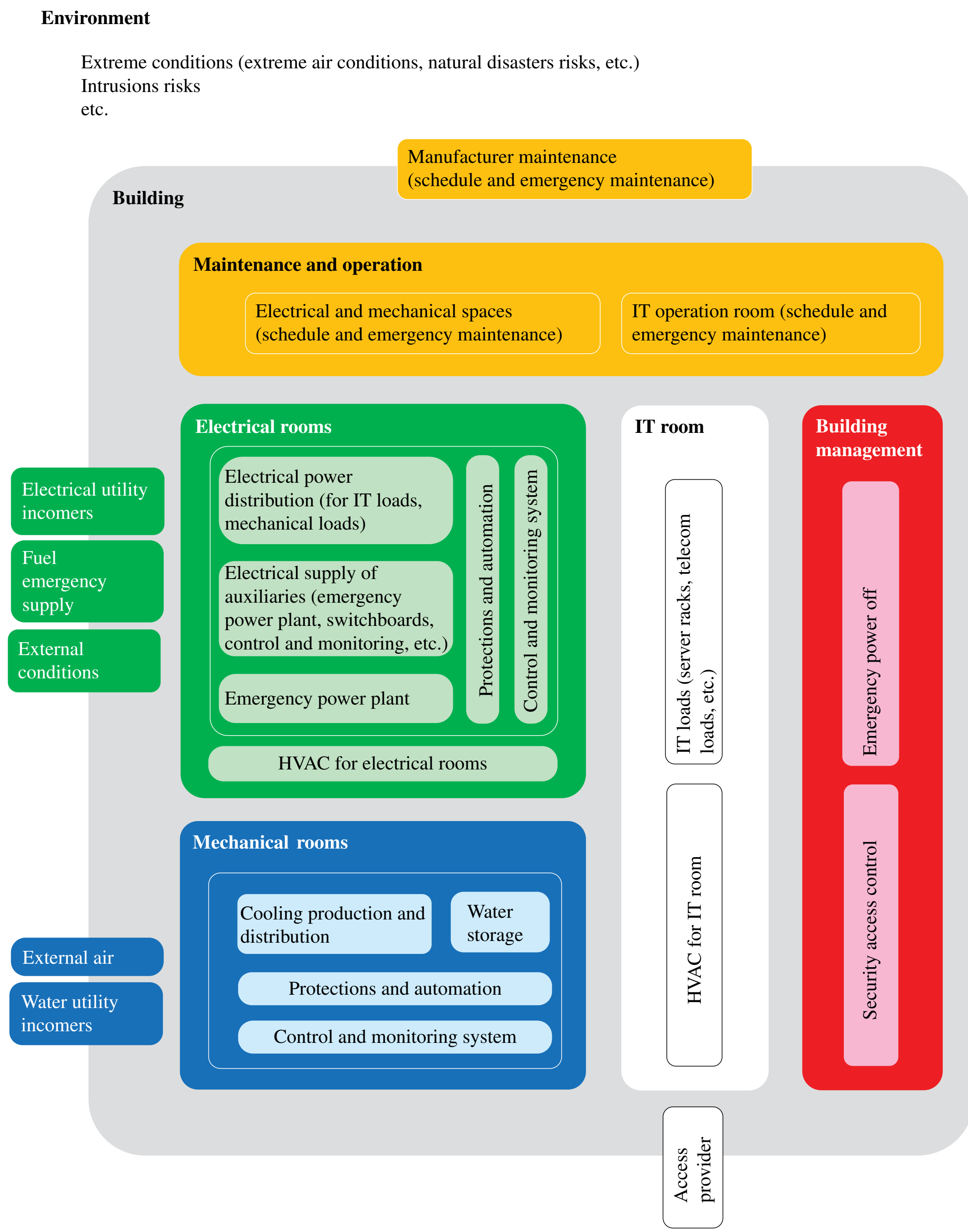


FIGURE 15.27 Overall dependability analysis technical scope for a data center infrastructure.

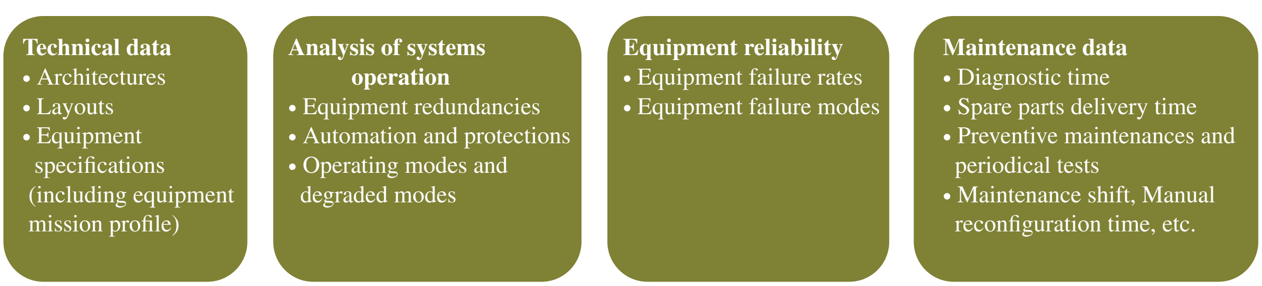


FIGURE 15.28 Reliability calculation taking into account ageing failures.

equipment based on equipment internal architecture data and field experience of the equipment can bring sufficient information to determine failure mode contributions with great accuracy.

CMF

As data center infrastructures are highly reliable/available systems, CMF identification and quantification are major key points for an accurate dependability analysis.

Maintenance Data

To determine equipment mean downtime, it is essential to take into account:

- The time for failure diagnostic
- The time for spare parts delivery

Also, programmed unavailability of equipment for preventive maintenance (verification, cleaning, preventive replacement) or for installation work during installation evolution phase needs to be taken into account.

Lack of Data

Sometimes, some data are not available. In this case, assumptions have to be made, highlighted, and verified further in the project.

15.4.2.3 Dependability Management during Project Phases During the project cycle, dependability is a main customer requirement that is applied on each project stage as mentioned in Figure 15.29.

Preliminary Outlines

During preliminary studies, the customer defines his/her needs and the global architecture of the data center with the help of a design office:

- Site selection
- Building main characteristics
- IT process definition
- IT rack rated power

At this stage, a preliminary risk analysis is performed to identify unwanted events and determine dependability targets. This step is performed by the customer, and a competency in dependability engineering is provided by the customer or an external design office.

Basic Design

Within its proposal, the contractor provides a simplified dependability analysis to confirm that its basic design reaches the dependability requirements. At this step, a

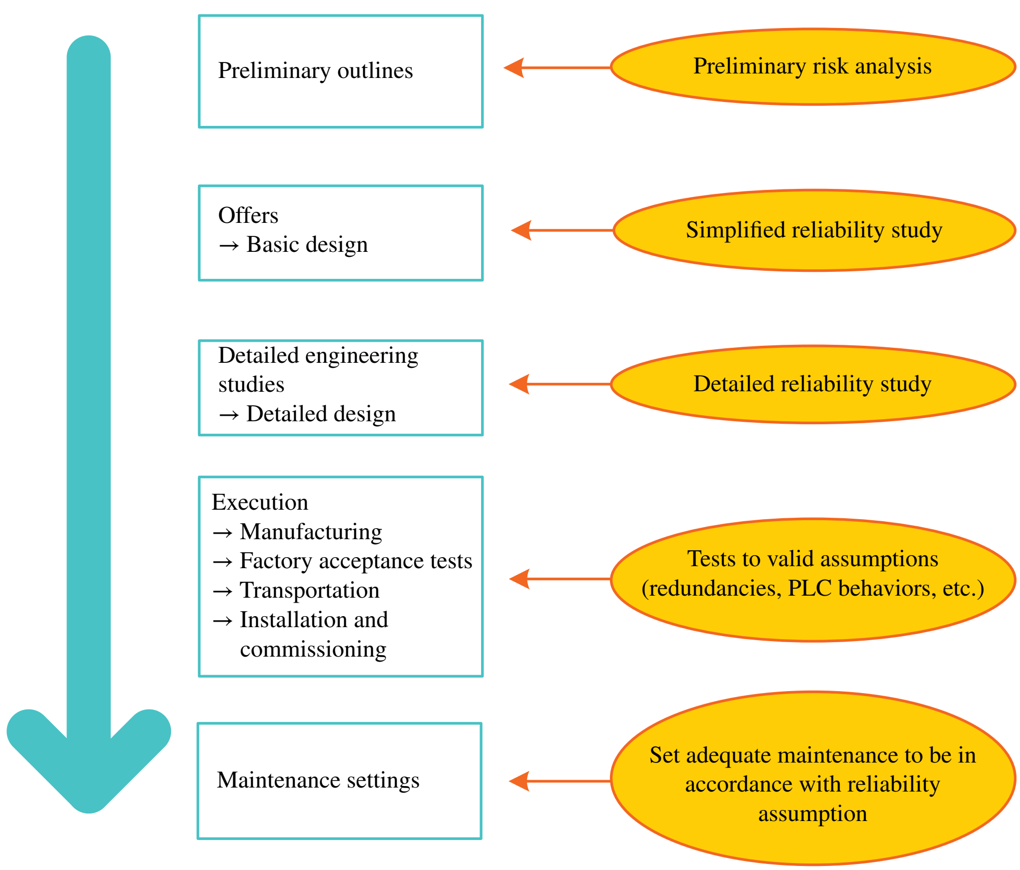


FIGURE 15.29 Dependability management during project cycle.

simplified analysis is sufficient as the design may change during detailed studies due to technical issues or customer requirement modifications. Moreover, a simplified analysis is useful when making iteration with the design team.

The simplified analysis can be limited to a simplified FMEA analysis but including the entire data center infrastructure (Fig. 15.25). In addition, few calculations of multiple-contingency analysis on nonreliable equipment (utility/power plant, chillers, pumps) will provide that redundancies are correctly designed.

Detailed Studies

During the detailed design phase, an efficient way to proceed is to provide dependability advices and checking to help the design team to make the decisions. This permits to include efficiently dependability in the design phase with multiple iterations.

When the detailed design is sufficiently defined, the dependability detailed analysis can be performed using assumptions for maintenance data.

Project Execution

Even if detailed engineering stage has been completed, some modifications may happen and lead to an update of dependability analysis. (It is a common issue that design modifications are done at this stage without checking dependability consequences.)

During installation and commissioning phases, assumptions of the dependability analysis need to be confirmed by inspection and tests to ensure that the system satisfies the dependability requirements.

On-Site Maintenance Setting

The on-site maintenance has to be set to match dependability analysis assumptions. Some iteration may happen: maintenance assumption modifications and dependability analysis update to ensure the dependability level remains unchanged.

Some difficulties of managing overall dependability analysis during project phases are summed up below:

- During basic design and detailed design phases, the contractor is responsible for the overall dependability assessment. Some difficulties may happen when the contractor need to perform the overall dependability study based on several dependability analyses of each subsystems (e.g., electrical, HVAC, and security) provided by each design offices. To minimize these difficulties, the dependability targets have to be clearly defined for each system at the beginning of the project (at design phase). Moreover, during the detailed design phase, as the different systems have many interdependencies, it is better if only one entity performs for the overall detailed dependability analysis.
- The accuracy of dependability assessment may be degraded if:
  - Some parts of the systems are not included in the system, particularly for auxiliary systems. Typically, the IT process is frequently separated from the rest of the infrastructure; this can lead to misunderstandings and architecture design problems (oversize redundancy on terminal electrical power distribution or minimize the risk to “lose at the same time the entire data center IT rooms”).
  - The reliability expert who performs the analysis is not experienced with engineering and exploitation of each system.
- A common problem is that dependability is considered during design phase but not after project execution and operation phase. The customer shall keep its dependability analysis updated during all the phases of its installation.
- As data center infrastructures are intended to be upgraded several times during its lifetime, overall dependability analysis needs to be updated for each phase to ensure dependability level during all phases.

15.4.3 TIA Level Classification and Dependability Assessment

A basic description of level classification according to TIA-942 standard is given below:

Level I: “Basic capacity”—no redundancy required
Level II: “Redundant capacity components”—redundancy on nonreliable equipment
Level III: “Concurrently maintainable”—Level II requirements + each equipment can be removed and repaired without a data center blackout
Level IV: “Fault tolerant”—Level III + fault-tolerant architecture (no single-failure point)
For more information, see Ref. 1

According to the data center IT business criticality, the level classification is a powerful tool:

- To set adequate level according to the customer’s business
- To set equipment redundancy for infrastructure design (electrical, mechanical, building, etc.)

Benefits of Level classification

Simple classification

- ⇨ Understandable and accessible to everyone
- ⇨ Powerful to perform a quick assessment at design phase

Take into account all critical systems (electrical power system, HVAC systems, critical auxiliaries, …)

Good levels of dependability classification that provides a good frame of references

Disadvantages of Level classification

Pessimistic assumption on utility dependability that leads to oversizing of some redundancies

The huge gap between Level III and Level IV sometimes leads to designing Level III infrastructure with additional redundancies without matching Tier IV requirements

In some cases, "N+1" design of some equipment is not enough, but the classification does not take it into account

Emergency and preventive maintenance are not taken into account

All equipment failure modes are not systematically taken into account as well as the failure detection

During project phase, an efficient procedure is described below:

- During preliminary study, the data center owner expresses the criticality of its IT business and then is able to set the adequate tier level according to the tier standard.
- During basic design phase, the contractor provides a simplified FMECA to prove that the tier requirements are reached.
- During detailed engineering studies, the contractor provides a detailed FMECA study to prove that tier requirements are satisfied.

REFERENCE

[1] DiMinico C, Jew J. 2005. ANSI TIA-942: Telecommunications infrastructure standard for data centers. Available at www.tiaonline.org. Accessed on June 24, 2014.

FURTHER READING

Billinton R, Alan RN. *Reliability Evaluation of Power Systems*. 2nd ed. New York: Plenum Press; 1994.

Cabau E. Introduction à la Sûreté de Fonctionnement. Cahier Technique Schneider Electric nr 144. Grenoble: Schneider Electric; June 1999.

International Electrotechnical Commission (IEC). CEI 300-3-1: Gestion de la sûreté de fonctionnement, technique d'analyse de la sûreté de fonctionnement—Guide méthodologique. 2nd ed. Geneva: IEC; 2003.

IEC. CEI-IEC-61165: Application des techniques de Markov. 2nd ed. Geneva: IEC; 2006.

IEC. IEC 60812: Analysis techniques for system reliability—procedure for failure mode and effects analysis (FMEA). 2nd ed. Geneva: IEC; 2006.

Logiaco S. Electrical installation dependability studies. Cahier Technique Schneider Electric nr 184. Grenoble: Schneider Electric; December 1996.

Lonchampt A, Gatine G. High availability electrical power distribution. Cahier Technique Schneider Electric nr 148. Grenoble: Schneider Electric; 1990.

Villemeur A. *Sûreté de Fonctionnement des systèmes industriels*. Paris: Eyrolles; 1988.

16

PARTICULATE AND GASEOUS CONTAMINATION IN DATA CENTERS

Taewon Han

Rutgers, The State University of New Jersey, New Brunswick, NJ, USA

16.1 INTRODUCTION

Are IT equipment failure rates higher for data centers cooled with outside air compared to "closed" data centers? Data center cooling efficiency has a large impact on overall electricity use. A simple cooling solution is to implement free cooling, using outside air to provide direct cooling of IT equipment. However, many owners and operators are hesitant to use this because they are concerned that the use of outside air for cooling increases the risk of IT equipment failure due to airborne contamination, either particulate or gaseous contamination. A number of research studies have investigated the corrosion phenomena of IT hardware components at varied conditions [1–8]. IT environment control in data centers is subject to many regulations and guidelines such as ISO 14644-1 [9], ISA-71.04 [10], and ASHRAE papers [11–14]. In fact, in greater part of these papers, the authors have repeatedly published information about specific technical issues (e.g., corrosion on circuit boards) or have tried to help designers and operators manage their data centers more effectively. However, finding actual hardware failure rates comparing outside air-cooled to traditional closed data centers for limited test conditions (e.g., temperature, humidity, gas concentration, type, mixture in air contaminants) is difficult. Finding the root cause of IT equipment failure is challenging. The Restriction of Hazardous Substances (RoHS) directive adopted in 2006 by the European Union has led to the use of silver-based materials instead of lead-based materials for manufacturing printed wiring boards (PWB) in IT equipment. This chapter will briefly introduce standards, airborne contaminants (particulate and gaseous), and, most of all, measurement methods that are described along with a study (i.e., survey type study in real operating data centers) on how the conventional technologies should be approached to improve air quality monitoring by a simple economical method.

16.2 STANDARDS AND GUIDELINES

Most data centers adhere to the following standards or guidelines: ISO 14644-1 [9] and ANSI/ISA-71.04 [10]. A worldwide standard, the ISO-14644 consists of eight parts; especially, the part 1 (i.e., ISO 14644-1) covers the classification of air cleanliness in clean rooms and associated controlled environments exclusively in terms of concentration of airborne particles. ISO 14644-1 mainly specifies the quantity and size of particulates (Table 16.1) as well as a measurement methodology. However, the mass of particulates wasn't shown in the standard in spite of the need for total mass limitation for data center. Air cleanliness in data centers often complies with ISO Class 8 [15], which is simply achieved with MERV 8 or 11 or 13 filters depending on the cooling method (ASHRAE Standard and ASHRAE Book).

While ISO 14644-1 addresses particulate contaminants, ANSI/ISA-71.04 addresses gaseous composition environmental limits in 1985. This method describes how to determine the gaseous corrosivity of a data center environment. It is termed "reactive monitoring." This reactivity monitoring is defined by placing metal strips into the environment. They are exposed for a period of time and then analyzed to

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

TABLE 16.1 Air cleanliness classification as a function of particle size

Maximum number concentration of airborne particles						
Particles/m^3						
Class	Particle size (μm)					
	>0.1	>0.2	>0.3	>0.5	>1.0	>5.0
1	10	2				
2	100	24	10	4		
3	1,000	237	102	35	8	
4	10,000	2,370	1,020	352	83	
5	100,000	23,700	10,200	3,520	832	29
6	1,000,000	237,000	102,000	35,200	8,320	293
7				352,000	83,200	2,930
8				3,520,000	832,000	29,300
9					8,320,000	293,000

Source: ISO 14644-1, 1999 [9].

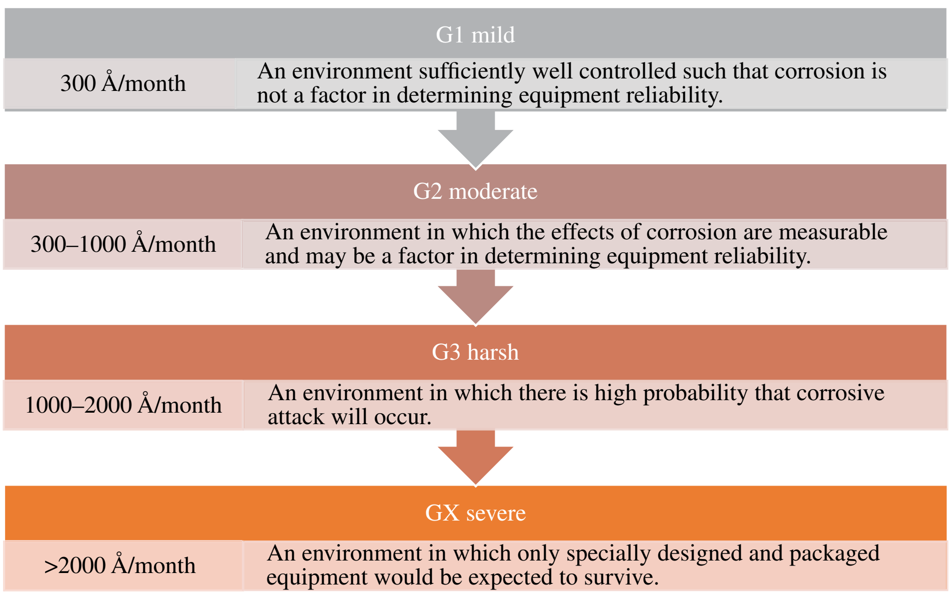


FIGURE 16.1 Four levels of gaseous corrosivity established by the ANSI/ISA-71.04 standard (in angstroms, Å). From Ref. [10].

determine the thickness (in angstroms, Å) of the corrosion films on the metal strips. This analysis method indicates the classification of the total amount of corrosion according to the thicknesses of individual corrosion films. Figure 16.1 shows gaseous corrosivity levels for copper coupons. The corrosivity levels vary considerably depending on the combination of corrosive gases (e.g., H_2S, SO_2, SO_3, Cl_2, NO_x, HF, NH_3, and O_3). The levels may also be affected by other synergistic effects such as temperature and humidity. Based on this standard and corrosion monitoring technology, much research has been performed in the area of environmental classification. However, one of the critical limitations to date was that the corrosion monitoring and assessment is only available for copper.

ASHRAE's Guideline [14] initially was titled "Particulate and gaseous contamination guidelines for data centers" and was prepared by the ASHRAE TC 9.9 committee's members in 2009. It was then updated based on an AHRAE survey of the air quality in data centers and on lessons learned in cleaning the air in contaminated data centers in 2011. The TC 9.9 committee's members consisted of 12 IT equipment manufacturers: AMD, Cisco, Cray, Dell, EMC, Hitachi, HP, IBM, Intel, Oracle, Seagate, and SGI. ASHRAE recommends that data centers be kept clean to ISO Class 8 for particulate (dust) contamination, which may be achieved by MERV 8 filters [16]. For data centers utilizing free cooling, MERV 11 or MERV 13 filters may be preferred to achieve ISO Class 8 level [13].

TABLE 16.2 Temperature and relative humidity operating conditions in data center

	2004 version	2008/2011 version
Low-end temperature	20°C (68°F)	18°C (64.4°F)
High-end temperature	25°C (77°F)	27°C (80.6°F)
Low-end humidity	40% RH	5.5°C (41.9°F) dew point
High-end humidity	55% RH	60% RH and 15°C (59°F) dew point

From Refs. 11 and 14.

In addition, gas-phase filtration systems are recommended by ASHRAE for data centers with higher gaseous contamination levels. In 2008 and 2011, expanded thermal guidelines for data processing environments increasing the range of temperature and relative humidity (Table 16.2) were prepared.

16.3 AIRBORNE CONTAMINATION

Airborne contaminants can be split into two distinct categories: particulate and gaseous. They can penetrate into the data centers whether they are of traditional closed design or utilize free air cooling, and in large concentrations, the contaminants may degrade electronic components. However, in practice, it is very difficult to find a direct relationship between airborne contamination and hardware failure due to many factors. There are no publically available statistics on IT equipment failure rates due to contamination. It is also unknown how many data centers in the United States or in the world have experienced IT equipment failure.

Particulate contaminants can be effectively captured by a proper filtration system (e.g., MERV 8, or MERV 11 or 13 filters); thus, contamination due to particulates is not considered to be a problem at most data centers [14, 17, 18]. Minimum efficiency reporting value (MERV) rating is a measurement scale designed in 1987 by ASHRAE to rate the effectiveness of air filters (Table 16.3). ASHRAE standard 52.2 provides the procedure for measuring filter efficiency as a function of particle size (e.g., 12 particle size ranges). These 12 ranges are grouped into three ranges for rating purposes: E_1 (0.3–1.0 μm), E_2 (1.0–3.0 μm), and E_3 (3.0–10.0 μm).

The effects of gaseous contaminants such as hydrogen sulfide (H_2S), sulfur dioxide (SO_2), chlorine (Cl_2), ozone (O_3), and nitrogen dioxide (NO_2) are not as well understood. Gaseous contamination is often difficult to diagnose due to the interaction of gases, humidity, temperature, and other complicated conditions [19]. Determining the corrosion rate as a function of gaseous concentration in actual environments is generally difficult because many environments contain a complex mixture of contaminants that interact differently with the corrosive action of individual gas species. For example, sulfur dioxide and hydrogen sulfide alone are not very corrosive to silver or to copper, but can be very corrosive when combined with nitrogen dioxide or ozone [20, 21]. Silver is extremely sensitive to Cl_2 [17].

TABLE 16.3 Minimum efficiency reporting values (MERV), ASHRAE standard 52.2

MERV	E_1	E_2	E_3
	0.3–1.0 μm	1.0–3.0 μm	3.0–10.0 μm
8	—	—	70–84.9%
11	—	65–79.9%	≥85%
13	≤75%	≥90%	≥90%

From ASHRAE standard 52.2.

16.4 A CONVENTIONAL SOLUTION

As previously mentioned, gaseous contamination is more of a concern for data center operators in general. *Why do we worry about gaseous contamination?* The most common damage of metallic components in the IT industry has been known to be the result of copper or silver corrosion on circuit boards from the effects of gaseous pollutants [4, 19, 20–24]. IT failures in some data centers, anecdotally believed to be caused by gaseous corrosion, have been observed by IT manufacturers; however, these data are not generally made public. In response to this concern, the electronics industry has adopted a method to measure corrosion rates of metallic material affected by the environment. This method, common reactivity monitoring for gaseous contaminants, is referred to as the corrosion classification coupon (CCC) method and is described in ANSI/ISA-71.04.

16.4.1 CCC Measurement

This method is a convenient and common way to determine the gaseous corrosivity of a data center. Use of metal coupons is the best known and simplest of all corrosion monitoring techniques. The method exposes a copper coupon to the environment for 1 month and analyzes the corrosion product thickness using coulometric reduction to classify the environment into one of four severity levels (ISA-71.04): G1 (mild, <300 Å/month; corrosion is not a factor in determining equipment reliability), G2 (moderate, 300–1000 Å/month; corrosion may be a factor in determining equipment reliability), G3 (harsh, 1000–2000 Å/month; high probability that corrosive attack will occur), and GX (severe, >2000 Å/month; only specially designed and packaged equipment would be expected to survive). But the use of copper coupons alone has some limitations including the following: copper is not sensitive to chlorine, a particularly corrosive contaminant to many metals, and copper corrosion may be overly sensitive to relative humidity [19]. It is now common practice to include silver coupons (G1, mild, <200 Å/month) along with copper coupons to gain greater insight into the chemistry of the corrosive gases in the environment [14].

16.4.2 Application of the Conventional CCC Method

This is a description of an exploratory study using the conventional coupon method [25]. The coupon placement strategy included a total of 19 data centers (California, Texas, Illinois, New Jersey, Georgia, North Carolina, and Massachusetts) located in the United States and 2 in Bangalore, India (Fig. 16.2). The objective of the study was to investigate the following questions: (i) What are the approximate statistical distributions of copper and silver corrosion rates in data centers across the United States? (ii) Are corrosion coupon measurements repeatable? (iii) What is the relationship between copper and silver corrosion measurements? (iv) Are corrosion rates higher for outside air-cooled data centers compared to “closed” data centers? (v) Are corrosion measurements related to IT equipment failure rates?

As a means of environmental corrosion monitoring, a CCC consists of a combination of copper and silver metal strips that have been used to measure the severity level of corrosive gases (Fig. 16.3). The method involves exposing a coupon specimen to the environment for a given duration (e.g., 30 days). In the study, following exposure, the specimens were analyzed using cathodic/electrolytic reduction. The magnitude of corrosion, or corrosivity, was quantified by corrosion growth rate of angstroms (Å)/30 days. To minimize any background corrosion during transport, coupons were placed in a ziplock plastic bag with a special material that acts as a scavenger for any ambient contamination sealed inside the bag with the coupon.

Each coupon set (copper and silver) is attached to a Plexiglas support (~4 in. ×3 in. ×1/4 in.), and the coupon number, date, and location information are recorded on the attached label. The coupons at each data center were placed in three different placement categories (e.g., Fig. 16.4): (1) outside air at the building entry point prior to filtering or conditioning, (2) inside duct work or plenums feeding the data center room, and (3) inside the data center room. All data were collected from August to November 2010.

The results of the study indicated the following:

1. Copper and silver coupon corrosion rates were generally low in the United States compared to rates that are thought to be problematic.
2. Measurements within the same data center frequently differed by a factor of 2 or more.
3. Silver corrosion rates were poorly correlated with copper corrosion rates.
4. Copper corrosion rates were not higher for outside air-cooled data centers than for “closed” data centers. Silver corrosion rates were not higher in most air-cooled data centers, but may be higher in some.
5. Data centers with relatively high silver corrosion rates reported no unusual equipment failure rates, although detailed analysis of this question is not possible because of the low number of data centers with high silver rates encountered. The poor repeatability of the

FIGURE 16.2 Coupon placed locations [10]: United States (14 locations, 19 data centers: San Francisco, California (1); Dublin, California (1); Silicon Valley, California (5); Rocklin, California (1); Fresno, California (1); Los Angeles, California (1); Phoenix, Arizona (1); Chicago, Illinois (1); Boston, Massachusetts (1); Research Triangle Park, North Carolina (1); Richardson, Texas (1); Dallas, Texas (1); Atlanta Georgia (1); Piscataway, New Jersey (2)) and Bangalore, India (two data centers).

FIGURE 16.3 Photo of the CCC. The standard method for analyzing corrosion coupon is called cathodic/electrolytic reduction. The thickness of the corrosion film is determined by a laboratory analysis. The results of the report included photograph of returned coupon strips, ISA Environmental Classification, and film thickness/30 days.

(a)

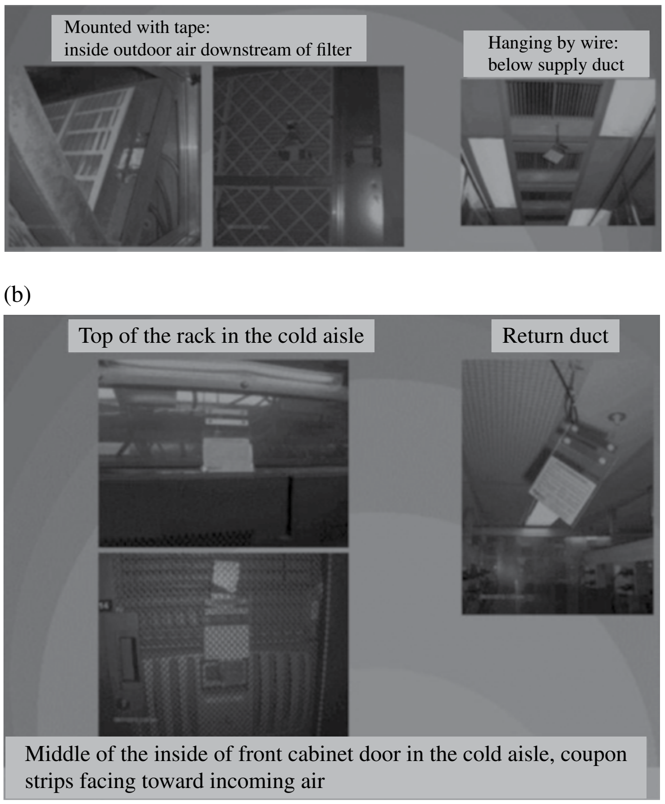


FIGURE 16.4 Photos of different coupon placements: (a) examples of inside outdoor air and (b) example of inside indoor air [10].

measurements and the apparent lack of an elevated equipment failure rate in the only facility with a high measured silver corrosion rate suggest that corrosion coupon measurements may not be useful for predicting equipment failure rates. Most facilities in the study, including outside air-cooled facilities, did not have elevated corrosion rates, so even if measured corrosion rates did correlate with failure rates, the use of outside air cooling did not seem problematic in the United States.

As already explained, copper and silver coupons used together may provide a more complete risk assessment for classifying data center corrosivity. A key question is whether data centers using large amounts of outside air for cooling have more risk of IT equipment failures. There was no evidence that a high corrosion rate measured by corrosivity monitoring implied a high equipment failure rate. Because quantitative data on failure rates were not readily available, a correlation between reactivity monitoring coupon measurements and IT equipment failure rates could not be determined. These corrosion rates are not thought to be problematic per the ANSI/ISA-71.04-1985 guidelines. Copper and silver corrosion measurements can differ substantially, which is not surprising since these elements react differently to corrosive gases. There is some correlation between these measurements, suggesting that the coupons are in fact measuring something real about the corrosivity of the environment, in spite of the substantial measurement errors.

16.5 CONCLUSIONS AND FUTURE TRENDS

There is considerable concern over the use of outside air for cooling data centers, but industry experts disagree on the severity of the concern. Fortunately, most data centers in the United States are located in regions with relatively clean environments [25]. Thus, there currently is no public information on failure rates of IT equipment due to contamination in the United States. There is no publicly available data linking use of outside air cooling with IT equipment failure rates. Anecdotal evidence suggests that equipment failures have occurred inside data centers that were closed.

In the future, (i) basic research on corrosion caused by gaseous contamination is still needed to attempt to find correlation to IT equipment component failure rates. (ii) A survey-type study is suggested to better understand the corrosivity monitoring method and continuously monitor gaseous contamination with a large number of coupons placed in airstreams located after the prefilter in the data center to measure conditions at the inlets of IT equipment. Once failure mechanisms are understood, a study of potential remedies should be undertaken.

ACKNOWLEDGMENT

The author acknowledges the organizations ISO, ANSI/ISA, and ASHRAE for providing the following standards or guidelines that served as major references for completing this chapter: ISO 14644-1 [9], ANSI/ISA-71.04 [10], and ASHRAE's Guideline [14]. The author also thanks Henry

Coles, William Tschudi, and Ashok J. Gadgil of Lawrence Berkeley National Laboratory for conducting the exploratory study and the staffs at the data center sites for participating in the study. This project was funded by the Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE) and the California Energy Commission (CEC) Public Interest Energy Research (PIER) Program.

REFERENCES

[1] Cullen DP, O'Brien G. Implementation of Immersion Silver PCB Surface Finish in Compliance with Underwriters Laboratories. Bannockburn: IPC Printed Circuits Expo; 2004.

[2] Hillman CA, Arnold J, Binfield S, Seppi J. Silver and Sulfur: Case Studies, Physics and Possible Solutions. College Park: SMTA International; 2007.

[3] Mazurkiewicz P. Accelerated corrosion of PCBs due to high levels of reduced sulfur gases in industrial environments. Proceedings of the 32nd ISTFA; November 12–16, 2006; Austin.

[4] Reid M, Punch J, Ryan C, Franey J, Derkits GE, Reents WD, Garfias LF. The corrosion of electronic resistors. IEEE Trans Comp Pack Technol 2007;30(4):666–672.

[5] Sahu AK. Present scenario of municipal solid waste dumping grounds in India. International Conference on Sustainable Solid Waste Management; September 5–7, 2007; Chennai, India.

[6] Schueller R. *Creep Corrosion of Lead-Free Printed Circuit Boards in High Sulfur Environments*. Edina: SMTA International; 2007.

[7] Veale R. *Reliability of PCB Alternate Surface Finishes in a Harsh Industrial Environment*. Edina: SMTA International; 2005.

[8] Xu C, Fleming D, Demirkan K, Derkits G, Franey J, Reents W. Corrosion resistance of PWB final finishes, Alcatel-Lucent, APEX; 2007; Los Angles.

[9] ISO. ISO 14644-1, *Cleanrooms Associated Controlled Environments—Part 1: Classification of Air Cleanliness*. Geneva: International Organization for Standardization; 1999.

[10] ISA. Environmental Conditions for Process Measurement and Control Systems: Airborne Contaminants. The Research Triangle Park: Instrumentation, Systems, and Automation Society; 1985. ANSI/ISA 71.04-1985.

[11] ASHRAE. 2008 ASHRAE environmental guidelines for Datacom equipment—expanding the recommended environmental envelope. Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers; 2008. TC 9.9 white paper.

[12] ASHRAE. Particulate and gaseous contamination guidelines for data centers. Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers; 2009a. TC 9.9 white paper.

[13] ASHRAE. *Particulate and Gaseous Contaminants in Datacom Environments*. Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc.; 2009b.

[14] ASHRAE. 2011 thermal guidelines for data processing environments—expanded data center classes and usage guidance. Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers; 2011. TC 9.9 white paper.

[15] Ortiz S. Data center cleaning services. Processor Soc 2006; 28(14):4.

[16] ASHRAE. ANSI/ASHRAE Standard 127-2007, *Method of Testing for Rating Computer and Data Processing Room Unitary Air Conditioners*. Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers; 2007.

[17] Shehabi A, Horvath A, Tschudi W, Gadgil AJ, Nazaroff WW. Particle concentrations in data centers. Atmos Environ 2007; 42:5978–5990.

[18] Shehabi A, Ganguly S, Gundel LA, Horvath A, Kirchstetter TW, Lunden MM, Tschudi W, Gadgil AJ, Nazaroff WW. Can combining economizers with improved 19 filtration save energy and protect equipment in data centers? Build Environ 2010;45:718–726.

[19] Rice DW, Peterson P, Rigby EB, Phipps PBP, Cappell RJ, Tremoureux R. Atmospheric corrosion of copper and silver. Electrochem Soc 1981;128(2):275–284.

[20] John WO. Corrosion-induced degradation of microelectronic devices. Semicond Sci Technol 1996;11:155–162.

[21] Volpe L. Environmental factors in indoor corrosion of metals. Technical report. Armonk: IBM International; 1989.

[22] Lopes BG, Valdez SB, Zlatev KR, Flores PJ, Carrillo BM, Schorr WM. Corrosion of metals at indoor conditions in the electronics manufacturing industry. Anti-Corros Meth Mater 2007;54(6):354–359.

[23] Vargas OL, Valdez SB, Veleva ML, Zlatev KR, Schorr WM, Terrazas GJ. The corrosion of silver in indoor conditions of an assembly process in the microelectronics industry. Anti-Corros Meth Mater 2009;56(4):218–225.

[24] Veleva L, Valdez B, Lopez G, Vargas L, Flores J. Atmospheric corrosion of electro-electronics metals in urban desert simulated indoor environment. Corros Eng Sci Technol 2005;43(2):149–155.

[25] Coles HC, Han T, Price PN, Gadgil AJ, Tschudi WF. Most U.S. Data centers can use outside air for cooling without causing failures, Part of the DOE Roadmap for compressor less cooling in data centers. Lawrence Berkeley National Laboratory Technical Publications; 2011.

FURTHER READING

Bennett HE, Peck RL, Burge DK, Bennet JM. Formation and growth of tarnish on evaporated silver films. J Appl Phys 1999;40(8):3351–3360.

Comizzoli RB, Frankenthal RP, Lobnig RE, Peins GA, Psato-Kelty LA, Siconolfi DJ, Sinclair JD. Corrosion of electronic materials and devices by submicron atmospheric particles. Electrochem Soc Interf 1993;2(3):26–34.

Miller SK. Whiskers in the data center. Processor 2007;29(30):1–24.

PURAFIL. Corrosion control for mission critical facilities: design guidelines for the assessment, control and testing of gaseous contamination in data centers and server rooms; 2011. Technical brochure TB1800-02.

17

COMPUTATIONAL FLUID DYNAMICS APPLICATIONS IN DATA CENTERS

Mark Seymour
Future Facilities Limited, London, UK

17.1 INTRODUCTION

One of the principal issues in maintaining the very high availability required for mission-critical facility data centers is how to deliver cooling effectively and efficiently to all the equipment, wherever it is in the room.

The significant power densities in modern data centers mean that this is a substantial task and justifies significant focus if it is to be implemented in such a way that achieves the cooling objective at low cost and without significantly interfering with operational objectives. People often forget that the primary purpose of the cooling system is to cool the electronics; historically, data centers were simply treated in the same way as other occupied spaces.

Figure 17.1 shows that to cool the electronic component it is important to take responsibility for configuration/design of cooling for the IT Equipment itself, how it is cooled in any rack/cabinet, and how those cabinets are cooled in the data hall. The broken line represents the break in ownership of the problem; at the electronics scale, the manufacturer is responsible; at the room scale, the facility manager is responsible. There is an obvious danger that the rack/cabinet configuration and cooling falls between the two.

Modern environmentally considerate designs are often very data center specific. They use fluids—most commonly air but sometimes liquid—to deliver the cooling to the IT and carry away the heat from it. This is not only a challenge for the initial design but also for the ongoing management because unlike a box of electronics where the internal components remain pretty much fixed for the life of the system, what is installed in many data centers changes frequently, in some cases even on a daily basis.

Where air is the medium employed for cooling in the rooms of the data center facility holding the IT equipment (commonly known as data halls), it is often beneficial to analyze and thus optimize the cooling design and IT equipment configuration to make best use of the cooling available. This helps avoid hot spots that otherwise tend to build up as the data center evolves over time. Analysis will also be useful in the design and configuration of liquid-cooled systems, but this will commonly be a task allocated to a cooling professional and so this is not the primary focus of this topic.

One of the key issues about a data hall is that almost every hall is unique. For one reason or another and unlike typical electronics cooling problems (where the equipment configuration is defined and fixed during design), the design of a data hall varies from one installation to another. What is more, a hall will continue to vary over time as new equipment items are deployed and older equipment items are removed, moved, or upgraded. This creates a scenario where a "one size" solution does not fit all.

Because the cooling is achieved via a fluid cooling medium, the only theoretical technique that can predict the complex performance of the cooling system—the cool air delivery and hot air scavenging—is a technique known as Computational Fluid Dynamics, or CFD for short.

17.2 FUNDAMENTALS OF CFD

CFD is, as the name suggests, the use of computers (using numerical methods) to analyze the likely behavior of a fluid (liquid or gas) as a result of the surrounding environment. It takes into account the stimuli that promote or restrict

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

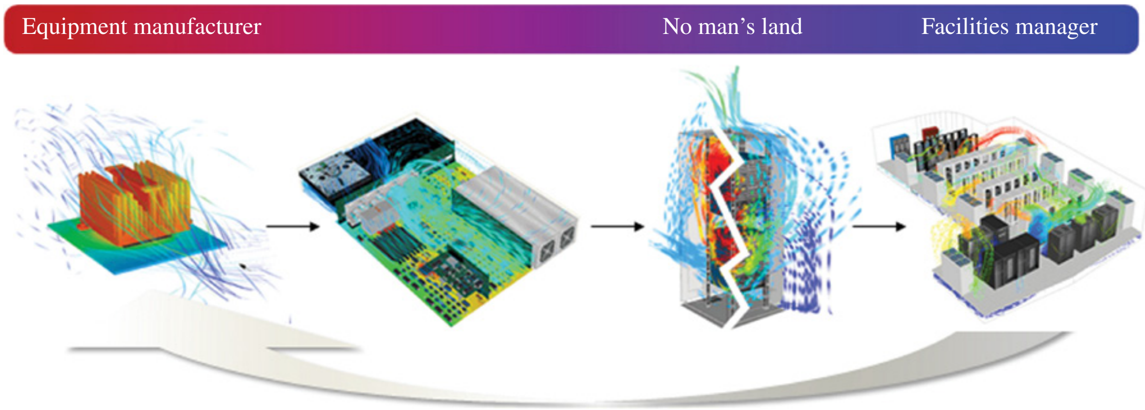


FIGURE 17.1 Data center cooling is required at all scales to effectively cool equipment heat dissipation. Courtesy of Future Facilities. Please visit companion website to access the color version of this figure.

movement of the fluid within and around geometrical objects in the region for which the analysis is being made.

The term "fluid" is used because the techniques can equally be applied to liquids or gases. For example, an analysis could be made for both the chilled water flow inside the chilled water circuits of a finned heat exchanger and for the airflow between the fins around the chilled water circuits of the heat exchanger. In fact, the analysis is not limited to the flow of only a *single* fluid, liquid or gas, but can in principle analyze both the chilled water flow (liquid) and airflow (gas), simultaneously accounting for the heat transfer between the two fluids (in this example, water and air). It can also account for the impact of temperature variations in the two fluids and indeed the solids comprising the heat exchanger itself.

Already from this simple example you will begin to realize that the CFD methodology is, in fact, very general and can be applied to many applications, not just cooling IT equipment in a data hall (largely using air as the fluid). In fact, the techniques are even more powerful than indicated thus far: they can be applied not just to segregated fluids but to fluid mixtures as well. They can allow for changes in state (e.g., evaporation and condensation of water), known as two-phase problems. Additionally, they can be applied to time-dependent variations in the fluid flow, rather than just the conditions at a snapshot in time as though nothing ever changes. The latter is commonly known as a "steady-state analysis."

It is, therefore, perhaps appropriate to outline some of the fundamental principles of CFD before considering, in somewhat more detail, the uses and application of CFD to data centers.

17.2.1 Basic Principles

When performing a CFD analysis, one uses a computer program to solve a set of equations commonly known as the Navier–Stokes equations. The equations mathematically define the conservation laws for mass, momentum, and energy and were first derived by Claude Louis Marie Navier in 1827. The equations were in fact independently derived by several people over the next two decades, and in 1845, George Gabriel Stokes published his derivation (from a different perspective) of the equations. Hence, they subsequently became known as the Navier–Stokes equations. However, for most practical applications, and in fact all but the simplest of scenarios, there is no analytic solution to these simultaneous equations.

In order to solve the equations, it is necessary to divide the space up into a grid of cells; for each cell, the Navier–Stokes conservation equations can be written down as nonlinear partial differential equations. The detailed form of this mathematics is not important in this context, but it is important to understand that they can now be used to understand the fluid flow and heat transfer.

There are several equations. The most fundamental are the basic equations describing the velocity of the fluid in the cell. There are generally three equations for three orthogonal velocity components (in a rectangular grid, one for each axis direction: *X*, *Y*, and *Z*). Additional equations can then be added representing the transport of other things such as thermal energy (for temperature) or contaminants. Finally, the equation for continuity (mass conservation) is added, from which pressure can be derived.

The property being calculated in each equation set is often termed a "variable," and the collection values for each variable (typically one value per cell) covering the entire calculation volume is often called a "field of data." From a conservation perspective, the fluid (or other transported properties) entering and leaving a cell through its faces must be consistent with each other and with its neighbors. This transfer of fluid through a surface is known as a "flux."

The actual formulation of the equations is dependent upon the numerical method employed (Section 17.2.2). As there is no analytic solution for most practical problems (i.e., the equations cannot be rearranged so that the answer can be calculated directly), the equations have to be solved numerically. Essentially, numerical solution is achieved by making a "guess" at the answer and inserting the values into the equations for every cell and every variable. Once inserted, the inconsistency in fluid flow or other variables (e.g., heat flow) into and out of any given cell can be calculated, allowing a correction to be made to the guess.

The process can then be repeated iteratively, and if successful, the error reduced to an acceptable level. This process is known as "convergence," and when successful, the solution is termed a "converged solution" (Section 17.2.2). It is important to note that being converged does not mean that the prediction of flow (or any other transported variable) is accurate, just that the errors have been reduced to less than a predefined "acceptable level."

There are other numerical reasons for the solution only to be an approximation to reality. The most common are as follows:

- Breaking the model (the representation of the data center, what is in it and affecting it) into pieces to form a discrete grid and the potential for numerical diffusion (Section 17.2.4).
- Not all the equations are a pure description of the physics. In particular, the turbulence model is an empirically derived relationship intended to capture the gross effects, such as mixing, that occur as a result of smaller scale fluctuations. For some applications, for example, aerodynamics, the equations may be tuned specifically for the particular niche (it may even be essential to do so for the technique to be accurate enough to be useful).
- The way in which we define the model: the physical geometry, airflow, and thermal boundary conditions that define features inside the calculation space (often known as "solution domain"), also the representation of the items that define interaction with the surrounding environment and indeed other elements of the model. These representations are often referred to as "boundary conditions" and will be discussed further in Section 17.2.3.

17.2.2 Numerical Methods

There are many approaches used in numerical mathematics to solve sets of differential equations like the Navier–Stokes equations. The method adopted affects the formulation of the equations for the solution, the way the space is broken down into the cells or grid, and how much computational power (in terms of processor and memory resources) is required. It even affects whether the solver (the computer program solving the equations) is likely to always produce a solution for any scenario or whether it will require special attention/control to achieve a solution at all.

The following are methods that the reader is most likely to encounter when considering CFD for application to data centers.

17.2.2.1 The Finite-Volume Method The finite-volume method is a numerical method based on dividing the space into control volumes (or cells), thereby discretizing the space into a mesh of cells each surrounding a data point. By casting the Navier–Stokes conservation equations onto the mesh of cells, by definition of conservation, the flux (airflow, heat flow, etc.) at any face of a cell leaving or entering the cell is equal to the flux entering or leaving the neighboring adjacent cell. Figure 17.2 shows for a 2D slice how, in what is known as a staggered grid approach, the scalar quantities (e.g., pressure (P) and temperature (T)) are calculated and stored at the cell center, while the velocities (u,v) and hence the fluxes (mass flow rate of the air) are calculated and stored on the cell face.

One of the advantages of this method is that it is easier to write down the equations: the fluxes through the cell faces are a direct result of the values in the adjacent cells. For example, the flow through two adjoining faces is directly dependent upon the difference in pressures in the two cells. The disadvantage is that it has just one point per cell—there is no information about gradient—and this results in numerical diffusions that can be addressed to some extent by the use of more sophisticated numerical algorithms such as higher-order differencing schemes.

At the time this chapter was written (January 2013), the finite-volume method is by far the most commonly used approach used for data center modeling; it has been demonstrated to be capable of producing usefully accurate prediction in an acceptable period of time. The reminder of this section will therefore assume this approach unless otherwise stated.

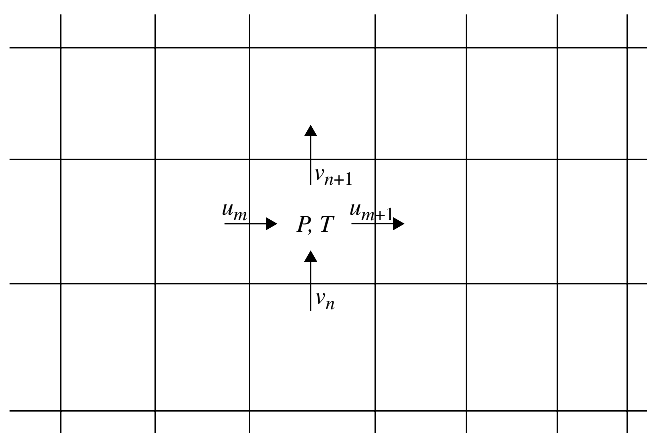


FIGURE 17.2 Slice through mesh showing connectivity, values, and fluxes. Courtesy of Future Facilities.

17.2.2.2 The Finite-Element Method The finite-element method is more commonly used for structural analysis problems, although it has also been used for commercially available CFD programs. The advantage of the finite-element method is that it has multiple points in each element (required in structural analysis to determine the stress): it carries more information per cell and, implicitly, the gradient or a variable.

Ironically, the use of multiple points is also the disadvantage of the finite-element method. This additional information makes the approach much more computationally expensive, not only in terms of memory but also in terms of computational speed. For fluid dynamics applications, it is often therefore too expensive, although it is used particularly where the CFD software may be intended to work alongside structural analysis software, making it easier to integrate the two technologies.

17.2.2.3 The Potential Flow Method The potential flow method is growing in popularity because the assumptions made result in simpler and quicker solutions. The key assumption is that the flow is a perfect fluid, that is, that it is inviscid (has no viscosity) and it is incompressible. The result is that the flow will be irrotational, such that there is no flow normal to the streamlines (streamlines show the convective flow path of the fluid/air or the conductive path for heat).

To use an analogy, it would be like a river flowing continuously along the river path with all the water flowing smoothly along it, around bends, and over and around rocks, all without vortices/circulations or cavities along the way. In practical terms, this will result in flow that cannot separate from a surface, and so, for example, flow may appear to unrealistically flow around corners.

17.2.2.4 Other Methods These are by no means all the numerical methods available to solve partial differential equations, and because of the size of a data hall and the number of objects and features inside it, there is continuing effort to identify faster solution techniques with acceptable accuracy. However, in the opinion of the author, these represent the techniques currently most commonly used for data centers.

17.2.3 What Defines a CFD Model

A fluid dynamics simulation is undertaken to understand the flow (and often the heat transfer) within the fluid volume. The primary variables are the mass, momentum, and energy.

To define a model for simulation, a physical description must be made in a way that can be represented through one of the methods described in the previous section. Plainly speaking, one builds a three-dimensional (3D), computer-generated model of the data hall.

The first thing to define is the size of the computational space in which the calculations are going to be made. This is commonly the shape of a shoe box—a rectangular block with dimensions appropriate for the applications concerned. This space—the solution domain—can be much more arbitrarily shaped in some cases, but for simplicity's sake, we will assume a rectangular box here.

Once defined, you can imagine that if this box were filled with air or any other fluid, absolutely nothing will happen inside it unless there is some input or force to disturb the equilibrium. Of course, the six faces of the solution domain represent an interface to the "world," which has an impact on the inside. Even as a sealed box, the faces can transfer heat, causing air close to the wall to heat up and rise. Once the air is moving, the walls of the box have another effect: friction from the surface will slow down the nearby air.

Where something is added, heat in the example mentioned earlier, this is a "source"; where something is taken away, it is a "negative source" or "sink." Of course, sources or sinks are not restricted to heat and friction but can be any source for mass, momentum (mass × velocity), and energy or indeed any other calculated/solved-for variable. Further, they can occur anywhere inside the solution domain as well as on the surfaces of the solution domain.

So, a fan *inside* the solution domain speeding up the air would be a momentum source because of the increase in velocity it creates. However, a fan bringing air *into* the solution domain would be both a mass source *and* a momentum source: it is adding air to the solution domain volume *and* bringing it in with velocity, thereby adding momentum. Furthermore, if the simulation accounts for temperature, then a source or sink of temperature (or "enthalpy") will also be present. The general description in CFD terms of these source and sink terms is "boundary condition."

Boundary conditions can be defined on a surface or defined over a volume, depending on what they represent. Most CFD programs include an extension to allow what is known as "conjugate heat transfer." This means that solid objects inside the solution domain can exchange heat with the fluid (or other solid objects) while the conduction of heat within the solid objects and the heat distribution in the fluid is calculated.

A key part of the boundary condition is representation of the effect of surface friction on local velocity and any resulting heat transfer. A common method is to assume a logarithmic variation of velocity with distance from the wall. This is unsurprisingly known as the "log law of the wall," and it assumes that the flow is turbulent. To represent more completely the surface boundary condition, the logarithmic profile is replaced by a linear relationship in the region immediately adjacent to the wall known as the "laminar sub layer" where the flow is no longer turbulent.

As well as direct mass and momentum sources representing a specified flow at a boundary condition, a flow boundary condition can be dependent on the conditions outside the solution domain. In this case, the flow depends on the conditions immediately outside the solution domain and adjacent to the boundary condition. The conditions could represent the following:

- Pressure due to pressurization in an adjacent space
- Pressure and momentum resulting from an external flow such as the wind
- Relative buoyancy due to thermal expansion (or contraction) of the air compared with the external (reference) condition

In fact, in some CFD simulations, it is quite common to treat the fluid as incompressible, where the expansion or contraction of the fluid does not significantly affect the mass of fluid in a cell. This is common in airflow and heat transfer simulations for the built environment, where the range of temperatures is usually small enough not to warrant treating the flow as compressible. Even so, the effect of temperature cannot be completely disregarded: the density variations that result do cause warmer air to rise and cooler air to fall. This can be accounted for by adding a force as a source term into each grid cell to reflect the local relative buoyancy. This approach is commonly referred to as the "Boussinesq Approximation."

17.2.4 Choosing a Solution Grid

Historically, one of the most time-consuming issues for a CFD modeler has been defining the solution grid or mesh. The choice and level of refinement can substantially impact the predicted airflow and heat transfer or even the ability for a converged solution to be achieved. Moreover, it also affects the level of resolution that can be studied.

The flow through an array of small obstructions could be predicted using a detailed model describing the geometry of each of the obstructions and the boundary layer around each. This would require a fine grid—a large number of cells in the area of the obstructions—and could therefore be computationally very expensive. However, if all that is of interest in relation to that group of obstructions is the degree to which they obstruct the airflow and the consequent pressure drop (with no interest in the details of the local flow), a simplified model is often employed. This uses empirically derived equations to calculate the pressure drop. It does so based on key characteristics of the obstructions: cylinder size and spacing for a group of pipes or cables, for example.

It is important to note that in most circumstances the only way to be confident that the model has sufficient grid to capture the features of interest is to undertake a grid-sensitivity study. This establishes whether or not the results of interest change as the grid is made finer or coarser.

The simplest form of grid is one where the space is divided up into an array of rectangular/brick-shaped elements. This can easily be visualized in two dimensions as a set of parallel (but not necessarily equally spaced) lines across the page and a similar set running vertically up the page. The result is an array of rectangles within an overall bounding rectangle. The lines are placed closer together where there is need for increased resolution, but the user should be aware that because the lines run all the way through the solution domain, a small gap between lines in one direction with large gaps in another will result. This may cause very long thin cells, and these may make it difficult to solve the equations and predict the flow solution. A long and thin cell is referred to as having a "high aspect ratio." A similar problem in solution can also occur if the lines in one direction are very close together and then suddenly they are very far apart. This is known as a "high expansion ratio."

In three dimensions, the process is identical. However, now the lines should be considered as planes and a third set must be drawn in the third orthogonal direction of the 3D axis system. This type of grid is often referred to as a "structured Cartesian grid." It is structured because each grid cell has six faces (one on each side of the box) and next to each cell face is another cell with a coincident face (Figure 17.3).

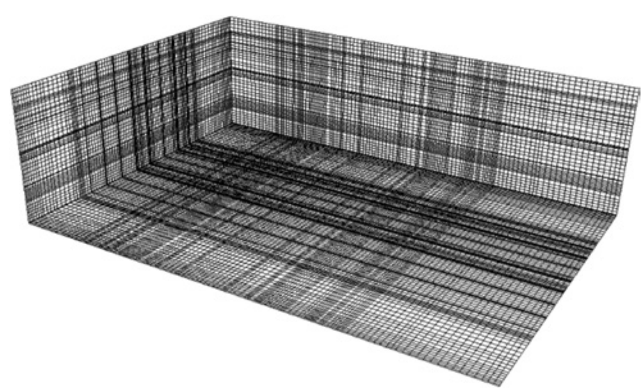


FIGURE 17.3 Structured Cartesian grid. Courtesy of Future Facilities.

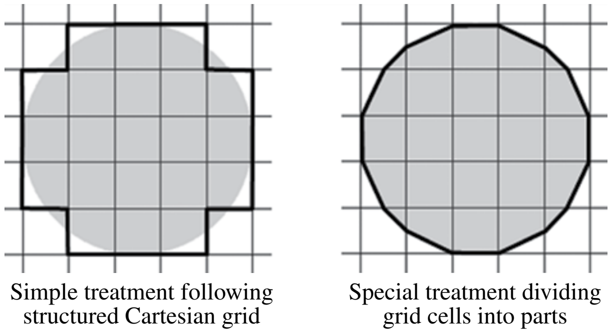


FIGURE 17.4 Surface approximation in a simple Cartesian grid approach and in a modified approach. Courtesy of Future Facilities.

There are two key challenges with a structured Cartesian grid:

1. The difficulty of capturing varying degrees of detail throughout the model without creating high aspect ratios (long, thin cells) or high expansion ratios (adjacent cell size changing very quickly).
2. Where objects or surfaces do not align with the grid (Figure 17.4), special treatment is required to avoid surface flows unrealistically detaching. Without special treatment, the surface has to be approximated by a line following the cell faces it cuts.

In some CFD tools, an "unstructured grid" is used in order to more easily describe complex shapes and more accurately predict surface effects. This can be achieved in a number of ways. The simplest way is to retain the Cartesian approach, but to no longer insist on the lines/planes extending all the way through the solution domain, and to allow a cell face to have more than one other cell faces next to it. This is often limited to dividing the cell into two in any given direction on a specific cell face. This limit is not essential, but when applied, this gridding approach is called "octri."

An alternative approach is not to insist on the cell shape being rectangular or box shaped (Figure 17.5). This allows the grid to follow more complex outlines and permits the cell faces to more closely reflect the true surface. The cells can be distorted brick-shaped elements but are more commonly tetrahedral in shape.

The fundamental disadvantage of unstructured grids is that they require more computer memory and are slower to calculate per cell. This is because if the cells are not a simple rectangular mesh of brick-shaped cells where the next cell is implicitly known and then additional connectivity information has to be stored and processed in order to know which cells are neighbors to each other. Of course, this disadvantage may be outweighed by the fact that an unstructured grid may allow the model to be constructed and calculated with fewer cells by only refining where necessary, but this is not guaranteed. Refinement is arguably more effective when using

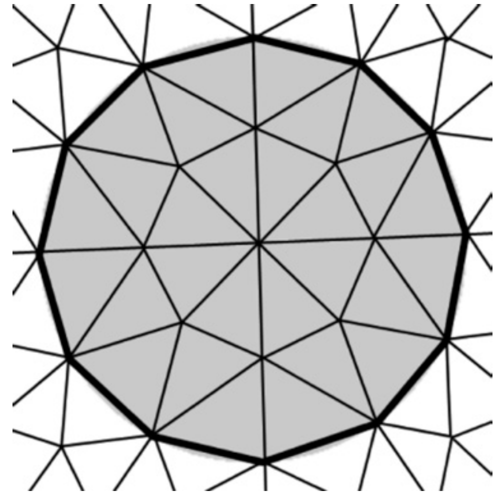


FIGURE 17.5 Shapes can be directly represented to coincide with the unstructured faces of the cells. Courtesy of Future Facilities.

the tetrahedral or body-fitted approaches, but here, gridding has an additional criteria: the cells must not become too distorted with small internal angles.

An advantage of CFD designed for a particular purpose is that the grid rules can be developed to suit the application involved. These rules are based on an awareness of the types of objects that will be in the model and their grid requirement.

17.2.5 Calculating the Solution

As there is no analytic solution for anything but the most simplistic configuration, the Navier–Stokes equations have to be solved numerically. The way this is done is essentially a "guess and correct" approach, where the solver is provided with an initial guess of the solution (normally zero velocity with a single value for temperature and pressure throughout the solution domain).

The boundary conditions are superimposed on this set of field data and provide sources and sinks of mass, momentum, and energy for the cells they are located in. Without correction, the addition of these sources and sinks locally will create errors in the conservation equations for the cells, and so the values need to be adjusted to account for the changes. These corrections are made on a variable-by-variable basis and on a cell-by-cell basis for each variable.

Of course, the values for a cell depend on the values in all its neighbors. So, because the values are recalculated cell by cell, the first variable will no longer be consistent with the other variables and, similarly, the first cell recalculated will not be consistent with its neighbors after they have been updated. As a consequence, this recalculation of the values for all the variables in all the cells has to be undertaken many times—as many times as required until the errors in the conservation equations are reduced to an acceptable level. At that point, the solution is deemed to have "converged." Each time all the values for all the cells are recalculated for all the variables is termed an "outer iteration." In practice, some variables (typically temperature and pressure) may be recalculated several times during any given outer iteration. These repeated recalculations for a single variable carried out within a single outer iteration are called "inner iterations."

The fact that these recalculations are made does not guarantee that the values calculated for each field of data will necessarily be closer to the true answer iteration by iteration. Consider riding in a car. If the car had no suspension, the ride would be very uncomfortable because even though the intention (desired solution) for the car is to ride smoothly along, it would deviate from this smooth path bump by bump. By adding springs, the severity of the deviation of the ride position in the car can be reduced by the spring absorbing some of the energy and starting a decaying oscillation. In most circumstances, the oscillation will gradually decay, but if the car hits bumps at a rate that excites the natural frequency of the spring–mass system that the car represents, the oscillation could go out of control. For a car's suspension, the solution is to add a shock absorber or damper to limit the motion on the spring, thus bringing the car back to equilibrium ride height more quickly. Making the damper too heavy results in a hard ride with little oscillation, but once the car deviates from ride height, it takes a long while to move back to equilibrium position. With a very light damper, much of the oscillation is allowed and the energy from the bump(s) will be only very slowly absorbed, giving long periods of oscillating ride height.

The convergence process for numerical solution in CFD can be considered similar to this analogy. If undamped, the solution may oscillate or even "diverge" (move away from the solution). So, damping is used to stabilize the solution. This is normally done in one of the following two ways:

1. **Linear relaxation**—where the calculation of the change required for a value is ΔV, then the change applied is $f\Delta V$, where f is the linear relaxation factor (between 0.0 and 1.0)
2. **False time step relaxation**—where the change is calculated as though it were the change that would happen if a small amount of time were to pass

Such false time step relaxation should not be confused with "time varying" or "transient simulation," where the CFD is used to predict time-varying flows. For time-varying calculations, the solution is undertaken by using a similar iterative process to solve the equations for the change that will occur over a small period of time. In the same way that the grid resolution is important in space, so is the time step size in time. The time step has to be small enough to capture the time-varying features of interest. Generally speaking, the smaller the time step, the fewer the number of outer iterations required to reach convergence.

17.2.6 When Is the Solution Ready for Use?

We have discussed the concept of convergence and a converged solution: the process of reducing the errors in the equations and therefore being closer to the principle of conservation of mass, momentum, and energy throughout the solution domain. The errors left in the equations are often termed "residual errors."

One way of measuring residual errors is to add up all the errors (imbalances) from all the cells for each variable and to then compare the error sum with some reference value. Commonly, the reference value is the incoming mass, momentum, or energy (as appropriate) for the variable. The solution is deemed converged when the errors for all the variables fall below a small proportion (e.g., 0.5%) of that reference value. The incoming mass, momentum, and energy is commonly estimated before the solution takes place, and so the final performance may not be a true reflection of the true incoming mass, momentum, and energy. Also, *where* the error occurs in the solution domain can be very important. If much of the error is localized in an area of the solution domain that is of no particular interest, then a higher error can be tolerated. If, on the other hand, the error occurs at a point of key importance, then the error may be less acceptable.

Another way of determining whether the solutions are acceptable (at least for non-time-varying calculations) is to monitor the change in key variable for important interest points and see whether they have achieved a steady condition. If an acceptable residual error has been achieved and the points of interest have stable conditions, then it is likely that the solution is sufficiently converged to be considered representative for final analysis.

Once you have established that from a numerical perspective the solution is ready to use, it is important to review the model and results with the possibility in mind that they may not be correct because the model is not a sufficient representation of reality. Put another way, one must review the model with a healthy degree of skepticism: a model is only as good as the input data that defines it. There are two possibilities for it not being correct:

1. The user has made a mistake building the model.
2. There is insufficient detail in the model to capture all the key features.

In the event this is a predictive model and there are no measured results to compare with, it is often good to have an expectation of what the solution will be and then, when the solution is different, to question why it is different—is there a mistake in input or is there something happening that is reasonable but unexpected? Of course, when the model is of something that exists and is being used for trouble shooting or onward development, it is always best to first model the existing scenario and develop a realistic model before using it for predictive simulation.

17.2.7 What Are the Results?

The solution methodology delivers values for each of the solved-for variables. The basic set of variables includes pressure (derived from continuity), temperature, and velocity components (and thus the resulting overall fluid velocity and direction). The data is available for a grid of points throughout the room, depending on the grid that is chosen (as described earlier).

In a structured Cartesian grid, the data is stored in a set of 3D arrays, one array for each solved-for variable and (in the finite-volume method) one value per grid cell. Given the rich data set, it is possible to undertake a wide range of postprocessing. Data center-specific postprocessing and metrics are described later; however, almost any CFD tool will provide the following.

17.2.7.1 Result Planes A "result plane" is a graphical depiction of the values of a calculated variable. It is displayed in the 3D model as a plot (normally in plan or elevation orientation, but not necessarily so), where each grid cell in a selected plane is colored. The color is set according to the value in that grid cell for the selected variable in question (Figure 17.6).

In the example shown, the plot is of temperature variation in grayscale, with white being hot and dark gray (almost black) being cold. On most computers or printouts, this would normally be in color, typically showing purple or blue as cold and red as hot.

Although it is normal to draw the variation in a smoothed way, interpolating the values between points and creating the impression of continuous variation, most tools also allow you to plot just the calculated value in each cell. This is sometimes helpful as some CFD programs are less clever when making the interpolations, especially near solid boundaries and a solid–fluid interface. In such a scenario, a simple interpolation may produce a misleading plot.

A result plane can also be used to plot airflow patterns (Figure 17.7) (or to plot heat fluxes for conduction, where appropriate) by combining the three orthogonal velocity components (or three fluxes). The magnitude of velocity (or flux) in the plane is normally indicated by the size of the arrow, while the magnitude of the 3D velocity is often represented by the color scale or the grayscale. Color is also often used to represent another variable so that the relationship between flow and other variables can be seen more easily.

17.2.7.2 Streamlines "Streamlines" are commonly used to understand the convective flow path of air (or conductive path for heat) (Figure 17.8). They are easy to relate to visually as most people have seen streamers, smoke carried in an airstream or dye in water. They simply follow the convective path of the fluid from a single point or set of points.

17.2.7.3 Surface Plots Another way of visualizing results is to graphically depict results on a surface. There are two basic types of surface plot:

1. **Distribution on an object surface**, such as surface temperature or surface pressure. In the first example, the surface temperature is commonly the temperature of a solid at its surface where it meets fluid, often air. This is useful, for example, when modeling electronics. Surface pressure, on the other hand, is normally the pressure in the fluid adjacent to the solid surface. An example of its use might be the pressure distribution on an aircraft or vehicle when optimizing lift (or down force) and drag.
2. **A surface representing a constant value of a calculated variable**. This is sometimes referred to as an

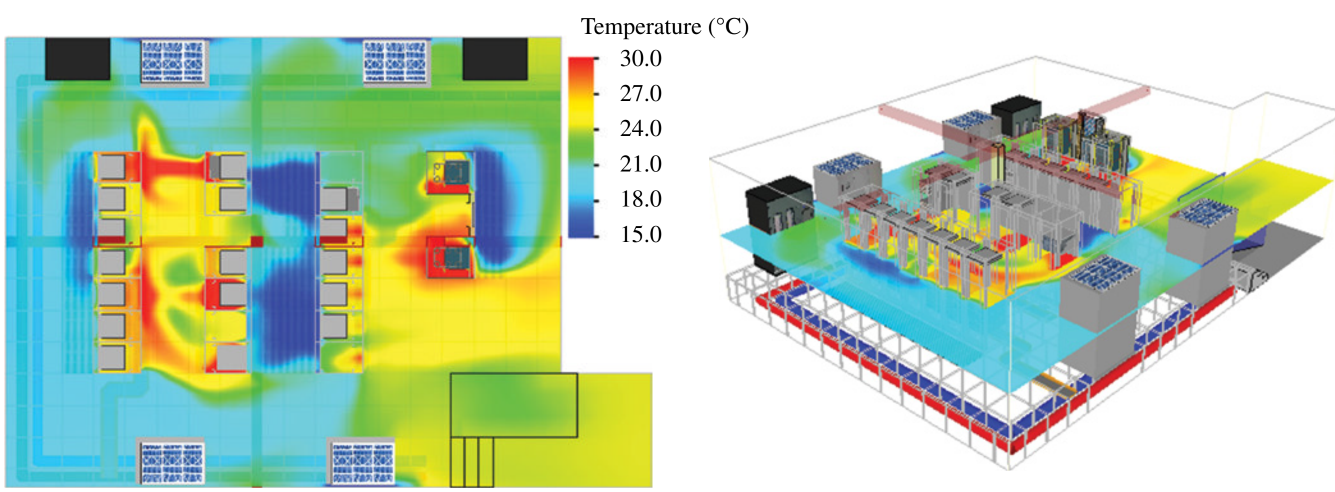


FIGURE 17.6 Result plane of temperature and halfway cabinet. Courtesy of Future Facilities. Please visit companion website to access the color version of this figure.

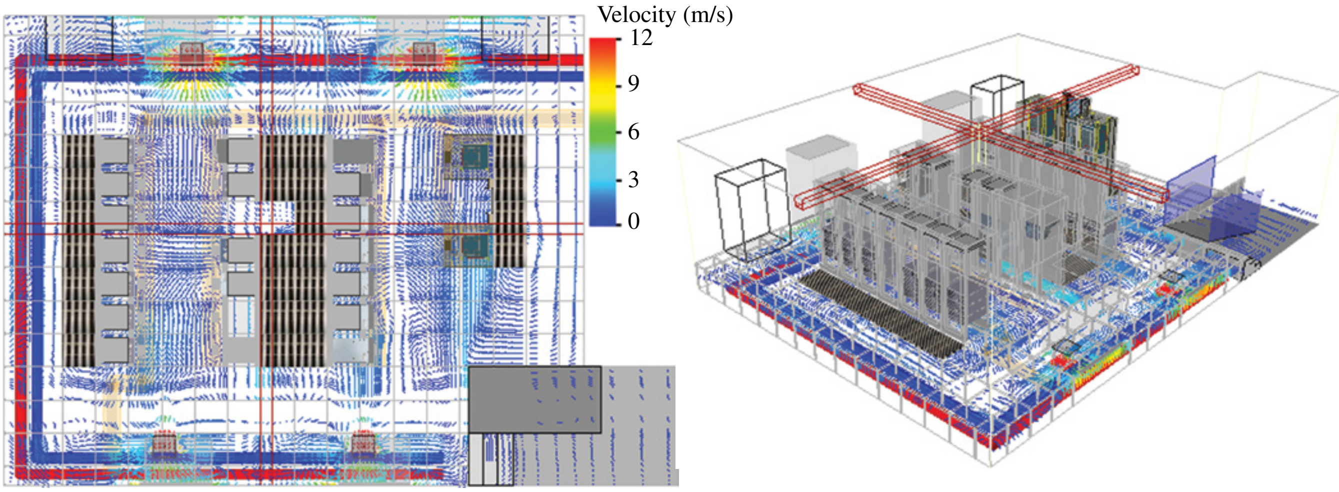


FIGURE 17.7 Result plane of flow pattern in a floor void. Courtesy of Future Facilities. Please visit companion website to access the color version of this figure.

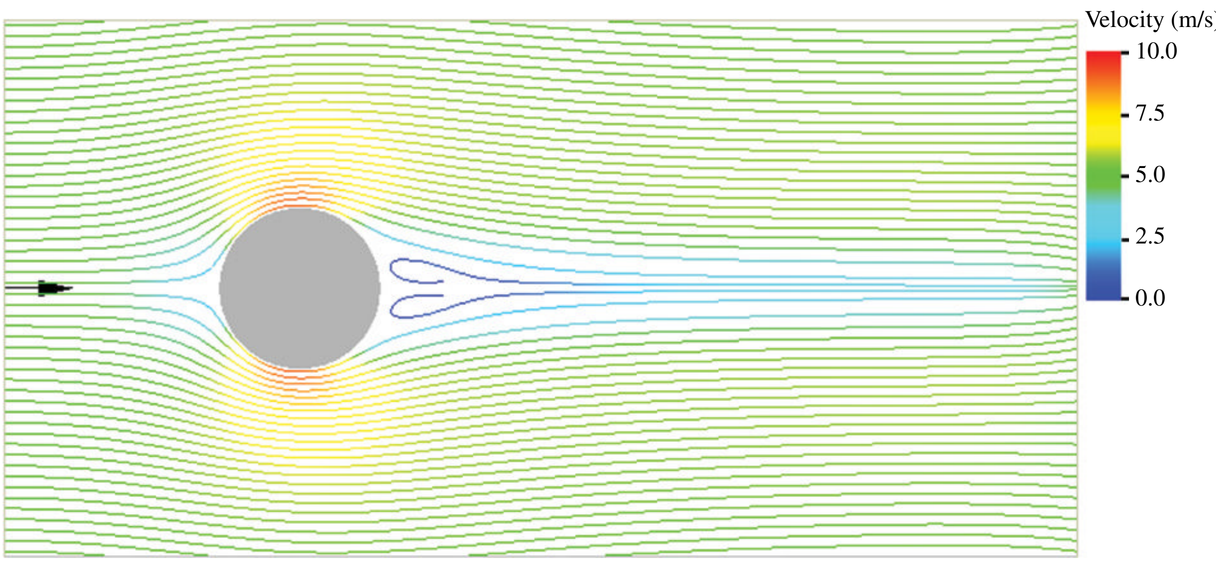


FIGURE 17.8 Streamline flow around a cylinder. Courtesy of Future Facilities. Please visit companion website to access the color version of this figure.

"iso surface." An example of its use might be to show the volume in which a contaminant/pollutant is at a hazardous level. The surface would be drawn at the critical level with outside the surface—away from the pollutant source, being below the critical level, and inside the surface—nearer the source—being above the critical level.

17.2.7.4 Postprocessed Data As described earlier, the large volume of field data lends itself to 2D and 3D visualization in graphical views. However, the data can also be postprocessed to provide aggregated data for quicker understanding. This is sometimes referred to as "derived data."

For example, the total flow rate through an inlet or outlet is often recorded perhaps with its average temperature and/or concentration. The flow is summed and the averages calculated using all the cells/cell faces that coincide with the inlet or outlet in question. The more tailored the CFD program is to an application, the more tailored this derived data can be.

17.3 APPLICATIONS OF CFD FOR DATA CENTERS

17.3.1 Typical Uses

Strictly speaking, CFD is most commonly applied to a data hall (the room where the IT equipment is housed) rather than a data center (the entire facility, including the support infrastructure outside the IT room). It is sometimes applied to other aspects of the data center, such as the following:

- Airflow around outdoor chillers to determine whether the hot air is exhausted effectively without being reentrained
- Airflow and cooling of batteries and infrastructural equipment such as UPS systems
- Generator halls
- In fact, almost any of the support spaces, whether occupied by equipment or people

Of course, the requirements of CFD vary for each type of space, especially if the focus is human comfort rather than equipment operating conditions.

This chapter will focus primarily on the data hall but will also address some of the other issues that may occasionally be addressed for a data hall or arise in other data center applications. For the data hall, there are a number of different tasks that may need to be considered. These are as follows:

1. Conceptual design
2. Detailed design
3. Assessment
4. Detailed assessment and troubleshooting
5. Operational management

In addition, it is also common to consider the cabinet scale—that is to say configuring IT equipment in a single cabinet or group of cabinets independently. This means that IT equipment can be deployed in cabinets with the confidence that the internal configuration is effective and will not undermine room-scale management. Rack/cabinet-scale simulations are considered important because some data center professionals report that equipment cooling problems occur as a result of internal cabinet configuration issues as frequently as they do from room configuration issues.

17.3.2 Use for Data Center Design

Most data center design scenarios do not need to consider the detail of specific equipment configurations. In general, they only consider the ability of the room cooling system to distribute the cool air throughout the facility and scavenge the warm IT equipment exhaust air effectively. In the author's experience, about half of overheating problems occur as a direct result of equipment configuration. Accordingly, a good design does not guarantee its successful operation, but subject to due care at the equipment configuration stage, it does increase the likelihood of success. For this reason, it does not matter that in early design the end user cannot normally say exactly what equipment will be installed. The end user only really needs to know, conceptually at least, the different types of equipment that will be installed from a cooling methodology perspective.

The level of detail required in the model will depend on the design decisions being considered. While adding unnecessary detail will not result in poor decisions, it may take more time than necessary for the modeler to create the model and more compute time to solve the model. This will limit the number of design iterations that can be studied in a given space of time, although some tools do have the option to undertake the calculation at a lower level of granularity, even when more detail has been included in the model.

17.3.2.1 Conceptual Design In conceptual design, CFD can be used to test and optimize the overall design philosophy. Historically, CFD simulations have been made for a full load scenario with a very uniform power and airflow distribution. As a consequence, the conceptual design scenarios often do not exercise the system for the intermediate and nonuniform loads that the system will have to address.

The challenge here is that it is necessary to create a range of scenarios that test the system for a range of realistic conditions when very little is known about either what IT equipment will be installed or when it will be installed. Consequently, it is not appropriate to be too detailed in modeling; only sufficient detail to capture the key characteristics and test sensitivity is required.

If the appropriate decisions are made (Section 17.4), a conceptual design model can be used to:

- Test different cooling strategies, be they conventional computer room air conditioning (CRAC) or computer room air handler (CRAH) systems, in-row or overhead cooling units, cooling units with economizers, or direct fresh air economizer cooling
- Test the number, size, and layout of cooling systems to optimize the cooling distribution, accounting for the room size and shape and other architectural features (such as columns), equipment layout, and power distribution
- Optimize cooling paths, including raised floor height, false/drop ceiling height, and ventilation duct size
- Optimize floor grille, ceiling grille, and duct grille layouts
- Test segregation concepts, such as cold aisle or hot aisle containment systems
- Allow for the effects of notional power distribution in the data hall and optimize IT equipment layout
- Test part load configurations
- Test redundant cooling configurations
- Optimize energy efficiency
- Evaluate the impact of generic representations of cable

It is normal for conceptual design models to make a number of assumptions about the IT equipment and practices to be deployed in the data hall, including the following:

- Generic cooling system such as CRAC/CRAH units with simplified/idealized cooling distribution and nominal cooling capacity
- Known cooling set points, such as supply air temperature and air flow rate with limited, if any, control of these parameters during solution to respond to the resulting conditions
- IT equipment is configured for front-to-back ventilation
- Cabinets are well configured and do not allow internal recirculation
- Cable penetrations are well managed and have little impact

17.3.2.2 Detailed Design Of course, a model created for a detailed design assessment can be used to make conceptual judgments, but often, it is unnecessary to develop the model to this level of detail until conceptual decisions have been completed. That said, adding detail to the model allows these decisions to be confirmed with greater confidence. It also allows the modeler to test additional design assumptions that may undermine or enhance the performance. Additional considerations will include the following:

- Detail of cooling system, such as specific CRAC/CRAH units with particular fan type and consequent airflow pattern
- Detail of control system, including sensor locations and control characteristics, to include condition-sensitive capacity and variable air volumes
- The impact of more realistic equipment choices allowing for the following:
  - Non-front-to-back configurations where appropriate (e.g., in switch cabinets)
  - End-user practices, such as top-of-rack switches and not using lower slots
  - Higher IT equipment power density so that IT equipment does not fill the cabinet
- The inclusion of more realistic cabinet and equipment configurations that may affect recirculation, such as the following:
  - Whether cabinets are mounted off the floor
  - Empty slot blanking policies
  - Gaps around the mounting rails to the sides, above the top, or below the bottom
- Details of aisle containment systems, including potential leakage paths and control measures to achieve cooling system–IT equipment airflow balance
- The effect of realistic cable management practices, including the following:
  - Variation in cable route sizes and densities based on type of IT equipment deployed
  - Realistic cable penetration sealing performance and other penetrations, such as raised floor holes

Detailed design models are still expected to use typical IT equipment types, notional cable routes and cabling densities, and uniform damper settings. Consequently, they generally will not contain the diversity of configuration typically seen in an established data hall. In addition to these more detailed models, to better understand and optimize the data hall configuration, similarly detailed simulations may be undertaken at a cabinet scale to understand the implications on internal cabinet airflow of user operational choices (e.g., the use of top of rack switches with side-to-side ventilation) and optimize/influence practices, where possible.

17.3.3 Use for Assessment, Troubleshooting, and Upgrade

CFD models created for existing data hall assessment troubleshooting and upgrade can be made at two levels of detail similar to conceptual design and detailed design. The consequences of the simplifications will be similar. As such, if the conceptual design approach is adopted, only high-level issues will be predicted: any changes proposed and modeled will therefore only be appropriate at the same high level.

In order to understand issues at rack level, more detail will be required in the same vein as the detailed design option. Often, calibration (Section 17.3.4) of the model will be necessary (and is always desirable) using monitored data for the systems in practice. Examples are as follows:

- Measurement of airflows and temperatures for the CRACs/CRAHs or other cooling systems to determine the actual control response.
- The power distribution at as near IT system level as possible to determine the impact of system use and resulting utilization. This is important since power consumption and consequent heat dissipation are strongly dependent upon the applications deployed on the IT equipment.

This level of detail and the fact that a model is "calibrated" (Section 17.3.4) should not bring the user to believe that the predictions will be perfect. Although the simulation results from a well-prepared model will provide useful qualitative results and, to a great extent, good quantitative results, a simulation cannot be guaranteed to predict individual equipment inlet temperatures precisely. The simulation may, in some instances, predict a problem in a different cabinet—perhaps an adjacent cabinet to where the problem really occurs. This happens because some airflow phenomena can be very sensitive to small details.

For example, when the jets from two opposing CRAC/CRAH units pass each other, this can cause a recirculation between the two jets. The recirculation is like a small tornado: low pressure at its center, while all around it is at a similar pressure to the typical pressure in the raised floor. If this low pressure is below a floor grille/perforated tile, then there will be little flow upward and potentially even flow downward. Such a flow feature can be disturbed by very small forces and can easily be predicted one tile away from its true placement. If this is the case, a tile that should have virtually no airflow in reality may have airflow of several hundred liters per second. Meanwhile, its neighbor that should have hundreds of liters per second is simulated as having almost none. In such an instance, the predicted flow could be significantly in error. However, the qualitative affect is likely to be predicted accurately, even if slightly in the wrong location.

With this acknowledged, CFD has been used to model many corporate data centers. Indeed, it has been able to identify and provide understanding of problems and consequently enable resolutions to be tested prior to implementation.

17.3.4 Use for Operational Management

Classic operational management using software tools has focused on what are known as Data Center Infrastructure Management (DCIM, pronounced *dee-sim*) tools, which claim to provide a tool set for the facility managers to deploy new equipment accounting for space, power cooling, and network.

The limitation with DCIM tools is that the cooling is based on the *design capacity* of any given IT rack/cabinet and can be an overprediction or an under prediction of actual capacity. The reality, however, is that *capacity varies from IT equipment rack to IT equipment rack*.

DCIM tools point to the ability to integrate data from live monitoring systems to see live power consumption and live temperatures, allowing the end user to understand the true environment. But in practice, the problem is that this information tells you *what has happened* rather than what *will happen*. What is required is for DCIM to become Predictive. In practice, the only way for this to be done for a data hall is to use CFD to analyze the impact of a change or changes on the air management and cooling performance.

Consequently, the use of CFD in operational management is perhaps the most demanding yet rewarding application of CFD to the data hall. At the time of writing, operational management using CFD is only really practiced for some enterprise-scale data halls. This is ironic, since one of the key challenges to the use of CFD operational management is the complexity and time-consuming nature of this simulation: it could be applied much more easily and quickly to the many smaller data halls that exist in almost every city all over the world. However, the rewards for any scale data hall are potentially very significant. The difficulty in deploying a variety of IT equipment systems in a data hall is that the impact of a change in one area of the hall may be felt somewhere else. In fact, it is a little like a water bed—if you press down in one location, the bed is likely to rise at locations quite some distance away.

Large data halls can have many IT equipment changes every day. Over time, it is therefore common to arrive at a configuration where placement of a new item of IT equipment can cause overheating in itself or some other, quite unconnected, item of equipment. Once this happens, the typical enterprise reaction is to protect the mission-critical operations in the data hall by refusing any further deployments. This can, and often does, occur at between 60 and 70% design capacity for the data hall. In a mission-critical facility, one of the key challenges is that once deployed, it is very hard to power down IT equipment: going back on a problematic deployment is almost impossible.

CFD provides a method by which something unseen, air movement, can be visualized and understood. This has provided an opportunity to address and resolve some of the cooling challenges in facilities where apparently substantial design capacity had seemingly been lost. Further, these data centers are often substantially overcooled in order to address overheating in a handful of items of IT equipment. In practice, CFD has been used to bring problematic data centers back under control, increasing available capacity to well in excess of 90% of the original design intent and allowing cooling to be undertaken in a much more energy-efficient manner. The cost savings can be tens or even hundreds of thousands of dollars per hall in any given calendar year. Likewise, it can avoid tens of millions of dollars being spent ahead of schedule to build another data hall, not to mention the less tangible, but still very real, reduction in risk and improvement in availability.

An often unrecognized benefit of a calibrated data hall CFD model is that to define the model, it is necessary to include all key items of infrastructure and IT equipment present within the hall. As a consequence, it is natural (as some software tools have done) to extend the model to provide inventory tracking, space, power, cooling, and network capacity planning, as well as providing the cooling simulation. CFD, therefore, provides an almost unique opportunity for predictive analysis to be applied as a matter of course for any chosen data hall. The main disadvantage of existing CFD software programs, when compared with current DCIM programs, is that the majority of current-generation CFD tools provide snapshot analysis that is primarily suited to data hall design rather than to operational management. Those CFD programs that are designed for operational management are still data hall focused rather than data center or indeed enterprise-wide tools.

17.3.4.1 Calibration Calibration is an essential part of data center operational management using CFD. Deployment tools need to be available for everyday use by the facility and IT management teams, so the traditional snapshot use of CFD by fluid dynamics specialists is not appropriate. In reality, however, while the tools can be given user-friendly interfaces that make them attractive to data center professionals, the complexity of a data hall means that it is easy for a data hall "virtual facility" model to provide results that may look convincing but at the same time may actually be misleading. For this reason, it is essential that the virtual facility model is periodically recalibrated if it is to be used with confidence to determine the most appropriate IT equipment deployment strategies.

Unlike calibration of an item of test equipment—where calibration of the equipment is the process of comparing what the instrument reads with readings from a more accurate tool, allowing the readings from the test equipment to be corrected to provide true readings in the field—calibration of a virtual facility model is used to provide data to see if the model is still sufficiently representative of reality to be used for operational management. If not, the measurements are used to enable the user to determine what features need updating in the model to make it sufficiently accurate.

Depending on the rate of change in the facility, such calibrations should be carried out periodically, normally no less than quarterly as an absolute minimum. Fortunately, the increase in built-in monitoring systems for data halls is resulting in more of the data required being available automatically.

To make a calibration, the following are typically monitored:

- CRAC/CRAH or other data hall cooling system temperature and airflows
- Perforated/slotted floor grille air flow rates and temperatures
- IT equipment power draws as near to the IT equipment as possible
- IT equipment inlet air temperatures

Probably, the most difficult of these to measure are the airflow parameters. In particular, airflow through modern high-open-area perforated floor tiles is difficult because of the low resistance of the tile itself combined with the large number of floor tiles on a single open-floor void. Consequently, the introduction of the measuring device can significantly affect flow measurement. Flow hoods with back pressure compensation (designed for flow measurement from grilles on more typical building ventilation ducts) can help in correcting the measurements, but even these cannot correct the measurement if the addition of the hood causes the flow to be reduced to nothing. Nor can they make satisfactory measurements where the flow is particularly turbulent/unsteady. This latter feature is commonly true at perforated tiles near CRAC/CRAH down flow units. Similarly, making velocity measurements in the outflows from CRAC/CRAH units is difficult for the same reason, and flow measurements for these units are normally most readily achieved by measuring the return airflows.

17.4 MODELING THE DATA CENTER

Like any simulation, the value of a data center model will depend on the quality of the particular model in question as much as the simulation tool of choice.

At present, the user of the tool is almost entirely responsible for the model, including the representation of proprietary items such as CRACs/CRAHs, PDUs, IT equipment, etc. Although some CFD tools offer libraries of equipment (commonly referred to as "symbol libraries"), these are limited in scope and often require review and tailoring for their use in the chosen facility model. The modeling decisions made can, therefore, critically affect the outcome of modeling—"garbage in–garbage out." It is therefore important that the user should use traditional approaches to gaining confidence in the model, including the following:

- Having an expectation of the result and questioning why the result is different. That is, is there a flaw in the model, or is something genuinely happening that was not anticipated?
- Undertaking sensitivity studies where there are uncertainties. One of the key advantages of simulation is that it can be run for a variety of conditions, and so, where there is uncertainty, parametric variations can be undertaken to test sensitivity.
- For a real facility, where possible, use calibration of the model to ensure it is well specified so that the effect of changes can be expected to reflect reality.
- When making models of items to be included in a virtual facility model, first test them independently (in a separate purpose-built test model) before using them in the virtual facility model.

There now follows some high-level (and definitely not exhaustive) guidance on data center modeling.

17.4.1 Architecture

In a data hall application, the architecture of the envelope is normally only important from a shape point of view. This is because the internal heat gains are normally so large that fabric heat transfer is at most a second-order effect. It is also common for data halls to be internal spaces and, as such, to be surrounded by other controlled environments, so the temperature differences are not significant. Internal architectures such as columns and partition walls need to be considered primarily from an airflow perspective where they impact the air distribution associated with the cooling system, particularly—but not exclusively—on the supply side. This section discusses the two classes of architecture identified focusing on when it is necessary to include additional details.

17.4.1.1 The Data Hall Envelope Consider a typical enterprise data hall with floor area of 1,000 m^2 (~10,000 ft^2). A typical plan might be 40 m in length and 25 m in width. It is common for the data hall to be quite tall, so, including a floor and ceiling plenum, let us consider the wall height to be 5 m.

Given modern building codes, it is likely that the *U*-value (heat transfer coefficient measured by how much heat is passed from the air on one side of the material to the other over an area of 1 m^2 for every 1°C difference in temperature) for the envelope will need to be significantly less than 1 W/m^2 K. Because this represents a worst case, we will use it as the heat transfer coefficient. If the walls were all exposed to the external environment, then total heat transfer through the vertical façade would be heat transfer coefficient (U) × area (A) × temperature difference (ΔT):

$$\begin{aligned} UA\,\Delta T &= 1.0\times\left(2\times 40\times 5+2\times 25\times 5\right)\times\Delta T \\ &= 650\times\Delta T\ \mathrm{W} \end{aligned}$$

This represents 650 W of heat transfer per degree of difference.

If the data hall is an internal building space, then, given a temperature difference of only a few degrees between inside the data hall and its surroundings, the heat gain is unlikely to be at all significant.

If the data hall is external and placed in a hot climate with an extreme temperature difference of 20 K,[1] the heat transfer through the walls would be 13,000°W. This is likely to be around 1% of the design internal full load heat gain.

The roof can be considered similarly. With an area of 1000 m^2, the heat transfer would be of the same order, giving a conducted heat gain of a few percent of the design internal full heat load.

The floor is likely to be relatively neutral in normal operation for a modern, low-energy data hall, since data halls are commonly on the ground floor and increasing supply air temperatures are likely to be similar to the ground temperature. This should be considered when considering cooling margins, but is probably not important to airflow and heat transfer calculations in the data hall.

[1]Temperature differences in SI (metric) units are normally quoted in Kelvin (K). In practice, this has no impact since 1 K = 1°C.

Now, let's consider a situation where the façade is exposed to solar radiation. We will again consider a scenario that will be close to worst case in order to understand the magnitude of the solar radiation effect. For a rectangular data hall, only two of the walls and the roof can be exposed to direct short-wave solar radiation at any given time. Assuming:

- The angle of incidence onto all surfaces was 45°, reducing the incident radiation per m^2 by a factor of 0.707.
- If the sun were immediately above the data hall, then the area subjected to short-wave radiation would be reduced from 1325 to 1000 m^2. If the solar intensity were 1000 W/m^2, then the heat gain falling on the building could be of the order of 1 MW. If all the heat were absorbed by the surface and given the U value of the surface of 1 W/m^2 K, then it would be reasonable to assume that the resistance of the wall plus internal surface resistance is the order of 10 times the resistance resulting from the external surface heat transfer coefficient. As a consequence, it would be possible for 10% of the solar radiation to occur in the data hall. The heat gain from solar radiation could be of the order of 10% of the internal heat gain. This clearly cannot be ignored from a capacity standpoint at least. If the data hall has a substantial area of glazing or is an older building with lower thermal performance, then the thermal performance of the façade must always be considered from a capacity and internal temperature perspective.

There is one more circumstance where building surfaces starts to play a role thermally. This is when a complete cooling failure scenario is considered. With a surface heat transfer coefficient of the order of 10 W/m^2, a surface of 1000 m^2 could absorb 10 kW per degrees of difference. As the temperature rises in the data hall, while this will be insufficient to offset the heat load, it can contribute toward a slowing in rate of temperature rise.

17.4.1.2 Internal Architecture Again, temperature differences across elements of internal architecture are not normally of key importance. However, the presence of internal architectural elements can significantly affect airflow.

For internal partitions, be they to provide independent airflow plenums or to segregate areas of the data hall for other reasons, their segregation efficiency is of great importance. As a consequence, the user should pay particular attention to any leakage paths. For conceptual design, the designer may assume that the segregation is perfect, but for detailed design or assessment and operational management in particular, care must be taken to ensure a good understanding of any leakage paths.

Over recent years, considerable focus has been paid to the segregation efficiency of the raised floor. The piecewise construction itself does allow leakage between the floor tiles, but in a typical data center, this leakage is small—of the order of 1%. However, in legacy data centers or indeed any data center where little or no attention is paid to the management of floor penetrations, open cable penetrations and other poor tile cuts (e.g., around cooling and power infrastructure) can result in unmanaged leakage of the order of 50% of the cooling airflow. From a modeling perspective, it is critical that the CFD tool can account for this leakage not only in magnitude but also in location: the location may control whether or not the cooling air is useful.

The introduction of segregation as a key element of energy-efficient cooling, such as aisle containment, has made this aspect even more critical. Incorrect assumptions about the leakage can completely undermine the theoretical performance of the cooling system by allowing unexpected recirculation (potentially hotter and more dangerous than without containment in place) and can also result in flow control challenges that otherwise might not be experienced.

Detailed design and assessment models and operational management models must also be able to capture the behavior of the flow devices such as perforated floor tiles. It is important to recognize that one tile of 40% open area will not necessarily perform in the same way as another tile of 40% open area. This is because the air volume passing through the tile will depend not only on the resistance to airflow and the pressure difference but also on the geometry and its effect on turning the air through the tile and on the tile's outlet flow pattern/velocity distribution. The ability of a CFD program to characterize key airflow devices is also, therefore, critical not only to the prediction of airflow in the room but also to the fundamental prediction of airflow distribution across the array of perforated tiles.

Other internal obstructions should be considered on the basis of their position relative to key airflows. For example, columns are often important because they interfere with the under floor airflow and may result in an inability to have regular perforated floor tile and equipment distributions. Beams, on the other hand, are often unimportant because they are typically at high level where there is little air movement. An exception to this is where the return air systems are at high level and the beams create segregation that prevents, or aids, the hot air in its return to the cooling system.

17.4.2 CRAC/CRAH Units and Cooling Infrastructure

The cooling and cooling distribution system are particularly critical in a CFD model of a data hall. For most data hall modeling, however, the CRAC or CRAH units, be they conventional down flow units or more recent alternatives such as in-row, in-cabinet, or overhead cooling systems, are represented as a black box. The assumption is made that any temperature variation on the return, and any cooling that is applied, results in a homogeneous, fully mixed airflow at the CRAC/CRAH outlets/supplies. For most scenarios, this is a perfectly adequate representation, as the airflow leaving CRAC/CRAH units is usually highly turbulent and so any nonuniformity at the outlets normally mixes out very quickly. Of greater importance is the representation of cooling capacity and the controls for air volume, cooling, and any humidification/dehumidification, if applicable.

For conceptual design or concept assessment, all that is generally required is the ability for the black box cooling unit to deliver the design airflow with the cooling to match the heat load in the space. Most data center CFD tools provide simple controls to govern the cooling applied based on either a user-specified set point for the average supply air temperature or the average return air temperature. Given the lack of knowledge of the heat load distribution for these types of simulation, such a model is a fit-for-purpose approach that can distinguish between basic assumption changes, such as return air temperature control versus supply air temperature control.

For more detailed analysis, especially detailed assessment or operational management, it is important to a use a more representative model of the chosen cooling unit and any associated controls. The following are critical to the effective prediction of the cooling performance:

1. A representative location for the cooling sensor(s): temperature distribution can vary dramatically over a small distance. For example, it is not unusual for the air temperature at the return to a down flow unit to vary by 5°C across the entire return area (Figure 17.9).
2. Careful representation of the outlet/supply airflow distribution: particularly, since the introduction of radial blowers can provide a very different air distribution (depending on configuration) compared with more conventional centrifugal blowers (Figure 17.10).
3. The ability to vary airflow and cooling in response to controls based on feedback from temperature, pressure, or flow sensors, as well as accounting for the physical capabilities of the system (e.g., the fan curve).
4. The ability for cooling capacity to vary based on on-coil conditions: during failure scenarios as the return air temperature rises and, therefore, the air temperature reaching the coil rises. This increased air temperature and the consequent rise in coolant temperature often result in increased capacity due to higher temperature differences at the coil and/or at the external heat exchanger.
5. Flow distribution from local cooling resources such as in-row coolers: the pattern and velocity distribution of the supply air can dramatically alter how much room air mixes with the cool air entering the cold aisle.

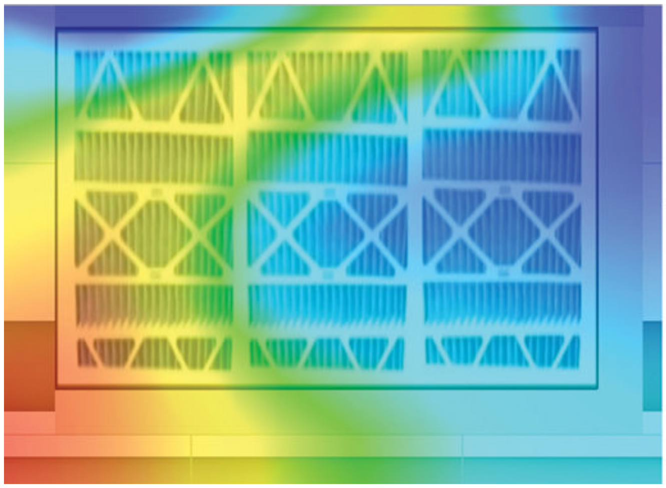


FIGURE 17.9 Typical return air temperature variation at a down flow unit return. Courtesy of Future Facilities. Please visit companion website to access the color version of this figure.

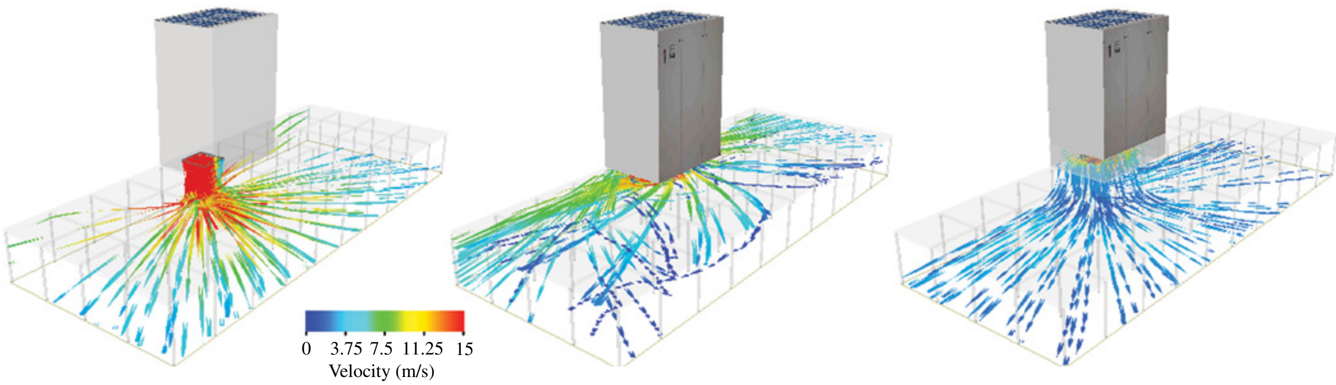


FIGURE 17.10 Typical flow patterns—centrifugal blowers (left) and radial blowers (center and right). Courtesy of Future Facilities. Please visit companion website to access the color version of this figure.

17.4.3 Other Infrastructure

The presence of other infrastructure in the data hall can critically affect the cooling performance, but this is easily overlooked. This can be important because the infrastructure uses cooling resources that would otherwise be intended for cooling the IT equipment. Alternatively, the infrastructure may be important because of its geometry and its impact on the path of the cooling airflow.

17.4.3.1 Use of Cooling Resources Some infrastructure that requires cooling will normally take it from the air intended to cool the IT equipment. This is particularly so when uninterruptible power supply (UPS), power distribution unit (PDU), and transformer systems are placed inside the data hall. Likewise, other heat sources such as lighting should also be considered, even if only from the point of view of their impact on reducing the cooling capacity available for IT.

For example, in-room PDUs can be cooled directly from the raised floor, with some units having pressure-driven flow through openings in the bottom to vents in the top to cool internal components such as transformers. Other PDUs may have built-in cooling systems that draw air from the room and return it to the room.

It is also worth noting that these infrastructure items, because of their weight, often stand on frames that are mounted on the floor slab rather than being mounted on the raised floor itself. If insufficient care is taken during installation, unintended and uncontrolled leakage can occur through the hole in the raised floor under and around the infrastructure item. To make detailed assessments and undertake operational management using CFD, it is important to inspect and account for these systems.

17.4.3.2 Obstructions to Airflow In addition to the items of infrastructural equipment that are typically placed above the raised floor, the common items to allow for are cooling pipes, power cables, and data cables. These were historically installed in the raised floor and can significantly impact the air distribution throughout the room. Problems with airflow often resulted from poor cable management and the raised floor becoming blocked.

The natural response to these issues has tended to be an overreaction—entirely removing almost all of these obstructions from the raised floor (at least the power and data cable elements) to leave the raised floor available as an air distribution plenum. CFD analysis can, however, be used to assess whether this is indeed the appropriate thing to do; experience suggests that some degree of distributed obstruction is helpful in achieving a more uniform airflow through raised floor perforated tiles. This is because they help to break up the highly directional airflows from the cooling units, thus generating a more uniform static pressure throughout the raised floor plenum.

For conceptual design, it is normally straightforward to account for infrastructural obstructions to airflow. However, for more detailed analysis and detailed assessment, troubleshooting and operational management in particular, more care is required in order to capture the true influence.

Structured pipe and cable routes are normally laid out during design, with only the main routes being critical in a conceptual design model.

Cooling Pipes The main cooling pipes are often significant enough in size to be included individually. Individual branches to each cooling unit are often ignored for conceptual design and may also be ignored even in more detailed analysis. Whether they are included will depend on their size and importantly their number and location. The decision whether or not to include smaller pipes can usually be made based on whether they, together with any other objects close by, represent any significant obstruction to airflow.

Small-percentage obstructions of the air path by one-off circular cross-sectional pipes are unlikely to significantly affect air distribution. The placement of the main cooling pipes in the path of the cooling jet from the cooling unit is likely to significantly affect the path and penetration of the jet.

Power Cables Power distribution can be made in several forms, from individual cables to significant-sized power conduits. Power conduits can be treated in a similar manner to cooling pipes. However, care must be taken to account for conduit junction boxes. These can be significant in size and, when placed in a raised floor, may block half the height of the raised floor and significantly disturb the local airflows.

Power cables, when laid in bundles or groups on the raised floor to be distributed around the data hall, are normally satisfactorily represented by solid obstructions capturing the height, width, length, and depth of the bundles. Often, the most difficult power cables to model are groups of cables descending vertically from the PDU or UPS to the solid floor. They are difficult to model because they are often in a loose bundle where air can pass between the cables. Visual inspection of the bundle will give the appearance of the region being more heavily blocked than it really is because of the visual obscuration. Consider the example shown later.

In one direction, cables are separated from each other by 150% of the cable diameter (Fig. 17.11 *left*). From the perpendicular direction, the cables are separated by the same distance as the cable diameter. Taking any row in either direction, the open area is 50% or more. Inspecting the bundle from the side from some directions would show the gaps to be blocked by the next row of cables (Fig. 17.11 *center*), while from others the bundles may appear to have a small gap (Fig. 17.11 *left*). In actual fact, a calculation will show that the volume is actually only around 32% blocked (Fig. 17.11 *right*). So, the judgments made using visual inspection should be treated with care!

If your CFD tool of choice does not have a loose cable bundle object, then use its porous obstruction modeling object to represent the partially obstructed volume. Coefficients for general purpose volume obstruction of this nature can be found in [1].

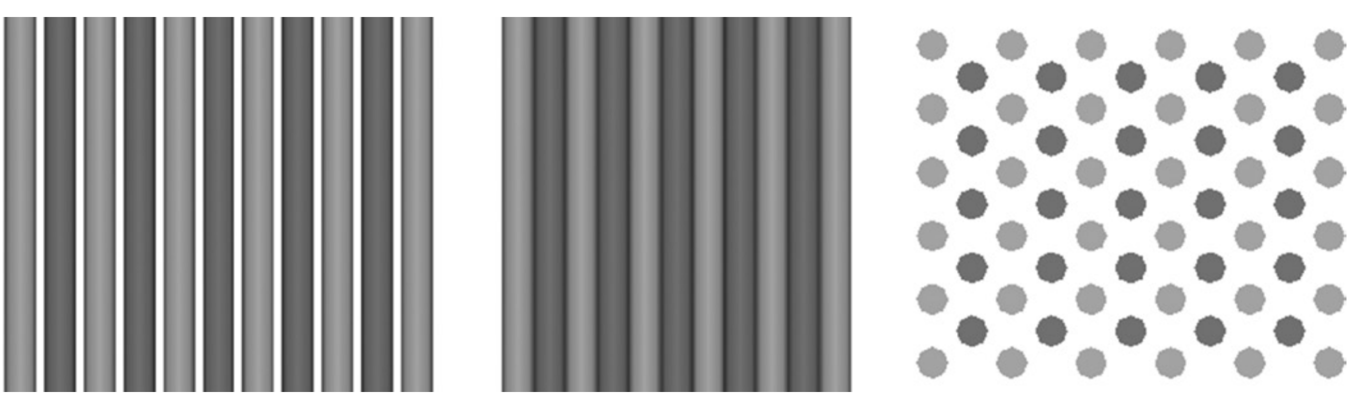


FIGURE 17.11 Staggered array of cylindrical obstructions. Courtesy of Future Facilities.

Data Cables For conceptual modeling, as the number and size of data cables will not be known in detail, it is normal to include a simplified obstruction represented as a solid: a bundle of cable tied together. For structured cabling, this is a reasonable representation since, unlike power cables, data cables are typically much more slender and can be bundled together and be changed in direction on a reasonable radius of curvature while still retaining the bundle properties. In some instances, it is necessary to also include the cable tray that follows the cable route and supports the cable bundles. This is particularly so when the cable tray has solid faces, either as a result of a sheet metal construction or because the wire mesh construction is supplemented by solid sheet of a material (like Correx, a corrugated polypropylene sheet) to prevent damage that may otherwise occur when cables are supported on a large open mesh wire basket tray.

For more realistic modeling in assessment, troubleshooting, or operational management, it is normal to capture a more realistic distribution of cable densities based on what is actually installed. Like conceptual design, cable route modeling is relatively straightforward when the cabling is structured. However, where the cabling is not structured (Figure 17.12), (particularly so for legacy data centers), then a similar challenge occurs in determining the true degree of obstruction when the cables are haphazardly deployed.

Data hall-specific modeling tools, particularly when they are being used for operational management, including the cable routes, can also have the added benefit that the tool may be able to automatically route cables when new equipment is added to the models. This itself has a number of additional benefits, including better representation of the cable obstructions and a calculation of the cable lengths required.

17.4.4 White Space/IT Equipment

Clearly of great importance, modeling the white space is critical to the purpose of CFD for data centers. This area of modeling poses the greatest dilemma to the data center CFD modeler. It is known that the characteristics of the IT equipment installed can change not only the local flow and heat transfer but also the overall cooling performance within the data hall. Yet, at the design stage for a data center, it is nearly always the case that little, or nothing, will be known about the actual IT equipment that will be installed. During conceptual design, it is therefore common to assume the following:

- The equipment ventilates front to back.
- The cabinets are filled with IT equipment.
- The entire data hall is populated with cabinets in which equipment is installed to the average design limit of heat load per cabinet.

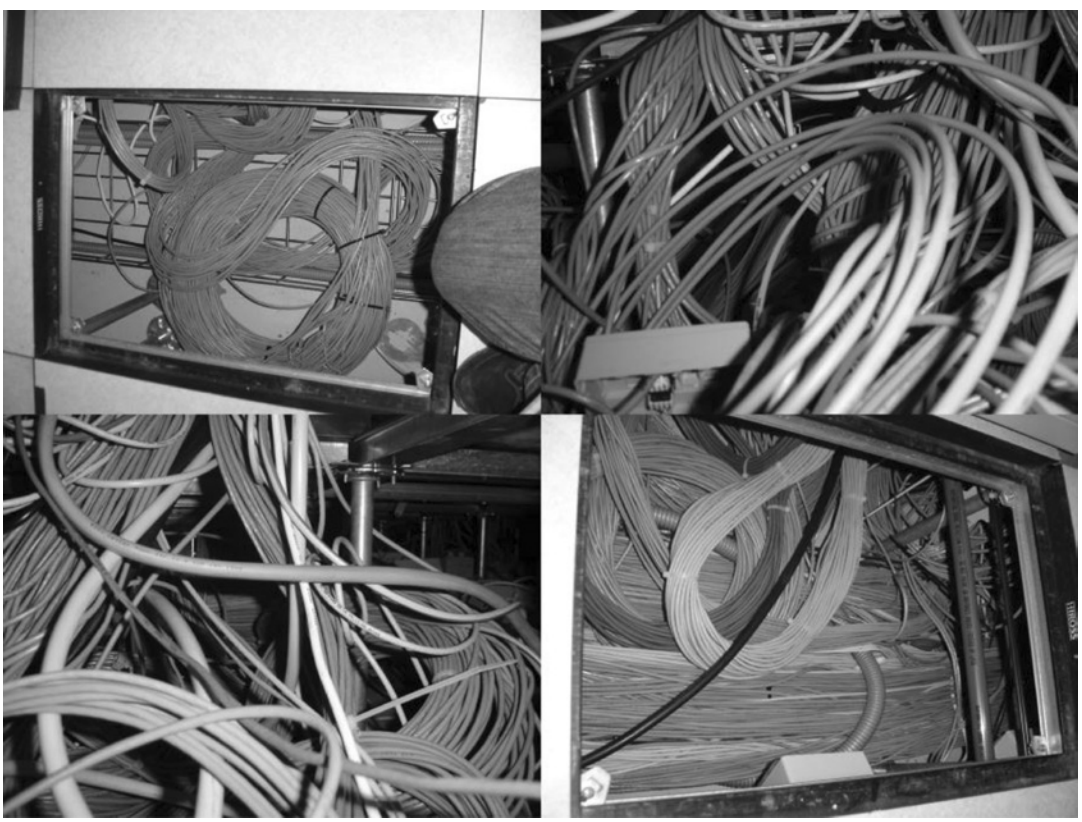


FIGURE 17.12 Classic unstructured cabling. Courtesy of Future Facilities.

- A nominal airflow per kW of IT load typically of the order of 120 cfm/kW (56.67 l/s/kW) of IT load.
- There is no opportunity for warm air recirculation in or under the cabinet.

Given the lack of certainty about the IT equipment to be installed, it is reasonable to make such gross assumptions provided that the user of CFD for data center conceptual design recognizes that the cooling design may be sensitive to variety of design parameters. A more comprehensive and reliable design study can be undertaken if the sensitivity studies are made for the following:

1. Airflow rate per kW—typical values range from about 80 cfm/kW (modern high-density equipment operating at high utilization) to 160 cfm/kW (low-power legacy equipment or modern equipment operating at low utilization). This is important because it can test the ability of the cooling system to match IT demand in terms of air flow (m^3/s) as well as power (kW).
2. How dense the equipment is in the cabinet—is the design sensitive to whether the cabinet is loaded over its entire height or whether it is half filled at the bottom or the top of the cabinet?
3. Variation of heat load per cabinet in different zones of the data center to represent some high-density zones and some low-density areas (but with the same overall average cabinet power density). This tests the ability of the cooling to adapt to more realistic nonuniform equipment distributions.
4. Partially loaded scenarios to understand how the design will cope when it is only part occupied in the early years.
5. Cooling system failure scenarios to investigate whether the redundancy strategy employed is effective.

A classic modeling error for IT equipment is to assume that the airflow exhausted from the IT equipment is distributed across the full face of the cabinet or at least the full area of the perforated door. This often results in unrealistically low momentum from the IT equipment, and so the warm exhaust air tends to rise more rapidly and travel less distance horizontally than it will in reality.

17.4.4.1 IT Equipment Detail When modeling real installations for detailed assessment or operational management, more attention has to be paid to the specific IT equipment. One of the key challenges is that the IT equipment will have variable heat output depending on its utilization. It is also likely to have variable airflow that can be dependent on the utilization but also on the operating environment and its consequent impact on the temperature of key electronic components. Further, knowing the manufacturer and model of many items of IT equipment is an insufficient specification since many platforms can be configured not only in terms of memory installed but also in terms of the number and type of processors, I/O cards, etc.

For most deployment decisions, what the IT manager or facility manager is most interested in are the maximum demands that will be placed on the facility infrastructure. As a consequence, most of the readily available data for IT equipment power and airflow is based upon these maximum requirements that, in normal operation, are never realized. Hence, for detailed design, assessment, and operational management, it is important that the IT equipment model behaves appropriately for the conditions that are likely to be present. As a consequence, some enterprise organizations will bench test their standard configurations of hardware running their chosen applications to better understand their operational characteristics.

The adoption of measures to make the data center more energy efficient is making the need for such bench test data all the more important. For example, the rise in ambient operating temperature makes it more likely that temperature thresholds that result in IT airflow changes will be reached. Also, the adoption of aisle containment requires greater attention on airflow balance between the IT equipment and cooling system.

Ideally, any IT equipment will be defined by the following:

1. The physical geometry of the IT equipment
2. The location and size of inlets and outlets
3. The power consumption of IT equipment and any dependence upon configuration and utilization
4. The airflow and its dependence upon pressure and temperature
5. Distribution of airflow and heat dissipation into separated flow paths through the IT equipment

Without this data, what can practically be done to achieve a useful model of the IT equipment?

- In the worst case, where design/modeling has to be done based entirely on available published data
  - Use nominal operating power consumption, if provided. Otherwise, use the name plate power (the quoted maximum power consumption) provided for electrical safety factored by a multiplier to reflect the fact that this power is not normally achieved in normal operation. For legacy equipment, this factor can be as low as 25%, but typically, modelers use a factor of 50%, which reflects increasing utilization and the current state of technology.
  - Use the published nominal flow rate and, where no data is published, assume a flow per kW. Typically, modeling often assumes a value of

about 56.67 l/s/kW (120 cfm/kW), but values can vary dramatically. Modern high-density equipment can have flow rates as low as 35 l/s/kW (and the current trend is toward these lower flow rates), while older/low-utilization equipment may have flow rates as high as 75 l/s/kW. Indeed, this latter figure can be considerably higher if the equipment is low power.

- Where there is access to an operating data hall
  - Use measured data for power to improve the estimate of power consumption at the IT equipment. In almost any data hall, the power consumption is known at PDU level. With the advent of intelligent power strips, power consumption is now often monitored down to the cabinet power strip level and even the individual socket level. In any case, the data available should be used at its most refined level to adjust the estimated power dissipation to match the measured data. So, if a cabinet has an estimated power consumption of 3.2 kW but the measured power consumption is only 2.4 kW, then the power for each item of IT equipment in the cabinet should be reduced to ¾ (2.4/3.2) of the estimated value. This has the added advantage of enabling a much more realistic power distribution in the virtual facility model.
  - Take sample measurements of inlet and outlet temperatures on selected items of IT equipment and, given the power consumed, use the temperature difference to check on the airflow. So, for example, a server using 400 W of power with a measured temperature difference of 12°C has a flow rate of 27.9 l/s based on the following formula:

$$Q = \dot{m} C p \Delta T$$

where Q is the power in Watts, $\dot{m}$ is the mass flow rate in kg/s (which in turn can be converted to volume flow rate by dividing by the air density), Cp is the specific heat capacity of air in J/kg K, and ΔT is the temperature rise in °C (or K).

To get the flow rate per kW (1000 W), simply multiply the flow rate by 1000/400: in this case, 69.7 l/s/kW.

17.4.4.2 Cabinet Detail As has been said, for conceptual design, it is normal for modelers to assume that the cabinet prevents recirculation from the hot exhaust side of the equipment to the cooler inlet side (and that it does not allow cool air bypass). In this scenario, many simulation tools assume that certain vents on the cabinet are inflows and others are outflows, even though, in practice, a perforated vent can easily allow both in different locations.

Further, some argue that this is a sufficient assumption for assessment and operational management, given that one is primarily interested in providing a good room (operational) environment for the IT equipment. However, in the author's opinion, this is not the case:

- First, the primary interest is not the room conditions but the local conditions for each and every piece of IT equipment.
- Second, if recirculation or bypass occurs in the cabinet (Figure 17.13), then the air volume passing through the cabinet and the temperature rise across the cabinet from inlet to outlet will not be the same as for the IT equipment itself.
- Finally, the inlet conditions at the IT equipment inlets are likely to be different from the surrounding room conditions.

Let us consider an example where 20% of the air flowing through the IT equipment recirculates inside the cabinet. As a consequence, only 80% of the airflow demand from the IT equipment is taken from the room. This means that the average inlet temperature will be increased by 25% of the IT equipment temperature rise due to the recirculation air supplying a quarter as much air as there is cooling air. It also means that the exhausted air volume from the cabinet will be only 80% of the IT equipment flow rate but at 25% greater temperature rise from the room inlet condition than the temperature rise in the IT equipment itself.

It is therefore critical for detailed assessment, troubleshooting, and operational management to model the potential

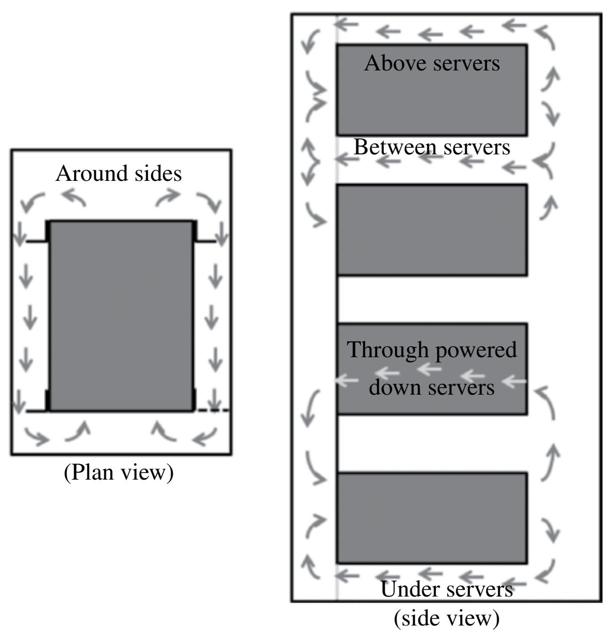


FIGURE 17.13 Schematic of internal cabinet recirculation. Courtesy of Future Facilities.

for recirculation and bypass inside the cabinet. The most common locations for this to occur are as follows:

- Through empty unblanked slots in the mounting rails
- Around the sides of the mounting rails
- Under the cabinet and/or equipment in the bottom U-slot
- Over equipment in the top U-slot
- Between IT equipment that do not fully occupy complete U-slots
- Through IT equipment that is installed but that is switched off

One cabinet detail that is often overlooked is the presence or absence of cabinet sides separating adjacent cabinets from each other. This is particularly important if the cabinets have large open cable penetrations (Section 17.4.4.3) in the raised floor beneath them, a characteristic that normally only occurs in legacy installations. In such circumstances, high-velocity jets can generate a recirculation in the cabinets and the raised floor, drawing warm air out of the bottom of some cabinets before the mixed air reenters cabinets further down the row.

17.4.4.3 Cable Penetrations Cable penetrations in the cabinet itself are normally not particularly important when compared with the vents intended for ventilation in the cabinet sides and top, although there are some exceptions, for example, cabinets with little ventilation—cabinets with glass doors—but these are no longer normally used to house IT equipment with significant power.

However, the location, size, and management of cable penetrations in the raised floor beneath the equipment can (as mentioned in Section 17.4.1.2) significantly affect the room cooling system performance. They can allow as much as half the cooling to leak unintentionally into areas of the facility that do not require cooling—the back of cabinets and the hot aisle, for example. It is now generally accepted that such cable penetrations should be managed by limiting the airflow with the introduction of foam, brushes, or gaskets. However, claims for the efficiency of such seals are often not achieved. This is because a cable turning from a horizontal distribution path in the raised floor to the vertical distribution entering the cabinet tends to disrupt the seal. It is therefore important to include this leakage when creating a CFD model for assessment, troubleshooting, or operational management.

A role where CFD modeling can be particularly important in relation to cable penetrations is where the intention is to upgrade a facility by introducing cable penetration management in a legacy data center where the cable penetrations previously went unmanaged. While the rule of thumb that introducing management generally improves performance, unilateral introduction without analysis of the impact can be dangerous: in some cabinets, IT equipment may be relying on cooling arriving through the cable penetrations.

It should be noted that a CFD tool that does not offer the user any control over where cable penetrations are is of little value in this type of analysis. This is because a cable penetration discharging air into the inlet side of a cabinet will have a very different impact compared with a cable penetration discharging air into hot exhaust side of the cabinet, where it will probably offer no cooling value to the IT equipment at all. Similarly, care should be taken with tools that do not enable such discrimination to ensure that they do not overpredict cooling by assuming that air entering through cable penetrations is providing cooling when it is not.

17.4.5 Modeling Control Systems

Controls have already been discussed to some extent in relation to cooling systems (Section 17.4.2) and IT Equipment (Section 17.4.4.1), and it is indeed these two areas where control is most important. The primary challenge with CFD models is the replication of the configuration and behavior of the real control system.

For conceptual design studies, most CFD tools have simple control models that allow cooling, or airflow through cooling systems, to be controlled on the basis of the average return air temperature or the average supply air temperature. This is normally sufficient for conceptual studies, where details of the cooling systems, control systems, and IT equipment are unknown.

However, as the design proceeds into the detailed design phase or where the model represents a real facility and is to be used for studies of a number of different configurations or conditions, the response to the control system becomes more important. The following list (not exhaustive) represents features that a modeler might look for to enable better representation of the facility behavior in the model:

1. Point sensors, in addition to average sensors, that can be placed at a user specified location.
2. The ability for controller response to be based on multiple sensor values and to select functions such as average, minimum, maximum, difference or indeed other user-specified functions of the values.
3. The ability for multiple items to be controlled by a single controller.
4. For a controller to control based on sensor values for one variable (e.g., temperature) but the controller output or object response to be limited by another variable (e.g., pressure).
5. For the object response to be conditional on the conditions in which it operates. For example, when the controller output is demanding full cooling from a cooling unit, the maximum cooling should be dependent

on the air temperature onto the coil, the coil coolant parameters, and the air volume.

6. The ability to control items other than cooling systems, including IT equipment flow, fans, dampers, etc.
7. The ability to control multiple variables including temperature, airflow, pressure, and relative humidity/moisture content.
8. The ability to sense any of the variables and use them for control.
9. Linear controller response and more general controller response.

Although this list represents features that are already present in a variety of data hall systems, it is unlikely that any CFD tool currently provides all these features in an easy-to-use, data center-focused toolset, out of the box. However, this is probably not the main challenge, as tools are being enhanced rapidly and making these features more accessible.

The key issue is that equipment manufacturers tend to regard control systems as proprietary, and the control system response can be quite complex because of the number of fans and components monitored. As a consequence, equipment manufacturers tend not to openly publish the data for their equipment. This makes it difficult for software suppliers to produce good library models of equipment. Instead, they have to provide libraries with limited characterizations. The user should therefore not assume that just because a tool has the capability to represent all the characteristics of the library models supplied or built by other users, it will necessarily contain a complete characterization.

Even so, it is anticipated that, over time, market forces, and in particular pressure from end users of the equipment, will result in more data and better library models being available.

17.4.6 Low-Energy Designs

The focus on low-energy design is resulting in new technologies that need to be modeled using CFD. In fact, CFD in its general form has been used to model the fundamental processes used in low-energy design for many years. However, for data hall applications, it is more appropriate to use simplified models rather than to necessarily include the full detail of the entire physics. As a consequence, CFD tools for data centers are permanently evolving to capture the new approaches being used in the drive for lower-energy, more efficient designs.

At the time of writing this chapter, the technologies being employed for state-of-the-art data halls are largely associated with air-side and coolant-side economizers. The features that are now being introduced to data hall CFD tools include the following:

1. Water mist sprays and subsequent evaporative cooling. The model probably does not need to be a full model of the nozzle and droplet motion with full two-phase treatment of the liquid-to-gaseous transition. However, the model does need to account for where the moisture will be evaporated and how much will be evaporated since the evaporation process absorbs energy (heat), providing cooling that does not have to be provided by a chiller. This evaporative cooling, "adiabatic cooling," is often termed "free cooling" because it does not require a mechanical cooling system such as a chiller.
2. Wet media as a source of moisture with similar intention to water sprays: to predict the location and extent of adiabatic cooling that results from the change in phase of the water droplets into gaseous phase.
3. Mist eliminators—where adiabatic cooling is applied to the air side of the data hall cooling system, any excess droplets that are not evaporated in the air handling section could be carried by the airstream into the data hall. Consequently, mist eliminators are deployed to remove any excess droplets.
4. Controllable dampers and vents for dynamic control of recirculation and free cooling.

These technologies, and particularly the simplified models for them, are currently in their infancy for data hall modeling. As a consequence, options available vary widely from one CFD tool to another. The user should, therefore, consider carefully what they hope to achieve with the tools and should look for clarification of capability from the CFD tool supplier.

17.4.7 Challenges

As mentioned in Section 7.2, one of the main challenges to data hall modeling is that the size of enterprise data halls is such that to capture the full details requires detail on a scale that results in very large calculation grids. The consequence of this is the requirement for significant computer resources and the resultant lengthy simulation times. There are currently two solutions being pursued to try and get away from this challenge:

1. Using alternative numerical techniques that try to simplify the physics in order to streamline and speed up the calculation. While the advantage of this approach is speed, the disadvantage is that it requires expertise from the user to know either when the simplified approaches are likely to deliver acceptable predictions or what to look for in the predictions to determine whether or not they are likely to be valid for the scenario(s) being considered.
2. Simplifying the data center models using current state-of-the-art CFD tools but ignoring certain details that add complexity and, in the view of the user, do not need to be considered for the decisions under consideration.

Similar to a), the key benefit is solution speed, but the key disadvantage is that it relies on the user understanding the impact of the simplifications applied. They must also understand what decisions the simulation results can be used to make or can contribute to.

Given an acceptance of the intensive nature of current data center CFD, probably the next greatest challenge to CFD is accuracy of the predictions. There are now numerous examples of CFD being used effectively to predict conditions for enterprise data halls; indeed, the tools are now being applied in day-to-day operational management.

While CFD is by nature an approximation, introducing errors into the prediction as a result of its formulation and methodology (e.g., the numerical discretization to solve the equations and the empirical formulae to represent turbulent mixing), probably the biggest uncertainties occur in data center modeling as a result of a failure to truly describe the boundary conditions.

There are two key sources of error resulting from boundary condition description:

1. The use of simplified representations of the boundary conditions does not truly capture the behavior of the object because the true behavior is not well understood. A good example of this (which occurs for almost any data center using a raised floor to deliver cooling) is the modeling of perforated tiles.

   The traditional approach (as described briefly in Section 17.4.1.2) is to represent the perforated tile by a resistance to airflow to capture the pressure difference required to drive air through it. In practice, this is only part of the story because a real perforated tile is made up of several layers or elements.

   The top surface of the perforated tile does act, to a large extent, in the manner represented by a flow resistance to capture the pressure drop characteristic, but it also has other effects. It normally produces a localized region immediately above the perforated tile where the discrete jets form the individual openings in the tile coalesce. In this region, an additional pressure drop is produced, which can be recovered either by sucking additional air from the surroundings into the jet, leaving the perforated tile, or by the jet being forced to contract and slow down as it consolidates.

   Further, depending on the construction of this top surface, the direction of departure of this jet may vary from the underlying direction resulting from the approach direction and turning that occurs as the air passes through the perforated tile substructure. As implied by this latter statement, this top layer of the tile is not the only layer of significance. There is also normally a considerable substructure present for a perforated tile that is largely intended to provide the physical integrity required for the tile to withstand the loads created by heavy equipment being transported through the data hall.

   This substructure can dramatically influence how much air from the raised floor passes through the perforated tile and whether the flow direction under the raised floor plays a role in the air volume and leaving direction of airflow. In addition, a flow control damper layer may also be present, adding further geometric and flow complexity.

   At present, there are no scientifically documented methods for the creation of simplified models of such perforated floor tiles. Current research indicates that for a representative model, it is commonly necessary to add details of all the layers to a lesser or greater extent, depending on the tile construction. The user is, therefore, dependent on the CFD software provider to undertake characterization and provide simplified models on an individual perforated floor tile by perforated floor tile basis, or the user has to undertake this characterization himself.

   This is perhaps an extreme example of lack of understanding of characterization of objects, but in fact, when a user adopts a tailored CFD tool, they are accepting that each and every object characterization has been made appropriately.
2. The use of simplified representations of the boundary conditions does not adequately capture the true state of the data center. This is because the features that define it are either too detailed to survey and understand or too detailed to practically model using current CFD modeling techniques and typical hardware available to practitioners.

   A good example of this category is a group of unstructured cables. While in principle the user could survey and record the precise path of every cable, the time and expense of doing so, not to mention the computational expense of modeling simulation of such a detailed representation, make such a strategy completely unviable and the improved accuracy not justified.

   It is, however, this set of judgments—what details to include in a model—that will be critical to whether or not it is sufficiently representative for the task in hand. These same choices drive the need for model calibration for a real facility model (Section 17.3.4.1).

Other challenges that occur in data hall modeling result as an extension of the large scale of a data hall and the time-consuming nature of modeling and simulation. Key extensions that exacerbate this performance issue are as follows:

1. The high rate of change of IT equipment, meaning that it may not be practical to run a simulation or simulations for each and every proposed deployment

2. The variety of equipment types available, each with unique characteristics, making it difficult to maintain a drag-and-drop library of all possible models and configurations
3. The dependence of IT equipment heat dissipation and airflow rate on utilization, compounding the lack of information available to fully characterize equipment
4. The increased resource that would be required to fully consider the numerous redundant and failure scenarios that could occur

Despite all these challenges, CFD modeling is now a part of the data hall design and operational management, because it has been found to help in the delivery of more energy-efficient and lower-risk strategies.

17.4.8 Time-Dependent Simulation and Failure Scenarios

As mentioned, the Navier–Stokes equations and CFD by its nature are able to predict not only steady-state predictions of how a data hall will perform from a thermal perspective but also to allow for time-dependent variations. This is particularly important in the case of an entire cooling failure, but first, it is appropriate to consider less extreme failure scenarios.

17.4.8.1 Redundancy Failure Scenarios Mission-critical data halls and their support infrastructure are generally designed with redundancy in place to protect the IT equipment, processes, and data from failure. For cooling, the common strategy is to install additional cooling units so that should one or more fail, there is still sufficient cooling.

In a small data hall, it is normal to use $N+1$ redundancy. That is, if N cooling units are required to deliver the cooling air to the data hall, one additional unit needs to be installed "just in case." From an energy and maintenance perspective, it would normally be most efficient to operate with the additional unit in standby and to make wear and tear even for all units—to periodically change which unit is on standby. However, from a cooling perspective, this increases the risk. There will now be different $N+1$ cooling patterns for the data hall, and in principle, all configurations should be tested. Some may suggest that the changeover is another state that should be tested by running a time-dependent simulation. However, the changeover period should be relatively short lived, and unless the data hall is being operated right at the limit, it is unlikely that a cooling-induced failure will occur in the time for switchover between two acceptable cooling configurations.

For larger data halls, it is normal to include at least two redundant units so that while one is not in use due to maintenance, it is still possible for another unit to fail, leaving N units still operating. However, it will be apparent that as the number of redundant units increases, the number of flow configurations increases astronomically. Although in principle it would be easy to ask the computer to run simulations for every possible combination, this may not be practical. In any case, it is probably not worthwhile either, since the chance of an increasing number of units failing at the same time becomes less likely. In this instance, it is therefore more normal to use diagnostic evaluations of cooling performance under normal operation to look for mission-critical equipment that may be heavily reliant on one or two cooling units and to select a small number of critical failure scenarios to consider. Examples of commonly analyzed configurations include where a number of cooling units are bunched together because of the physical architecture and failure of the cooling units in a high-density area of the data hall.

Simulation of a $2N$ redundant cooling system should not be confused with the scenarios described earlier. In a $2N$ scenario, while the redundancy could be achieved by having twice as many cooling units as needed, the redundancy is commonly achieved by each N being cooled using independent cooling circuits. So, two independent failure scenarios should be considered, one for each of the two cooling circuits failing, leaving the units operational on the other cooling circuit.

In all these failure scenarios, it is important to recognize that when fully loaded, the data hall may provide quite an uneven load on the different cooling units. It is therefore important to ensure that the capacity variation, as a function of on-coil temperature for the cooling units, is known. Only then can the simulation behave accordingly should the on-coil temperature rise.

17.4.8.2 Total Cooling Failure Total cooling failure can only be addressed using a time-dependent solution. This is because the data center will change from being in a constantly cooled state to one where no cooling is actively applied. Accordingly, the model must continuously change with time.

For a time-dependent simulation, the CFD program will calculate the change in conditions over small steps in time. The solution is still iterative and consequently expensive from a computational perspective. However, it should not be imagined that one time step will take as long as one steady-state calculation: a single time step normally takes a small number of iterations to converge if the time step is chosen appropriately.

Simulation of total cooling failure in a data hall by CFD programs is a special case time-dependent simulation. It is undertaken quite frequently because of the critical need to understand whether the design is likely to manage such a catastrophic failure well. However, due to the way in which data halls are currently modeled in CFD programs, there are a number of limitations that many, if not all, CFD tools are likely to suffer from and that require special attention if the predictions are to be most useful.

In practice, it is likely that CFD simulations of the entire cooling system failure are likely to be conservative—they will predict faster temperature rise of the system than will occur in reality. It is also reasonable to assume that CFD tools designed for data hall modeling are more likely to have strategies in place to offset the unrealistic effects of modeling approximations and lack of data, and so they can reasonably be expected to perform better.

Classic simplifications and approximations that will need addressing for a realistic simulation include the following:

1. The data hall model knows little about the external elements of the cooling system such as the chilled water or refrigerant loops or chillers themselves. While pumps are still running, these systems will add thermal inertia to the system and slow down the rate of temperature rise.
2. The cooling systems in the data hall have considerable thermal inertia, particularly in the heat exchangers. If the fans are connected to standby power systems, they will continue to provide cooling. If not, the large fans continue to run for a short period due to the inertial mass in rotation. Even when stopped, if there are no nonreturn dampers fitted to the cooling units, cooling will still be transferred into the air as it falls through the coils under natural convection or is drawn through the coils by the IT equipment fans that are running on stand by (UPS power).
3. The building architecture included for a steady-state calculation of data hall conditions does not need to include the thermal inertia. So, when a failure scenario is modeled, special attention must be paid to these surfaces to allow for their thermal inertia.
4. Sheet metal/thin objects, such as cabinet walls, are normally modeled as thin. This is because their physical thickness does not need to be included explicitly to correctly represent the thermal resistance for a steady-state calculation. As a consequence, and in line with the architecture, special treatment will be required to account for the heat capacity of these items.
5. The IT equipment itself does not immediately heat up as a result of air temperature rise at the IT equipment air inlets. Consequently, the temperature rise at the IT equipment outlet will be delayed. As these items are treated as "black boxes," generating heat and airflow for steady-state normal operation analysis, the time delays occurring in the temperature rise resulting from what may be thousands of IT items cannot easily be represented without special treatment.

If a user intends to use a CFD tool for this type of analysis, it would be prudent to investigate if and how the tool addresses some of these issues and, if not, whether there are practical methods that can be applied to the CFD models to obtain more realistic results.

When a CFD tool is used for a time-dependent analysis, it will calculate full results for each time step. Given the size of each data set and the fact that the time step is likely to be of the order of 1 s, an analysis over several minutes would result in hundreds of data sets. It is, therefore, common practice to save the full data sets only for selected times rather than saving the data for every time step. The data is normally saved after completion of time steps in a list or sometimes simply at a specified time step frequency.

For every full set of simulation data saved for a selected time step, the user can use all the standard tools (Section 17.2.7) for inspecting the conditions in the data center as well as the data hall-specific analysis tools (Section 17.4.9). In addition, the result views can normally be animated so that the user can visualize the change in conditions over time.

In addition to views and data from the complete saved sets of results, a time dependent analysis usually automatically records the data at every time step for selected items. Data hall-specific CFD tools have the advantage that they are normally able to automatically identify key data to be recorded as full time histories, but the user can normally identify locations/data that they would like to record. In non-data hall-specific tools, the user is likely to have to identify all the data points that are to be stored.

These cooling failure-specific analyses are only likely to be available for data hall-specific CFD tools. However, most CFD tools have general time-dependent analysis capabilities. The advantage of these general time-dependent analyses is that they are not limited to cooling failure analysis and can be used for other variations over time. This includes changes in IT utilization and the consequential changes in heat dissipation and airflow. The disadvantage is that all the special treatments that are included automatically for cooling failure would have to be manually implemented to obtain an equivalent solution.

17.4.9 Data Hall-Specific Analysis

Once the software development team is focused on data centers rather than any other fluid flow and heat transfer application, it is natural to include some data center-specific analysis of the wealth of data that is produced by CFD.

There are a significant number of useful metrics that can be used to distill the data to something that the mechanical engineer, facility manager, or IT manager can quickly understand. Of course, the most critical interest is whether or not the IT is efficiently cooled. One can simply plot the mean or maximum inlet temperatures for each cabinet, but this still means the user has to compare the data with a reference (Figure 17.14).

Published indices such as ASHRAE Temperature Compliance can be calculated, but for a live data center, it may be more appropriate to test the inlet temperature against the manufacturer's recommendations. In this case, CFD software can produce a plot of how close to overheating the cabinet-mounted IT equipment is by comparing the maximum inlet temperature of the IT equipment with the maximum temperature recommended for the IT equipment by the manufacturer (Figure 17.15).

To understand and optimize cooling efficiency, it is helpful to know where the cooling is coming from and where it is going. Simple plots can be made showing the load on each cooling system (Figure 17.16).

Streamlines, as previously discussed, can be used more intelligently to trace air and show which air from which cooling sources goes into IT equipment inlets and which air is unused. This methodology can be extended to quantitatively measure the supply effectiveness of the cooling system delivering cool air to IT equipment. A similar methodology can be applied to the hot air exhausted by the equipment returning to a cooling system and how effective the cooling systems are at scavenging the hot air. The same methodology can be used to track how much air recirculates either inside the cabinet or in the main body of the room.

There are several quantitative measures:

1. The Supply Heat Index and Return Heat Index are displayed once the model has simulation results. These indices, developed by HP [2], have been adopted by such organizations as ASHRAE and Green Grid.
   - In simple terms, SHI is a measure of the extent to which cold air from the air condition unit (ACUs) is diluted by warm, recirculated air before it reaches the inlets of the IT Equipment. A value of zero indicates "perfect" behavior, that is, no dilution occurs, and the equipment inlet temperature is, therefore, equal to the ACU supply temperature.

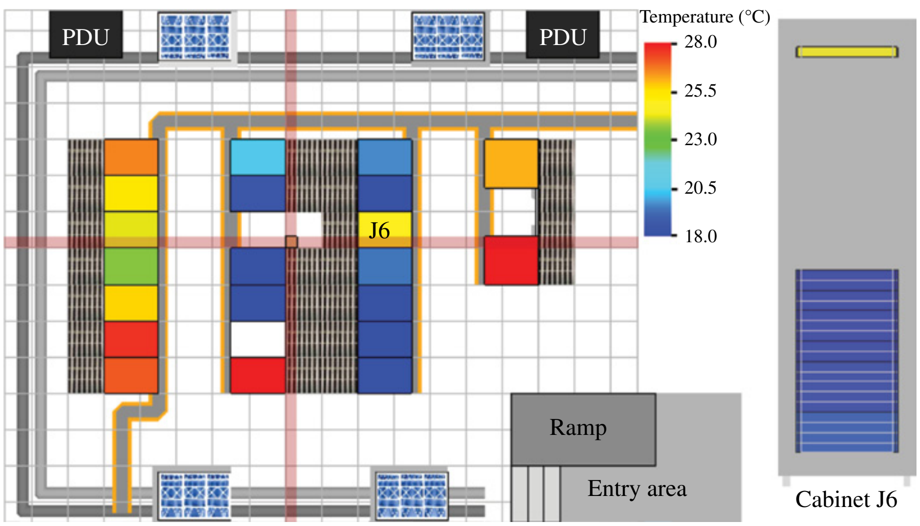


FIGURE 17.14 Maximum inlet temperature—room plan view (left) and individual cabinet view (right). Courtesy of Future Facilities. Please visit companion website to access the color version of this figure.

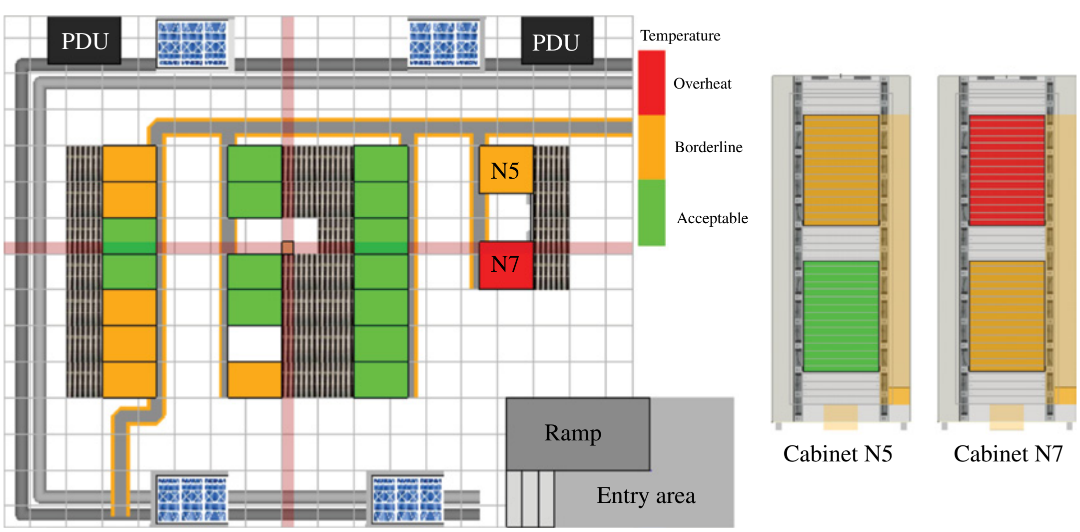


FIGURE 17.15 Plot of overheat risk. Courtesy of Future Facilities. Please visit companion website to access the color version of this figure.

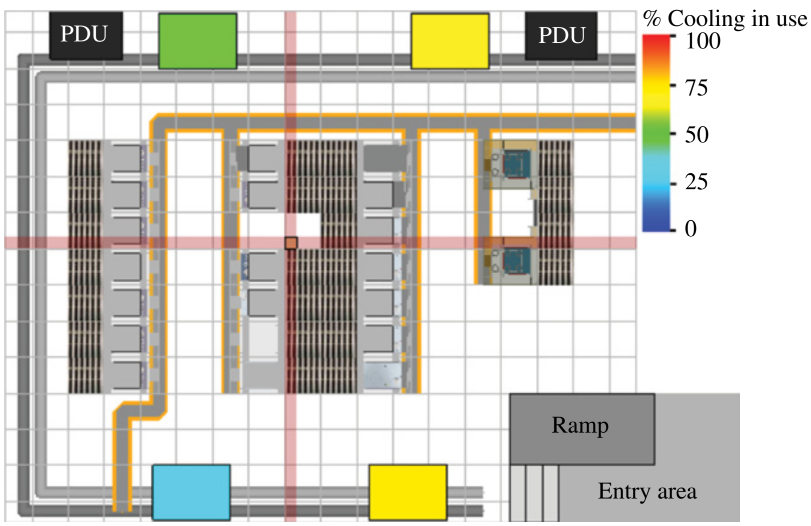


FIGURE 17.16 Cooling unit load. Courtesy of Future Facilities. Please visit companion website to access the color version of this figure.

- Similarly, RHI is an indication of how much of the cold supply air mixes into the ACU return air stream without ever reaching any IT Equipment. A value of 1 represents the "perfect" behavior the hot air returns undiluted to the equipment.

2. The Rack Cooling Index (RCI)® is a measure of how well the system cools the electronics within the manufacturers specifications, and Return Temperature Index (RTI)™ is a measure of the energy performance of the air management system. RCI® is a best practice performance metric for quantifying the conformance with IT equipment intake temperature guidelines such as those from ASHRAE and NEBS. RTI™ is a measure of net by-pass air or net recirculation air in the IT equipment room. The indices [3] can be used under license from ANCIS Incorporated (www.anis.us). RCI is calculated in two parts—RCI_{HI}™ and RCI_{LO}™:
   - $RCI_{HI} = [1 - (\text{Total Overtemp/Max Allowable Overtemp})] \times 100$ resulting in a %. A value of 100% represents ideal conditions: no overtemperatures. A value below 90% is often considered poor.
   - $RCI_{LO} = [1 - (\text{Total Undertemp/Max Allowable Undertemp}] \times 100$ resulting in a %. A value of 100% represents ideal conditions: no undertemperatures. A value below 90% is often considered poor.

   where:

   - Total Overtemperature represents a summation of over temperatures across all equipment intakes. An overtemperature condition exists when an intake temperature exceeds the Max Recommended temperature (ASHRAE Thermal Guidelines (2012) Class A1: 27°C).
   - Max Allowable Overtemperature is defined as Max Allowable temperature (32°C) minus the Max Recommended temperature (27°C) times total number of intakes.
   - Total Undertemperature represents a summation of undertemperatures across all equipment intakes. An undertemperature condition exists when an intake temperature drops below the Min Recommended temperature (ASHRAE Thermal Guidelines (2012) Class A1: 18°C).
   - Max Allowable Undertemperature is defined as Min Recommended temperature (18°C) minus the Min Allowable temperature (15°C) times the total number of intakes.

RTI is defined as

- $RTI = [(T_{Return} - T_{Supply})/(T_{EquipOut} - T_{EquipIn})] \times 100$ resulting in a %. A value between 80% (net by pass air) and 120% (net recirculation air) is often considered near-balanced airflow.

where:

- T_{Return} is the airflow (by volume) weighted average return air temperature to the cooling system(s).
- T_{Supply} is the airflow (by volume) weighted average supply air temperature from the cooling system(s).

  - $T_{EquipOut}$ is the airflow (by volume) weighted average equipment exhaust temperature from the IT equipment.
  - $T_{EquipIn}$ is the airflow (by volume) weighted average equipment inlet temperature to the IT equipment.
3. Simulation using CFD provides the inherent ability to trace where the air has come from and where it is going. Several indices utilize this functionality to characterize cooling system performance. Capture Index is the property of APC by Schneider Electric. Capture Index [4] includes the "Cold Aisle Capture Index (CACI)" and the "Hot Aisle Capture Index (HACI)," defined as follows:
  - CACI: the fraction of air ingested by a rack that originated from local cooling resources (e.g., perforated floor tiles, local overhead cooling supplies, or local coolers serving the same cold aisle cluster of equipment).
  - HACI: the fraction of air exhausted by a rack that is captured by local extracts (e.g., local coolers or return vents serving the same hot aisle cluster of equipment).
4. Cooling Unit Zone of Influence: Since the air can be traced to determine how much of it reaches which IT equipment, CFD can show which IT equipment each cooling systems affects. This is helpful in understanding the potential impact of a cooling system failure, although it should be noted that if even only one cooling unit fails, the airflow will change. Accordingly, this measure is only an indicator the actual effect.
5. Tailored analysis can also be extended to more general metrics: for example, if performance data is provided for external parts of the system such as the chiller and associated chilled water distribution systems, Power Usage Effectiveness (PUE) can be calculated. This is possible because the CFD model knows how much power is consumed by the IT, and using this with the external data and the internal calculated cooling system performance, the model can calculate how much power is used by the entire system. Similarly, the data can be used to summarize energy consumption and associated costs.

It is anticipated that as CFD is further adopted in day-to-day operational management and becomes more closely integrated with other tools, these purpose-specific analyses will be further developed.

17.5 POTENTIAL ADDITIONAL BENEFITS OF A CFD/VIRTUAL FACILITY MODEL

By definition, CFD models must describe the physical geometry of the facility, the cooling system configuration and operation, and the heat loads that are to be cooled. For a facility in operation, if the simulations are to be capable of prediction of the detailed cooling performance, it is likely that the model will need to contain significant details of the infrastructure and IT equipment alike.

As a consequence, the models can be easily extended to address other data center equipment deployment issues relating to space, power, network, and weight issues. They could easily be used as predictive DCIM tools or integrated with other DCIM tools, monitoring systems, or even IT application-based tools.

One way of defining an operational CFD model is to include every item of IT equipment explicitly in its correct cabinet and U slot, to effectively manage how much heat it dissipates, and to know its connectivity into the power system. The result is that the CFD model will automatically provide an inventory of the equipment located in the data hall.

Another aspect of CFD modeling is the need to determine how much power is dissipated as heat by the IT equipment in different locations. Given that the applications installed and the loading of the IT equipment will significantly affect the heat dissipation, one way of identifying the power consumption of a live facility is to use live power monitoring data. When this is done, the model will probably have the power network connectivity stored to enable it to make use of available data. In such a case, the CFD data hall model can be used to analyze the power system, including load balancing and the impact of single (or multiple) points of failure on availability, etc.

In a similar way, it is natural for the 3D model to be extended to hold data that it can be used as a DCIM tool. However, given the predictive character of a CFD tool, it lends itself to being used for capacity planning as a Predictive DCIM tool.

All that said, bearing in mind the complexity of data centers in terms of construction, logistics and operation, it is likely that separate tools will continue to be provided and used in corporate organizations for individual or small groups of tasks, including asset management and inventory, monitoring deployment, etc. It is therefore extremely valuable if the different tools can share data. The advantages in respect of the CFD model are severalfold:

1. IT inventory can be drawn from a chosen asset (or DCIM) tool.
2. Planning changes can be received from a chosen DCIM tool or transmitted to the chosen DCIM tool following analysis in the CFD tool.
3. Power network and consequent heat dissipation can be derived from a third-party power modeling and/or monitoring system.
4. Monitored data can be displayed in the 3D view provided by the CFD tool.
5. Cable routing can be made in the 3D model prior to implementation.

6. Alternative plans can be analyzed with a view to avoiding lost capacity by considering space, power, network weight, and cooling holistically.

If this type of approach is adopted where several tools share data, the end user can treat the applications as being a whole and the result is likely to provide a value greater than the sum of the individual parts.

17.6 THE FUTURE OF VIRTUAL FACILITY MODELS

As computers become more and more a part of our daily life, the demand for data centers is growing astronomically. One of the primary concerns about CFD is the significant compute requirement necessary to represent a live facility and simulate changes. This has resulted in many people and organizations exploring the possibilities of using simplified methodologies (such as potential flow) in order to achieve faster computing, in an ideal world simulation in real time. Fortunately, the very driver for computing in so much of our everyday lives is also providing the potential for rapid computation. One example is the rapid development of graphic processing units (GPUs) with hundreds of processors. As they are developed and gain access to sufficient on-board memory, they potentially offer the opportunity for massive parallelization (and consequent high-speed solution) on a relatively low-cost platform, negating the need for simplified solution methodologies.

Since it seems likely that the computational challenge will be overcome, how will the scope of these tools change? Will they be adopted for management? It is already clear that CFD tools can be and are being, used for more than one-off simulation of a proposed or existing facility. They have already been applied to the management of real facilities on an ongoing basis, focusing on deployment and capacity planning and accounting for space, power, network, and weight as well as cooling. Whether CFD tools become the core platform, or whether other management tools will absorb CFD tools, is not yet clear. What is certain is that an integrated toolset including CFD will be the norm. The toolset will almost certainly be broader, being the central reference for anything you need to know about your data centers around the world. It is also likely to go deep into the application layer as well as the facilities and physical layers.

Since CFD is also used to design much of the equipment being housed in a data center, it also seems likely that the simplified models used in the virtual facility models described here will be automatically generated and be an output of the equipment design process. This could substantially improve model quality and uniformity, making the modeling process much quicker and simpler.

One of the questions often posed about the future of CFD is, "will it continue to be necessary?" These questions have been raised because people see changing technologies as taking away the design and operation issues. In recent times, the question has arisen because of the use of hot and cold aisle containment. After all, to the uninitiated, physical segregation will prevent air mixing. However, in practice, segregation cannot be perfect and, if recirculation is allowed to take place, can be much more serious owing to the higher temperature of the undiluted return air. The emerging technology that reraises the question is that of liquid cooling. If liquid cooling is used in place of air, why would CFD be needed? In practice, liquid-cooled systems are desirable where power densities are high. Most, if not all, liquid cooling systems still leave some of the heat in the room to be carried away by the room cooling system, and so, with high-power densities, this may still be a significant cooling load. Further, the liquid cooling systems themselves will become large and complex: they are therefore likely to require some sort of simulation of their own for the purposes of optimization.

Further, the data hall is just one part of a data center. There are many other parts of the system that may warrant CFD, such as free cooling systems, generator halls, UPS rooms, etc. In summary then, in the view of the author, for the foreseeable future at least, CFD is here to stay.

REFERENCES

[1] Idelchik IE. *Handbook of Hydraulic Resistance*. 3rd ed. London: Hemisphere; 2005.

[2] Sharma RK, Bash CE, Patel CD. Dimensionless parameters for evaluation of thermal design and performance of large-scale data centres. American Institute of Aeronautics and Astronautics; 2002. Jaico Publishing House. AIAA-2002–3091.

[3] Herrlin M. Airflow and cooling performance of data centers: two performance metrics. [PDF] ASHRAE Trans 2008;114(2):182–187. Available at http://wwwhg.ancis.us/images/SL-08-018_Final.pdf. Accessed on January 24, 2013.

[4] Shrivastava SK, Van Gilder JW. Capture index: an airflow-based rack cooling performance metric. ASHRAE Trans 2007;113(1):126–136.

18

ENVIRONMENTAL CONTROL OF DATA CENTERS

Veerendra Mulay
Facebook, Inc., Menlo Park, CA, USA

18.1 DATA CENTER POWER TRENDS

In recent years, data center facilities have witnessed rapidly increasing power trends that continue to rise at an alarming rate. The combination of increased power dissipation and increased packaging density has led to substantial increases in chip and module heat flux. As a result, heat load per square feet of server footprint in a data center has increased. Recent heat loads published by ASHRAE [1] as shown in Figure 18.1 indicate that for the period 2000–2004, heat load for storage servers has doubled, while for the same period, heat load for computer servers has tripled.

According to these trends, compute server rack heat fluxes in 2006 were around 4000 W/ft^2. This corresponds to 27 kW for a typical 19 in. rack. There are 19 in. racks commercially available in markets that dissipate more than 30 kW, which corresponds to 4800 W/ft^2 rack heat flux.

18.2 THERMAL MANAGEMENT OF DATA CENTERS

This rapid increase in the heat load per server footprint has resulted in an equal increase in the research on how best to tackle this problem. Numerous research articles, papers, studies, and guidelines have been presented [1–106] that describe the work done in the area of thermal management of the data centers. Some of those topics from these articles are summarized as follows:

- The cooling system configuration
- The structural parameters
- The placement of Computer Room Air Conditioning (CRAC) units
- The energy management
- Data center metrics
- Modeling of data centers
- Experimental investigations of data center systems

18.2.1 The Cooling System Configuration

Nakao et al. [13] modeled four variations of the data center cooling configurations in their study. These included the underfloor supply with ceiling exhaust, underfloor supply with horizontal exhaust, overhead supply with underfloor exhaust, and overhead supply with horizontal exhaust. Noh et al. [14] modeled three variations of the data center designs for 5–6 kW rack loads for telecommunication applications. These configurations were underfloor supply with ceiling exhaust, overhead supply with underfloor exhaust, and overhead supply with wall exhaust. Both these studies concurred that the underfloor supply with ceiling return is the best alternative. The underfloor supply with wall exhaust is also a good option when the exhaust location is near the top.

Shrivastava et al. [15] studied different data center configurations. Computational Fluid Dynamics (CFD) models were constructed to assess the effectiveness of those configurations. They characterized the data center performance based on average region rack inlet temperature (RIT) and mean region RIT. They reported that for given constraints, underfloor supply ceiling return configuration was found to be the most effective. They also reported that among the supply air flow fraction, the ceiling height, and the location of the return vent, the supply airflow fraction is the most

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

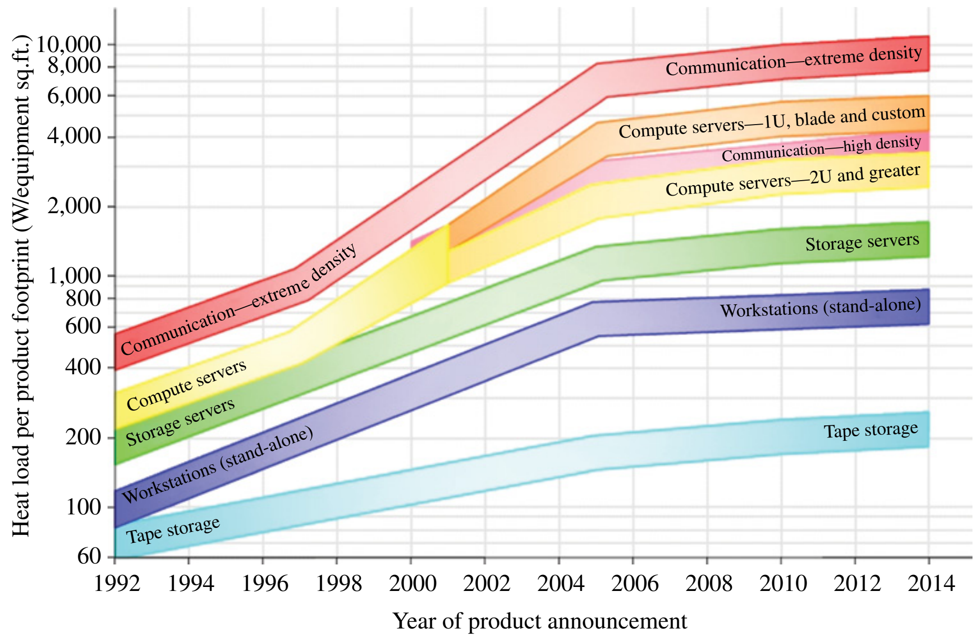


FIGURE 18.1 Heat load trends [1]. Courtesy of ASHRAE.

influential factor on RIT for various configurations. They also agreed with Nakao et al. [13] that the overhead supply with underfloor return represents the worst performing cooling configuration.

The suitability of airflow configurations for high-density data center clusters was discussed by Schmidt and Iyengar [6]. They considered two airflow configurations, namely, underfloor and overhead, which are prominently used. CFD models for both configurations are constructed, and the data is compared with respect to air supply fraction, rack location, and along the height of rack. They found high-temperature gradients in the inlet temperatures. They observed that these temperature gradients are, in some cases, more pronounced in underfloor configuration than the overhead supply design.

Similar studies were presented by Sorell et al. [16], Herrlin and Belady [17], and Mulay et al. [18] to compare the underfloor supply configuration with overhead supply configuration. Sorell et al. [16] were in agreement with Herrlin and Belady [17] that the underfloor supply can result in hot spots at the top of server racks due to severe recirculation patterns. The overhead supply design eliminates this drawback as the supplied air from top provides good mixing. Mulay et al. [18–20] presented the study of the two supply configurations in liquid cooling environment. Their observations that the underfloor supply configuration is preferred for high-power clusters over the overhead supply even in the liquid cooling environment agree with Schmidt and Iyengar.

A cooling technique requiring the control mechanism inside the racks was presented by Furihata et al. [21] and Hayama et al. [22, 23]. The control mechanism would monitor the temperature of exhaust air and then would adjust the airflow to yield the uniform temperature of the air exiting from servers. This technique resulted in reduction of airflow while providing adequate cooling to the servers.

In general, there is a strong consensus about the cold aisle–hot aisle layout. Beaty and Davidson [24, 25] and Beaty and Schmidt [26] recommend that the racks should be laid out in cold aisle–hot aisle arrangement, with racks drawing the air from cold aisle and releasing the hot air into hot aisle. Also, directing the hot air toward ceiling would be much better than simply having the high ceilings as observed by Beaty and Davidson [24]. Mulay et al. [27] presented the cabinet designs that agree with Beaty and Davidson [24].

18.2.2 The Structural Parameters

The structural parameters that can affect the airflow distribution are the plenum depth, the percentage opening of the perforated tiles, and the ceiling height. The plenum depth (the height to which the floor is raised) and the open area of a perforated tile are the crucial factors in determining the underfloor pressure distribution. The ceiling height may very well depend on the cooling configuration and may affect the scheme if not properly optimized.

18.2.2.1 The Plenum Depth With increased plenum depth, the velocities are reduced, which leads to more uniform subfloor pressure and subsequently uniform airflow distribution. A study conducted by Kang et al. [28] demonstrates the accuracy of pressurized plenum model with reference to CFD analysis. The CFD analysis of recirculating flow under plenum also indicates the limitations of validity of already-mentioned pressurized plenum model. The authors have used flow network modeling technique to predict flow rates of perforated tiles.

Karki et al. [29] presented the simulations of raised floor configuration with 25% open perforated tiles. They showed that relatively low plenum depth of up to 1 ft may lead to reverse flow near the CRAC units in some cases. However, when plenum depth is increased, the reverse flow is eliminated. Also, the flow variations across the tiles are reduced. Patankar and Karki [30] and Beaty and Davidson [25] present the cases that suggest a plenum depth with obstruction-free height of 2 ft. The plenum depth of 2 ft recommended by VanGilder and Schmidt [31] also falls in line with the suggestions by others as mentioned earlier.

Bhopte et al. [32] proposed multivariable approach to achieve optimal layout that will yield minimum RIT. The authors discussed the effect of plenum depth, floor tile placement, and ceiling height on RIT. These variables are used in multivariable optimization approach to study their interaction and the combined effect on the airflow distribution. The results are presented in the form of guidelines for optimal data center layout. These guidelines confirm that the airflow distribution becomes more uniform with increased plenum depth.

18.2.2.2 The Ceiling Height The ceiling height, among other factors, is dependent on the type of cooling configuration employed in the data center. A study performed by Schmidt [33] indicated the development of the hot spots over the racks where perforated tiles failed to deliver or exceed the required flow rate. For the underfloor supply configuration, these hot spots became more intense with the increased ceiling height. The increased ceiling heights lead to increased RIT.

In their parametric study, Shrivastava et al. [34] found out that the increased ceiling height has immense impact on the hot spot when reductions of up to 12°C were reported. The impact of ceiling height was, however, minimal in low-flux regions. In another study [35], the authors reported no impact on RIT when the ceiling height of a data center with underfloor supply and the room CRAC return was increased beyond 12 ft.

Sorell et al. [36] presented the three cases of cooling configurations with three different ceiling heights. The three configurations were underfloor supply with and without ceiling return and the overhead supply. The authors reported that with CRACs at 110%, increasing the ceiling height from 12 to 16 ft improved the performance of all the three data centers. They have also drawn the attention to the fact that increased ceiling height may lead to increased building costs.

18.2.2.3 The Perforated Tiles Schmidt [37] presented empirical and flow modeling data and a methodology to thermally characterize a data center. IT equipment power usage, airflow exiting perforated tiles, leakage flow escaping from cable cutouts, CRAC airflow, and air inlet temperatures were recorded for a 74 ft × 84 ft data center. Another study by Radmehr et al. [38] is focused on distributed leakage flow in raised floor data centers. The authors have outlined the procedure to measure airflow that escapes through the seams between panels, cable cutouts, and other gaps. The data is used to show the relationship between leakage area and the leakage flow. The authors reported leakage flow to be about 5–15% of the available cooling air.

Schmidt and Iyengar [39] measured IT equipment power usage, airflow exiting perforated tiles, leakage flow escaping from cable cutouts, and CRAC airflow and air inlet temperatures of three different data centers to study the patterns that will be helpful guidelines on data center layout. VanGilder and Schmidt [31], through the simulations of numerous raised floor data center models, quantified the impact of different parameters on the airflow distribution. They studied the factors such as underfloor blockages, tile layout, leakage flow, and the total airflow rate.

Bhopte et al. [40] presented a CFD model to demonstrate the impact of underfloor blockages on tile flow rates and RIT. They presented a parametric study identifying locations under floor where blockages, if installed, will have minimal effect on data center performance. Based on their case studies, the authors have presented guidelines on rearranging the blockages and still achieving improved performance.

18.2.3 Placement of the CRAC Units

The location of CRAC unit is an important factor in deciding the subfloor pressure distribution and will affect the airflow distribution in cold aisles. It has the potential of being the largest contributor to the energy inefficiency of the data center. Schmidt et al. [41] observed that the tuning vanes and baffles appeared to reduce the CRAC flow rate by 15%.

Koplin [42], in his study, indicated that the CRAC units should deliver the air in a way that will increase the subfloor pressure. When CRAC units are installed in parallel, they should not be so aligned that the plumes after delivery are colliding with one another, causing loss of static pressure. The study by Schmidt and Iyengar [43] agrees with this observation.

The work of Beaty and Davidson [25] and Schmidt and Iyengar [43] indicated the low inlet temperatures for those racks that have clear path for hot air from racks to the CRACs. They also recommended placing the CRAC facing hot aisles rather than facing the cold aisles.

18.2.4 Energy Management of the Data Centers

The increased energy consumption has equally opened up more opportunities for energy savings and efficient operations of the data centers. In their design guidelines, the

NREL [44] discusses, among other things on airside and waterside economizers, centralized air handling and liquid cooling. They have established the guidelines and some of these guidelines are discussed later.

18.2.4.1 The Airside Economizer An airside economizer uses outside air for cooling the data center when outside temperature is less than or equal to supply temperature. The cooler outer air is brought in and the hot air is exhausted into the ambient. In their proof-of-concept test, Intel IT has been running 900 production servers at very high rate of utilization [45]. This high-density data center used 100% air exchange at 90°F and without humidity restrictions. The filtration was kept at the minimal level. It was estimated that with economizer in use 91% of the time, 67% energy can be saved, which is estimated at US$2.87 million in a 10 MW data center. The proof-of-concept test by Intel also showed no significant rise in server failure rates when airside economizer is used.

A study by Shehabi et al. [46] compares the energy implications of conventional data centers with newer technologies employing waterside and airside economizers in five different climate zones in the state of California. They report that airside economizer performs consistently better in all climate zones. In fact, according to another study by Syska Hennessy Group [47], outside air can be used for almost an entire year in San Francisco.

18.2.4.2 The Waterside Economizer The waterside economizer uses the evaporative cooling capacity of a cooling tower and indirectly produces chilled water for data center cooling. Shehabi et al. [46], in their study, compared the five locations in California to judge the impact of waterside economizer. They observed that Sacramento has more potential benefits from waterside economization as compared to Los Angeles or San Francisco. The latent heat of the moisture content in San Francisco was overloading a chiller, causing another chiller to start to operate.

18.2.4.3 The Centralized Air Handling These centralized air handlers are ideal for the use of variable frequency drives, which enhances the part load efficiency. The centralized air handling units offer the following advantages over the conventional CRAC units:

1. They can be placed at some other locations than the data center rooms, thus freeing the space for IT equipment.
2. The sizing of centralized air handling unit can be designed to handle redundancy and reliability of operation.
3. The larger fans and equipment yield better efficiency.
4. The centralized systems have better part load efficiencies than the conventional CRAH units.

18.2.4.4 Liquid Cooling Schmidt et al. [47] reported the design of water-cooled rear door heat exchanger aimed to reduce exhaust air temperature in high-density racks. The impedance of rear door heat exchanger was reported to be matching with that of IBM standard rear door, thereby eliminating the need of extra fans. Mulay et al. [18–20], in their studies, studied the liquid cooling in data center for high-power clusters. They used different airflow supply fractions to study the impact of rear door heat exchanger. They also studied the deployment of rear door heat exchangers in both the overhead and the underfloor supply configuration for high-power density clusters. The rear door heat exchanger was found to be dissipating up to 55% of the heat.

The HP Modular Cooling Solution, as described in its Technology Brief [48], has three air-to-liquid heat exchangers and three hot swap blowers, which are mounted on the side of standard rack. The studies indicating substantial savings by the use of the liquid cooling in addition to the air cooling have been presented by Patel et al. [49], Schmidt et al. [47], and Leonard and Philips [50].

18.2.4.5 The Dynamic Cooling Patel et al. [51] introduced the concept of “Smart Cooling” by associating the local cooling to the work load allocation. With this holistic approach of cooling ensemble, the data centers would operate at the highest efficiency levels. Bash et al. [52] presented a distributed network of temperature sensors to provide real-time feedback to central controller. RIT at each rack is sensed. The temperature data is then used by controller to control the CRACs. This “Dynamic Smart Cooling” is shown to reduce power consumption.

Patel et al. [53] also discussed CRAC sizing and load balancing, rack layout, and the load distribution. In the study, the authors present the impact of the nonuniform nature of the heat load on the energy efficiency. The dynamic virtual data center and the algorithms to control the thermal management were presented by White and Abels [54].

18.3 COOLING SYSTEM DESIGN AND CONTROL

Facebook’s data center in Prineville, OR, has been one of the most energy-efficient data center facilities in the world since it became operational [107, 108]. Some of the innovative features of the electrical distribution system are direct current (DC) backup and high-voltage (480 VAC) distributions, which have eliminated the need for centralized UPS and 480–208 V transformation. The built-in penthouse houses the chiller-less air conditioning system that uses 100% airside economization and evaporative cooling to maintain the operating environment. These features have enabled significant reduction in energy consumption of the data center, which is reflected in Power Usage Effectiveness (PUE) of the facility. The PUE is defined as the ratio of total energy consumption of the data center to total energy consumed by IT equipment. The PUE of the Prineville data center is 1.07 at full load, which was verified during commissioning.

18.3.1 Data Center Design

The data center is a three-story building. The first floor holds the data hall and office space, along with the receiving yard and storage area. The second floor houses a large plenum for hot return air. The third floor is a built-up mechanical penthouse that holds the air handling equipment lineups. These lineups are divided into the intake corridor, the filter room, the Evaporative cooling/Humidification (EC/H) room, the fanwall room, the supply corridor, and, finally, the exhaust corridor. The airflow path is shown in Figure 18.2. This schematic also shows important points used in building management system (BMS) to control various components of air handling equipment lineups.

As indicated in Figure 18.3, the outside air enters into the intake corridor through vertical static louvers. These louvers have "S"-shaped cross section and are connected to a drain line. The cross section helps to keep the rain water or snow from getting into the corridor and facilitates the drainage of water so collected in drain pans. Figure 18.3 shows an inside view of the outside air intake corridor in which the vertical static louvers can be seen on the left side.

The outside air is then introduced into the filter room, which acts as a mixing chamber, as seen in Figure 18.4. On one side of this chamber are motorized dampers for outside air as well as return air. Depending on temperature and humidity of the outside air, these dampers modulate to vary

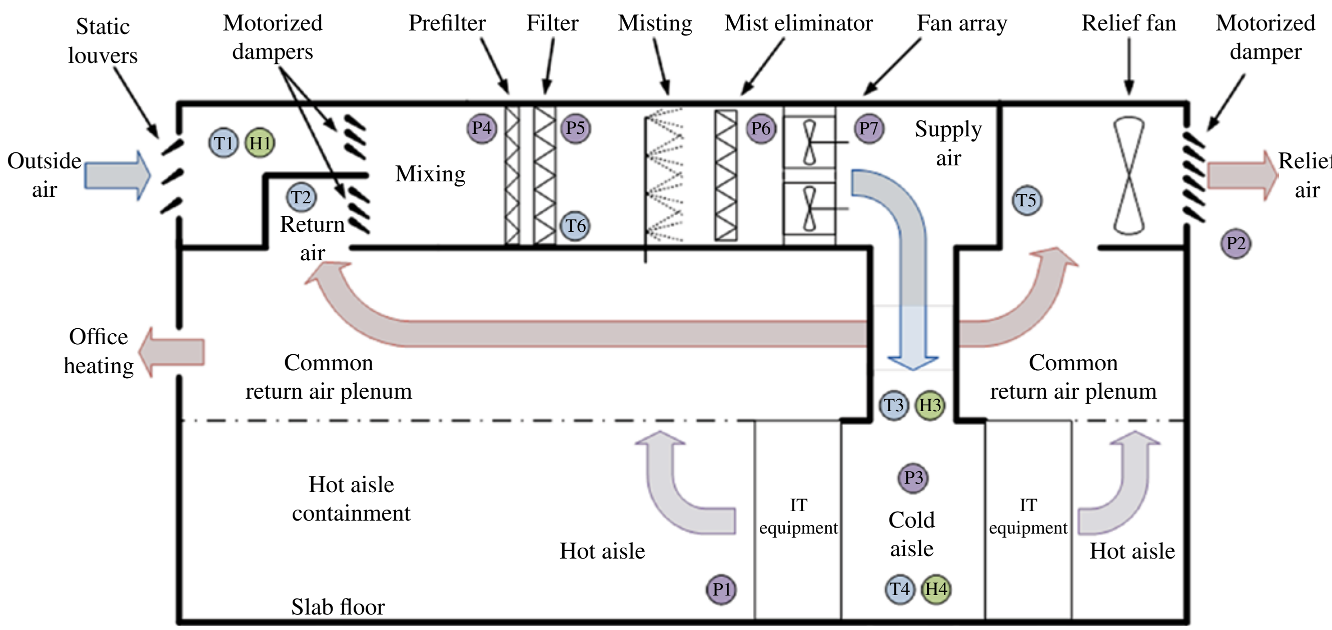


FIGURE 18.2 Side view of the data center indicating airflow path (L. Berkley, personal communication). Courtesy of Facebook.

FIGURE 18.3 Outside air intake corridor. Courtesy of Facebook.

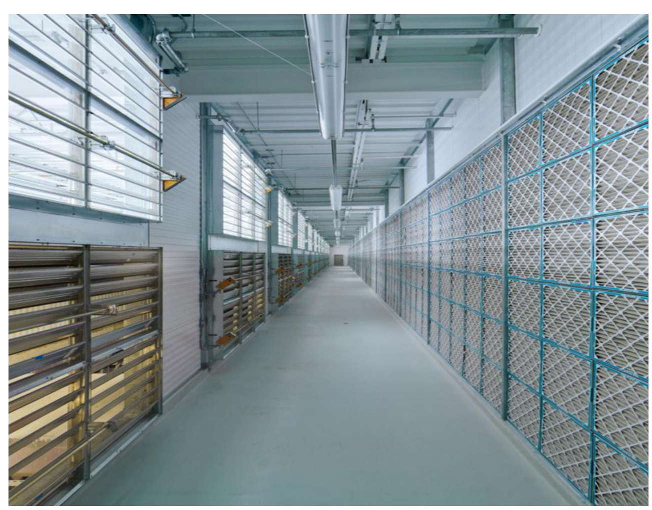


FIGURE 18.4 Filter room. Courtesy of Facebook.

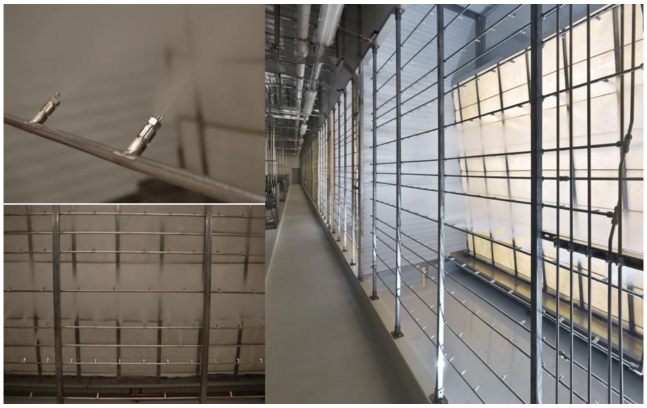


FIGURE 18.5 EC/H system. Courtesy of Facebook.

the proportion of outside air and return air. This mixed air then exits the mixing chamber through a filter wall. The filter wall consists of a 2 in. pleated prefilter followed by a filter of minimum efficiency reporting value (MERV) 13 (ASHRAE 85%). After passing through the tandem of these filters, the mixed air enters into the EC/H room.

The EC/H system uses high-pressure pumps and atomizing nozzles to spray atomized mist into the mixed air stream. The EC/H system has multiple stages to which the system modulates, depending upon the feedback temperature and humidity of the supply air. The water loop in this EC/H system is described later in another section. This sprayed air then passes through a mist eliminator media, which arrest any water atoms that are not evaporated, thereby preventing a moisture carry-over. The water thus collected in the drain pan of this mist eliminator is returned to the water loop for further processing and recirculation. Figure 18.5 shows the EC/H system in operation.

The next section of the air handling lineups is the fanwall section as depicted in Figure 18.6. The fanwall is essentially an array of plug fans assembled to form the matrix or a wall of fans. And this fanwall is actually pulling the air through all the aforementioned sections and then delivering it into the supply corridor.

The supply corridor contains supply shafts, which open into the data hall. The supply air is introduced into the data hall through these supply shafts.

In the data hall as shown in Figure 18.7, the cabinets are laid out in hot aisle–cold aisle arrangement. The hot aisles are contained to isolate the supply air from the air exiting the IT equipment, thereby avoiding the mixing of the two air streams as well as the recirculation of hot air and the bypass of the supply air. As such, the supply air then passes through the IT equipment and the hot air exits into the hot aisle. From hot aisle, the return air then enters into the return air plenum.

Once in return air plenum, the hot return air is introduced into the mixing chamber if the outside air conditions dictate it. The modulating dampers determine the quantity of the return air required for mixing, and the remainder of the hot return air is rejected to the atmosphere via relief fans in exhaust corridor as shown in Figure 18.8. In typical operation, these fans remain idle. During winter, the hot return air is used to partially heat the office space.

FIGURE 18.6 Fanwall section. Courtesy of Facebook.

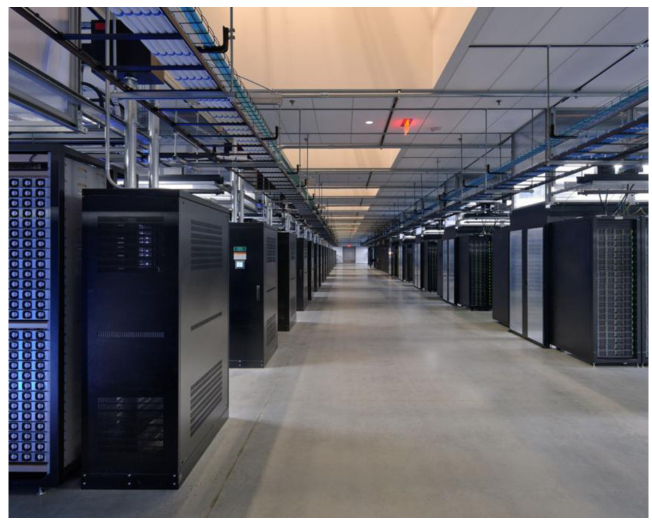


FIGURE 18.7 The data hall. Courtesy of Facebook.

18.3.2 Reverse Osmosis Makeup Water System

The water used in direct EC/H system is treated by a reverse osmosis (RO) process. The primary intent for the RO system is to take impurities out of the water to minimize the chances of clogging the misting nozzles, as the orifices on the misting nozzles are on the micron scale and the data center has several thousands of misting nozzles. Figure 18.9 shows schematic of the water loop. There are two sources of water feeding the RO system. The primary water supply is from a

FIGURE 18.8 Relief fans in exhaust corridor. Courtesy of Facebook.

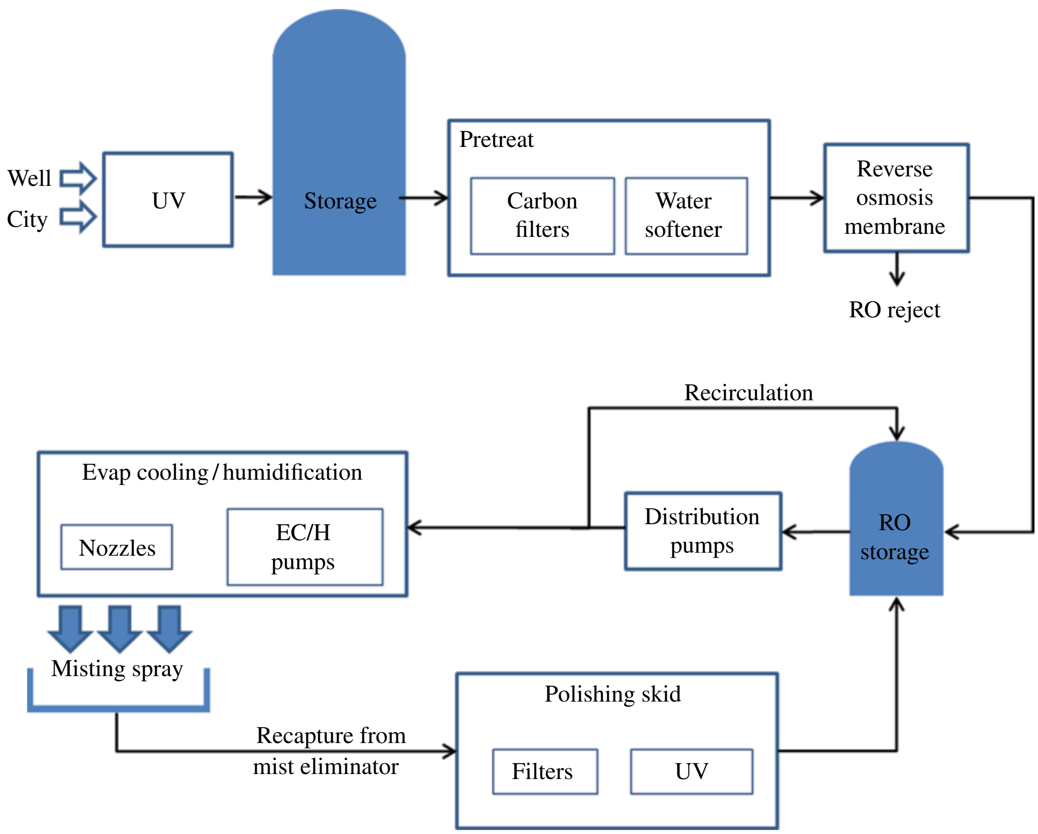


FIGURE 18.9 Water loop for EC/H system. Courtesy of Facebook.

well that was drilled on-site and the secondary is from the Prineville municipal water system.

The data center has two RO plants that supply half of the data center building each. The following is a description of the RO process for a single plant. An outdoor, aboveground, storage tank that is sized to support 48 h of operation during peak 50-year BIN weather conditions is filled via the on-site well as a primary source and via the city water as a secondary source. Furthermore, the outdoor storage tank is piped such that the water is intermittently circulated through an ultraviolet (UV) filter for the purpose of disinfecting and never allowing the water to stagnate for an extended period of time.

The next process occurs in the RO filter room. Booster pumps at 50 pound per square inch (psi), two pumps in parallel with a third redundant pump, and then pull water from the tank and pump the water through three sets of carbon filters and water softeners for the purpose of purifying and removing minerals from the water upstream of the RO membranes. Two RO pump skids, in parallel with a third skid, then receive the water and pump the water through the RO membranes at 50 psi, further removing large molecules and ions. It was found that the RO process makes two parts purified water from three parts well water. Next, the purified RO water is pumped into two RO Storage Tanks, sized for 1 h of operation in the event of an RO pump skid failure at peak load. Distribution pumps at 45 psi, two pumps in parallel with a third redundant pump, then circulate the RO water through another UV filter and up to the E/CH system pump skids in the mechanical penthouse. The EC/H pump skids then pressure the 45 psi water up to 1000 psi through the misting nozzles. Approximately 85% of the misted water is evaporated into the air stream, while 15% of the RO water is recaptured via a mist eliminator for the purpose of minimizing water carry-over. The 15% of water that isn't evaporated into the supply air stream is then brought back to the RO room, treated via a polishing skid with a UV and micron filter and then piped back into the RO storage tanks. The intent of reclaiming the RO water via the mist eliminator is to purify the RO water that has been potentially contaminated by the penthouse air stream. The air that has now been conditioned via the EC/H system is then supplied into the data center via fan arrays and dry wall shafts.

18.3.3 The Cooling System Basis of Design and Operational Envelope

The cooling system design was based on 50-year extreme data recorded for Redmond, OR, which is the closest weather station to Prineville, OR. In summer, the maximum dry bulb (DB) temperature recorded was 105.6°F, whereas the maximum wet bulb (WB) temperature recorded was 70.3°F. The winter extreme condition was recorded as −30.8°F DB temperature at 50% relative humidity (RH). This climate is advantageous for outside air and evaporative cooling; the coincident WB temperatures are low when the DB temperatures tend to be high, allowing free cooling most of the year and efficient use of evaporation when needed. As indicated on psychrometric chart in Figure 18.10, the system is sized to handle both these extreme conditions. In fact, DB temperature considered for summer design condition was 110°F instead of 105.6°F.

The supply air temperature in the data hall is controlled between 65 and 80°F. The moisture content is maintained between 65% at higher end and 41.9°F dew point (DP) temperature at the lower end. This operational envelope is compared with recommended operational envelopes by ASHRAE in Table 18.1.

From the table, we can see that the operating environment of Prineville data center is similar to ASHRAE's recommendations in 2008, except that the high-end moisture level is no more limited by the DP temperature.

18.3.4 The Cooling System Sequence of Operation

The psychrometric chart as indicated in Figure 18.11 is used to plot the state of air by using any two known properties such as DB temperature, WB temperature, DP temperature, RH, humidity ratio, etc. There are eight distinct operational regions as shown in Figure 18.11, which cover all possible outside air conditions. The sequence, in which the air handling lineups respond while in those regions, is as follows.

Region A
When outside air has a WB temperature lower than 52°F (11.1°C) and the DP temperature is below 41.9°F (5.5°C), the target supply air DB temperature is 65°F. The outside and return air dampers modulate to mix two airstreams. If required, EC/H system stages on to provide necessary humidification to maintain WB temperature of the supply air at 54°F (12.2°C) and the DP temperature at 41.9°F (5.5°C).

Region B
This region calls for 100% outside air. When WB temperature of the outside air is more than 52°F (11.1°C) and the DP temperature is below 41.9°F (5.5°C), the return air dampers are completely closed and the outside air dampers are fully open. EC/H stages on to provide required humidification or cooling. The supply air DB temperature is maintained between 65 and 80°F (18.3 and 26.7°C), while DP temperature is maintained at 41.9°F (5.5°C).

Region C
When the DB temperature of outside air is between 65 and 80°F (18.3 and 26.7°C), the DP temperature between 41.9 and 59°F (5.5 and 15°C), and the RH less than 65%, the outside air is delivered into data hall "as is" after filtration. In this region, again, the return air dampers are completely closed and the outside air dampers are fully open. One hundred percent outside air is admitted. The EC/H system remains OFF as no evaporative cooling or humidification is required.

Region D
When the DB temperature of outside air is exceeding 80°F (26.7°C), the WB temperature lower than 66°F (18.9°C), and its DP temperature above 41.9°F (5.5°C), the economizer is at 100%, meaning outside air without any mixing is admitted. EC/H stages on to provide required humidification or cooling. The supply air DB temperature is maintained at 80°F (26.7°C), while DP temperature is maintained between 41.9 and 59°F (5.5 and 15°C).

Region E
When the DB temperature of outside air is exceeding 80°F (26.7°C), the WB temperature exceeding 66°F (18.9°C),

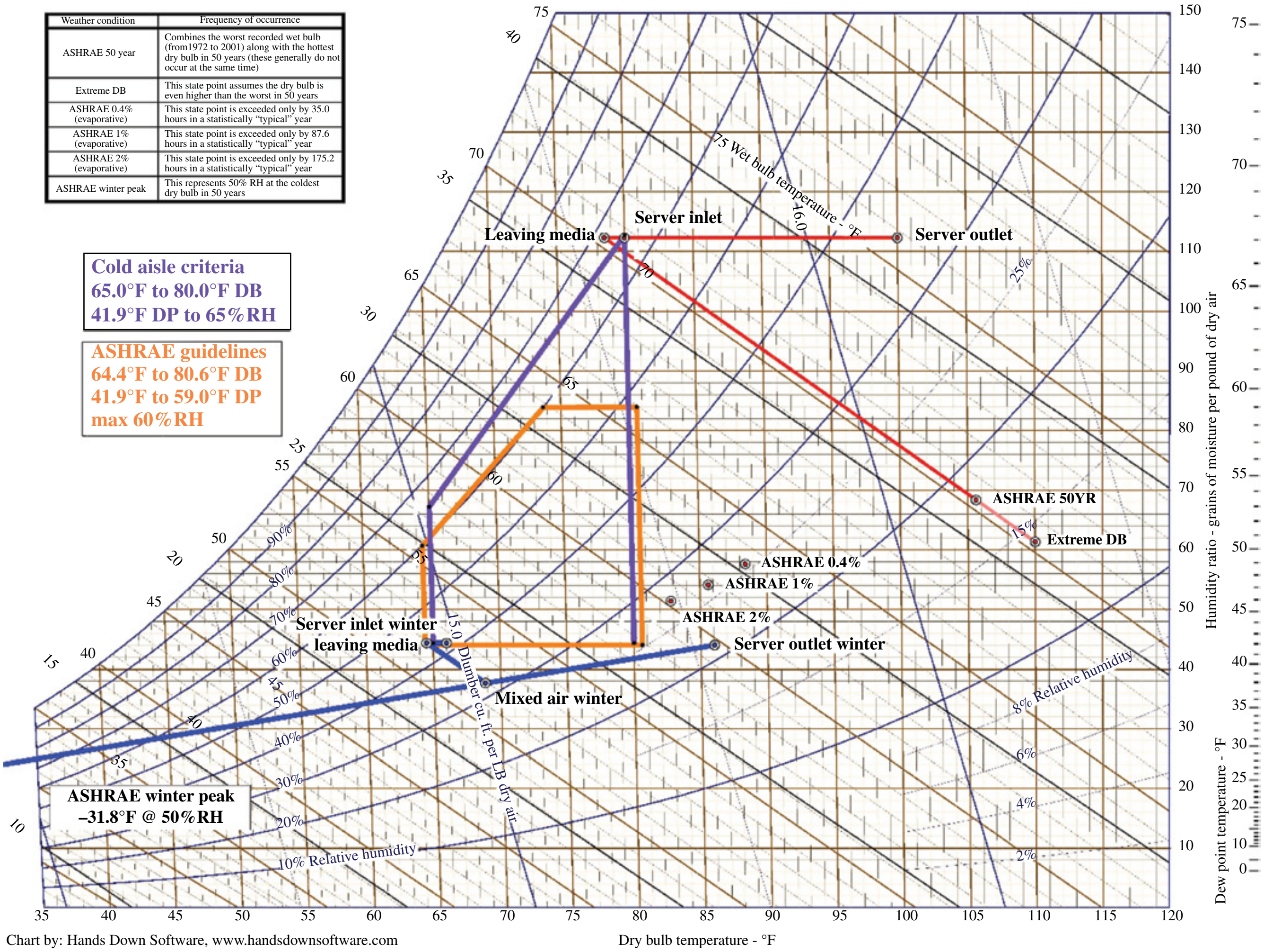


FIGURE 18.10 Operational envelope. Courtesy of Facebook.

TABLE 18.1 Operating environment of Prineville data center

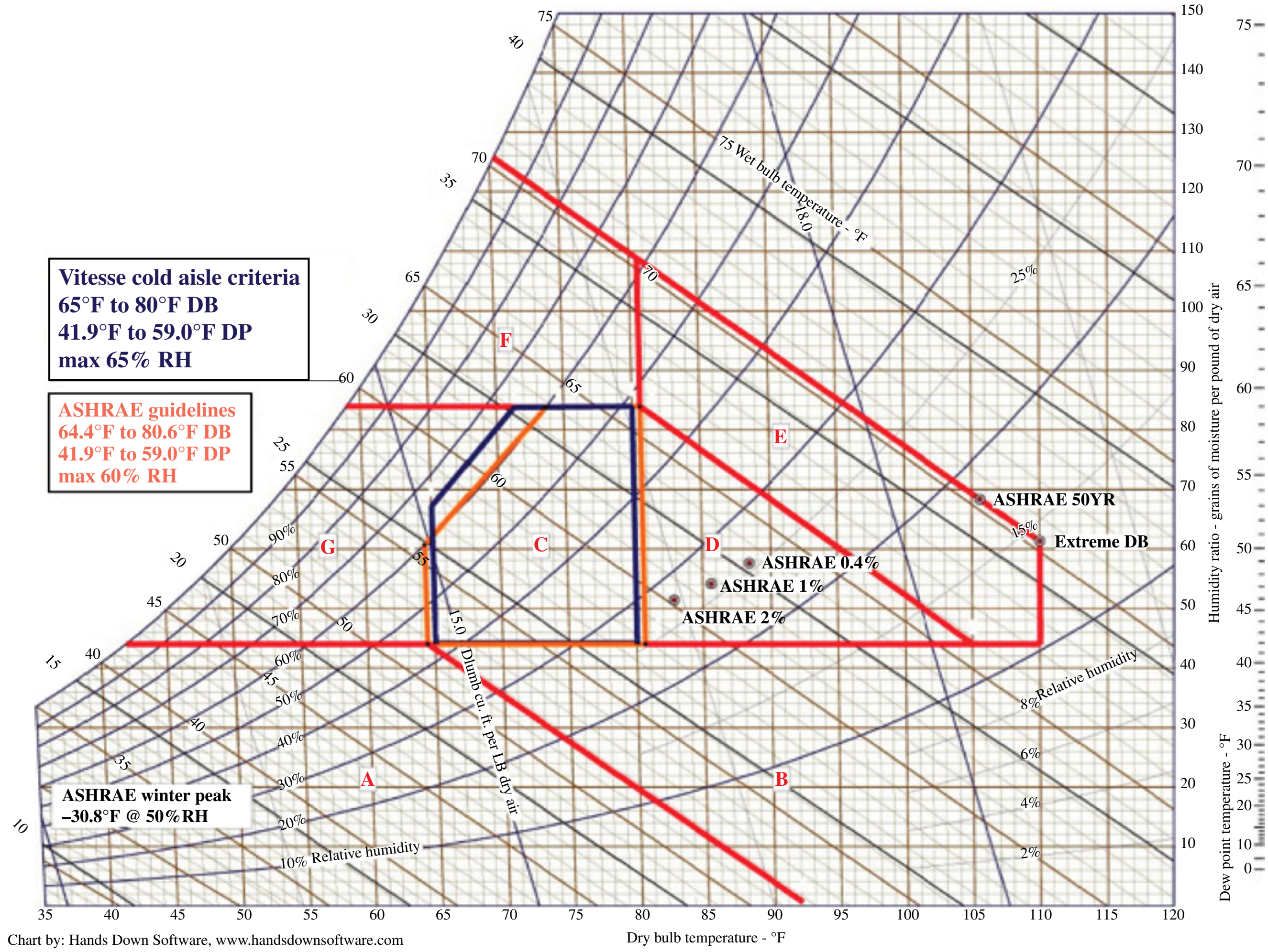
	ASHRAE 2004	ASHRAE 2008	Prineville DC
Temperature: low end	68°F (20°C)	64.4°F (18°C)	65°F (18.3°C)
Temperature: high end	77°F (25°C)	80.6°F (27°C)	80°F (26.7°C)
Moisture: low end	40% RH	41.9°F DP (5.5°C)	41.9°F DP (5.5°C)
Moisture: high end	55% RH	60% RH and 59°F DP (15°C)	65% RH

and its DP temperature above 41.9°F (5.5°C), the economizer is at 100%, meaning outside air without any mixing is admitted. EC/H stages on to provide required humidification or cooling. The supply air DB temperature is maintained at 80°F (26.7°C), while DP temperature is above 59°F (15°C).

Region F
When the DB temperature of outside air is lower than 80°F (26.7°C), the WB temperature less than 70.3°F (21.2°C), and its DP temperature above 59°F (15°C), the dampers modulate to mix outside air with return air to reduce supply air RH to 65% RH maximum. The supply air temperature will be maintained between 65 and 80°F (18.3 and 26.7°C). The DP temperature will be above 59°F (15°C). Direct evaporation system is bypassed. No evaporative cooling or humidification is required.

Region G
When the outside air has

1. DB temperature less than 65°F (18.3°C) and its DP temperature is between 41.9 and 59°F (5.5 and 15°C) or



Outside air operating conditions

FIGURE 18.11 Regions of operation. Courtesy of Facebook.

2. DB temperature exceeds 65°F (18.3°C), while its DP temperature is less than 59°F (15°C) and the RH is more than 65%, then

the dampers modulate to mix outside air with return air to increase cold aisle temperature as necessary to reduce supply air RH below 65%. The supply air temperature will be maintained above 65°F (18.3°C). The DP temperature will be below 59°F (15°C). Direct evaporation system is bypassed. No evaporative cooling or humidification is required.

Region H: Unacceptable OA Conditions (Smoke or Dust)
When outside air (OA) is inadmissible to the data center (such as excessive smoke or dust particulates in the air), the external dampers are closed.

18.4 PERFORMANCE METRICS

There are number of metrics that are used by data center professionals to measure the effectiveness of system performance. The PUE is the most commonly used metric, which was defined by the Green Grid as the ratio of total energy consumption of the data center to total energy consumed by IT equipment. The ideal PUE is 1.0 where all the energy supplied to a data center is consumed by the IT equipment. The EPA report to congress estimates the best practice PUE at 1.5. The Prineville data center with already-mentioned design and control scheme operates at much higher level of efficiency. At the end of Q3 2012, the 12-month trailing PUE for this data center was 1.09 [109].

Another metric defined by the Green Grid is the water usage effectiveness (WUE), which can be an indicator of how efficiently water is being consumed by the facility. It is defined as ratio of annual site water consumption to annual energy consumption of IT equipment. The Prineville data center began monitoring this metric in Q1 2012, and by end of Q3 2012, the trailing WUE for 6 months was 0.43 l/kWh [109].

REFERENCES

[1] ASHRAE. *Datacom Equipment Power Trends and Cooling Applications*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.; 2005.

[2] ASHRAE. *Thermal Guidelines for Data Processing Environments*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.; 2004.

[3] Schmidt R. Thermal profile of a high-density data center—methodology to thermally characterize a data center. ASHRAE Trans 2004;110(2):635–642.

[4] Schmidt R, Iyengar M, Beaty D, Shrivastava S. Thermal profile of a high-density data center—hot spot heat fluxes of 512 W/ft2. ASHRAE Trans 2005;111(2):765–777.

[5] Schmidt R, Iyengar M, Mayhugh S. Thermal profile of world's third fastest supercomputer. ASHRAE Trans 2006;112(2):209–219.

[6] Schmidt R, Iyengar M. Comparison between underfloor supply and overhead supply ventilation designs for data center high-density clusters. ASHRAE Trans 2007;113(1):115–125.

[7] Koomey JG. Estimating total power consumption by servers in the US and the world. Oakland: Analytics Press; 2007.

[8] Mitchell-Jackson J, Koomey JG, Nordman B, Blazek M. Energy needs in an internet economy: a closer look at data centers [Masters thesis]. Berkeley: University of California; 2001.

[9] Tschudi W, Xu T, Sartor D, Stein J. High performance data centers: a research roadmap. Berkeley: Lawrence Berkeley National Laboratory; 2004. Report nr LBNL 53483.

[10] Iyengar M, Schmidt R. Analytical modeling of energy consumption and thermal performance of data center cooling systems—from chip to environment. Proceedings of the ASME 2007 InterPACK Conference collocated with the ASME/JSME 2007 Thermal Engineering Heat Transfer Summer Conference, Volume 1; July 8–12; Vancouver, BC, Canada; 2007. p 877–886.

[11] Schmidt R, Shaukatullah H. Computer and telecommunications equipment room cooling: a review of literature. IEEE Trans Compon Packaging Technol 2003;26(1):89–98.

[12] Schmidt R, Iyengar M. Best practices for data center thermal and energy management: review of literature. ASHRAE Trans 2007;113:206.

[13] Nakao M, Hayama H, Nishioka M. Which cooling air supply system is better for a high heat density room: underfloor or overhead. Proc Int Telecomm Energy Conf 1991;12(4): 393–400.

[14] Noh H, Song K, Chun SK. The cooling characteristic on the air supply and return flow system in the telecommunication cabinet room. Proc Int Telecomm Energy Conf 1998;33(2): 777–784.

[15] Shrivastava S, Schmidt R, Sammakia B, Iyengar M. Comparative analysis of different data center airflow management configurations. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73234. p 329–336.

[16] Sorell V, Escalante S, Yang J. Comparison of overhead and underfloor air delivery systems in a data center environment using CFD modeling. ASHRAE Trans 2005;111(2):756–764.

[17] Herrlin M, Belady C. Gravity-assisted air mixing in data centers and how it affects the rack cooling effectiveness. Proceedings of the Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM '06); May 30–June 2; San Diego, CA 2006. p 438, 5 pp.

[18] Mulay V, Karajgikar S, Iyengar M, Agonafer D, Schmidt R. Computational study of hybrid cooling solution for thermal management of data centers. Proceedings of the ASME 2007 InterPACK Conference collocated with the ASME/JSME 2007 Thermal Engineering Heat Transfer Summer Conference, Volume 1; July 8–12; Vancouver, BC, Canada; 2007. Paper No. IPACK2007-33000. p 723–731.

[19] Mulay V, Karajgikar S, Agonafer D, Iyengar M, Schmidt R. Parametric study of hybrid cooling solution for thermal management of data centers. Proceedings of the ASME 2007 International Mechanical Engineering Congress and Exposition, Volume 8: Heat Transfer, Fluid Flows, and Thermal Systems (Parts A and B); November 11–15; Seattle, WA 2007. Paper No. IMECE2007-43761. p 519–526.

[20] Mulay V, Agonafer D, Schmidt R. Liquid cooling in data centers. Proceedings of the ASME 2008 International Mechanical Engineering Congress and Exposition, Volume 6: Electronics and Photonics; October 31–November 6; Boston, MA 2008. Paper No. IMECE2008-68743. p 95–101.

[21] Furihata Y, Hayama H, Enai M, Mori T. Efficient cooling system for IT equipment in a data center. Proceedings of the 25th International Telecommunications Energy Conference (INTELEC '03); 23–23 October; Yokohama, Japan. p 152–159.

[22] Hayama H, Enai M, Mori T, Kishita M. Planning of air-conditioning and circulation systems for data center. Proceedings of the 25th International Telecommunications Energy Conference (INTELEC '03); 23–23 October; Yokohama, Japan. p 140–146.

[23] Hayama H, Enai M, Mori T, Kishita M. Planning of air conditioning and circulation systems for data center. IEICE Trans Commun 2004;87(12):3443–3450.

[24] Beaty D, Davidson T. New guideline for data center cooling. ASHRAE J 2003;45(12):28–34.

[25] Beaty D, Davidson T. Data centers—datacom airflow patterns. ASHRAE J 2005;47(4):50–54.

[26] Beaty D, Schmidt R. Back to the future: liquid cooling data center considerations. ASHRAE J 2004;46(12):42–46.

[27] Mulay V, Agonafer D, Irwin G, Patell D. Effective thermal management of data centers using efficient cabinet designs. Proceedings of the ASME 2009 InterPACK Conference collocated with the ASME 2009 Summer Heat Transfer Conference and the ASME 2009 3rd International Conference on Energy Sustainability, Volume 2; July 19–23; San Francisco, CA 2009. Paper No. InterPACK2009-89351. p 993–999.

[28] Kang S, Schmidt R, Kelkar K, Radmehr A, Patankar S. A methodology for the design of perforated tiles in raised floor data centers using computational flow analysis. IEEE Trans Compon Packaging Technol 2001;24(2):177–183.

[29] Karki K, Patankar S, Radmehr A. Techniques for controlling airflow distribution in raised floor data centers. Proceedings of the ASME 2003 International Electronic Packaging Technical Conference and Exhibition, Volume 2; July 6–11; Maui, HI 2003. Paper No. IPACK2003-35282. p 621–628.

[30] Patankar SV, Karki KC. Distribution of cooling airflow in a raised flow data center. ASHRAE Trans 2004;110(2):629–634.

[31] VanGilder J, Schmidt R. Airflow uniformity through perforated tiles in a raised floor data center. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73375. p 493–501.

[32] Bhopte S, Agonafer D, Schmidt R, Sammakia B. Optimization of data center room layout to minimize rack inlet temperature. ASME J Electron Packag 2006;128(4):380–387.

[33] Schmidt R. Effect of data center characteristics on data processing equipment inlet temperatures. Proceedings of the Pacific Rim/ASME International Electronic Packaging Technical Conference and Exhibition (IPACK'01), Volume 2: Advances in Electronic Packaging; July 8–13; Kauai, HI 2001. Paper IPACK2001-15870. p 1097–1106.

[34] Shrivastava S, Iyengar M, Sammakia B, Schmidt R, VanGilder J. Experimental-numerical comparison for a high-density data center: hot spot heat fluxes in excess of 500 W/ft2. Proceedings of the Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM '06); May 30–June 2; San Diego, CA 2006. p 402–411..

[35] Shrivastava S, Sammakia B, Iyengar M, Schmidt R. Significance levels of factors for different airflow management configurations of data centers. Proceedings of the ASME 2005 International Mechanical Engineering Congress and Exposition; Heat Transfer, Part B; November 5–11; Orlando, FL 2005. Paper No. IMECE2005-81607. p 99–106..

[36] Sorell V, Abrogable Y, Khankari K, Gandhi V, Watve A. An analysis of the effects of ceiling height on air distribution in data centers. ASHRAE Trans 2006;112(1):623–631.

[37] Schmidt RR. Thermal profile of a high-density data center—methodology to thermally characterize a data center. ASHRAE Trans 2004;110(2):635.

[38] Radmehr A, Schmidt R, Karki K, Patankar S. Distributed leakage flow in raised-floor data centers. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73273. p 401–408.

[39] Schmidt R, Iyengar M. Effect of data center layout on rack inlet air temperatures. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73385. p 517–525..

[40] Bhopte S, Sammakia B, Iyengar M, Schmidt R. Guidelines on managing under floor blockages for improved data center performance. Proceedings of the ASME 2006 International Mechanical Engineering Congress and Exposition; Heat Transfer, Volume 3; November 5–10; Chicago, IL 2006. Paper No. IMECE2006-13711. p 83–91.

[41] Schmidt R, Karki K, Patankar S. Raised floor data center: perforated tile flow rates for various tile layouts. Proceedings of the Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM '04); Volume 1; 1–4 June; 2004. p 571–578.

[42] Koplin EC. Data center cooling. ASHRAE J 2003; 45(3):46–53.

[43] Schmidt R, Iyengar M. Effect of data center layout on rack inlet air temperatures. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73385. p 517–525.

[44] VanGeet O. FEMP best practices guide for energy efficient data center design. NREL Report nr NREL/BR-7A40-47201. Original publication February 2010; revised March 2011.

[45] Intel Information Technology. 2008. Reducing data center cost with an air economizer. Available at http://www.intel.com/content/www/us/en/data-center-efficiency/data-center-efficiency-xeon-reducing-data-center-cost-with-air-economizer-brief.htm. Accessed June 19, 2014.

[46] Shehabi A, Ganguly S, Traber K, Price H, Horvath A, Nazaroff W, Gadgil A. *Energy Implications of Economizer Use in California Data Centers*. Berkeley: Lawrence Berkley National Laboratory; 2008.

[47] Syska Hennessy. 2007. The Use of Outside Air Economizers In Data Center Environments. Available at http://www.syska.com/thought/whitepapers/wpabstract.asp?idWhitePaper=14 Accessed August 21, 2014.

[48] HP Technology Brief. 2006. HP Modular Cooling System: Technology overview and applications brief. Available at http://www.hp.com/go/mcs. Accessed May 23, 2014.

[49] Patel C, Bash C, Belady C. Computational fluid dynamics modeling of high compute density data centers to assure system inlet air specifications. Proceedings of the Pacific

Rim/ASME International Electronics Packaging Technical Conference and Exhibition (Inter-Pack); Kauai, HI; 2001. Paper IPACK2001–15622.

[50] Leonard PL, Phillips AL. Thermal bus opportunity—a quantum leap in data center cooling potential. ASHRAE Trans 2005;111(2):732–745.

[51] Patel C, Bash C, Sharma R, Beitelmal M, Friedrich R. Smart cooling of data centers. Proceedings of the ASME 2003 International Electronic Packaging Technical Conference and Exhibition, Volume 2; July 6–11; Maui, HI 2003. Paper No. IPACK2003-35059. p 129–137.

[52] Bash C, Patel C, Sharma R. Dynamic thermal management of air cooled data centers. Proceedings of the Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM '06); May 30–June 2; San Diego, CA 2006. p 452, 8 pp.

[53] Patel C, Sharma R, Bash C, Beitelmal A. Thermal considerations in cooling large scale high compute density data centers. Proceedings of the Eighth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM 2002); San Diego, CA; 2002. p 767–776.

[54] White R, Abels T. Energy resource management in the virtual data center. 2004 IEEE International Symposium on Electronics and the Environment. Conference Record; May 10–13; Scottsdale, AZ 2004. p 112–116.

[55] Anton R, Jonsson H, Palm B. Modeling of air conditioning systems for cooling of data centers. Proceedings of the Eighth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM 2002); San Diego, CA; 2002. p 552–558.

[56] ASHRAE. *Design Considerations for Datacom Equipment Centers*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.; 2005.

[57] Baer D. Managing data center heat density. HPAC Eng 2004;76(2):44–47.

[58] Bash C, Patel C, Sharma R. Dynamic thermal management of air cooled data centers. Proceedings of the Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM '06); May 30–June 2; San Diego, CA 2006. p 452, 8 pp.

[59] Boucher T, Auslander D, Bash C, Federspiel C, Patel C. Viability of dynamic cooling control in a data center environment. Proceedings of the Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM '04); Volume 1; 1–4 June; Las Vegas, NV 2004. p 593–600.

[60] Beaty D. Cooling data centers with raised-floor plenums. HPAC Eng 2005;77(9):58–65.

[61] Beaty D, Chauhan N, Dyer D. High density cooling of data centers and telecom facilities—part 1. ASHRAE Trans 2005;111(1):921–931.

[62] Beaty D, Chauhan N, Dyer D. High density cooling of data centers and telecom facilities—part 2. ASHRAE Trans 2005;111(1):932–944.

[63] Bedekar V, Karajgikar S, Agonafer D, Iyengar M, Schmidt R. Effect of CRAC location on fixed rack layout. Proceedings of the Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM '06); May 30–June 2; San Diego, CA 2006. p 425, 5 pp.

[64] Belady C, Beaty D. Data centers: roadmap for datacom cooling. ASHRAE J 2005;47(12):52–55.

[65] Belady C, Malone C. Data center power projections to 2014. Proceedings of the Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM '06); May 30–June 2; San Diego, CA 2006. p 439–444.

[66] Bhopte S, Schmidt R, Agonafer D, Sammakia B. Optimization of data center room layout to minimize rack inlet air temperature. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73027. p 33–41.

[67] Bhopte S, Sammakia B, Schmidt R, Iyengar M, Agonafer D. Effect of under floor blockages on data center performance. Proceedings of the Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM '06); May 30–June 2; San Diego, CA 2006. p 426–433.

[68] Furihata Y, Hayama H, Enai M, Mori T, Kishita M. Improving the efficiency of cooling systems in data centers considering equipment characteristics. Proceedings of the 26th Annual International Telecommunications Energy Conference (INTELEC 2004); September 19–23; Chicago, IL 2004. p 32–37.

[69] Furihata Y, Hayama H, Enai M, Mori T. The effect air intake format of equipment gives to air conditioning systems in a data center. IEICE Trans Commun 2004;87(12):3568–3575.

[70] Guggari S, Agonafer D, Belady C, Stahl L. A hybrid methodology for the optimization of data center room layout. Proceedings of the ASME 2003 International Electronic Packaging Technical Conference and Exhibition, Volume 2; July 6–11; Maui, HI 2003. Paper No. IPACK2003-35273. p 605–612.

[71] Hamann H, Lacey J, O'Boyle M, Schmidt R, Iyengar M. Rapid 3-dimensional thermal characterization of large scale computing facilities. IEEE Trans Comp Packaging Technol 2005;31(2):444–448.

[72] Herold K, Rademacher R. Integrated power and cooling systems for data centers. Proceedings of the Eighth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM 2002); 2002. p 808–811.

[73] Herrlin MK. Rack cooling effectiveness in data centers and telecom central offices: the rack cooling index (RCI). ASHRAE Trans 2005;111(2):725–731.

[74] Heydari A, Sabounchi P. Refrigeration assisted spot cooling of a high heat density data center. Proceedings of the Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM '04); Volume 1; June 1–4; 2004. p 601–606.

[75] Iyengar M, Schmidt R, Sharma A, McVicker G, Shrivastava S, Sri-Jayantha S, Anemiya Y, Dang H, Chainer T, Sammakia B. Thermal characterization of non-raised floor air cooled data centers using numerical modeling. Proceedings of the

ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73387. p 535–543.

[76] Kang S, Schmidt R, Kelkar K, Radmehr A, Patankar S. A methodology for the design of perforated tiles in raised floor data centers using computational flow analysis. IEEE Trans Compon Packaging Technol 2001;24(2):177–183.

[77] Karki K, Patankar S. Air flow distribution through perforated tiles in raised floor data centers. Trans Build Environ 2006;41(6):734–744.

[78] Karlsson JF, Moshfegh B. Investigation of indoor climate and power usage in a data center. Trans Energy Build 2003;37(10):1075–1083.

[79] Kurkjian C, Glass J. Air-conditioning design for data centers accommodating current loads and planning for the future. ASHRAE Trans 2004;111(2):715–724.

[80] NEMA. Metal cable tray installation guidelines. Rosslyn: National Electrical Manufacturers Association; 2001. NEMA VE 2-2001.

[81] Norota M, Hayama H, Enai M, Kishita M. Research on efficiency of air conditioning system for data center. Proceedings of the 25th International Telecommunications Energy Conference (INTELEC '03); October 19–23; Yokohama, Japan; 2003. p 147–151.

[82] Patterson M, Steinbrecher R, Montgomery S. Data centers: comparing data center and computer thermal design. ASHRAE J 2005;47(4):38–42.

[83] Radmehr A, Schmidt R, Karki K, Patankar S. Distributed leakage flow in raised floor data centers. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73273. p 401–408.

[84] Rambo J, Joshi Y. Multi-scale modeling of high power density data centers. Proceedings of the ASME 2003 International Electronic Packaging Technical Conference and Exhibition, Volume 1; July 6–11; Maui, HI 2003. Paper No. IPACK2003-35297. p 521–527.

[85] Rambo J, Joshi Y. Physical models in data center air flow simulations. Proceedings of the ASME 2003 International Mechanical Engineering Congress and Exposition; Heat Transfer, Volume 2; November 15–21; Washington, DC 2003. Paper No. IMECE2003-41381. p 153–159.

[86] Rambo J, Joshi Y. Reduced order modeling of steady turbulent flows using the POD. Proceedings of the ASME 2005 Summer Heat Transfer Conference collocated with the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems; Volume 3: Heat Transfer; July 17–22; San Francisco, CA 2005. Paper No. HT2005-72143. p 837–846.

[87] Schmidt R, Cruz E. Raised floor computer data center: effect on rack inlet temperatures of chilled air exiting both the hot and cold aisles. Proceedings of the ITherm Conference; 2002. p 580–594.

[88] Schmidt R, Cruz E. Raised floor computer data center: effect on rack inlet temperatures when high powered racks are situated amongst lower powered racks. Proceedings of the ASME 2002 International Mechanical Engineering Congress and Exposition; Electronic and Photonic Packaging, Electrical Systems Design and Photonics, and Nanotechnology; November 17–22; New Orleans, LA 2002. Paper No. IMECE2002-39652. p 297–309.

[89] Schmidt R, Cruz E. Cluster of high powered racks within a raised floor computer data center: effect of perforated tile flow distribution on rack inlet temperatures. Proceedings of the ASME 2003 International Mechanical Engineering Congress and Exposition; Heat Transfer, Volume 2; November 15–21; Washington, DC 2003. Paper No. IMECE2003-42240. p 245–262.

[90] Schmidt R, Cruz E. Raised floor computer data center: effect of rack inlet temperatures when rack flow rates are reduced. Proceedings of the ASME 2003 International Electronic Packaging Technical Conference and Exhibition, Volume 2; July 6–11; Maui, HI 2003. Paper No. IPACK2003-35241. p 495–508.

[91] Schmidt R, Cruz E. Raised floor computer data center: effect on rack inlet temperatures when adjacent racks are removed. Proceedings of the ASME 2003 International Electronic Packaging Technical Conference and Exhibition, Volume 2; July 6–11; Maui, HI 2003. Paper No. IPACK2003-35240. p 481–493.

[92] Schmidt R, Iyengar M, Chu R. Data centers: meeting data center temperature requirements. ASHRAE J 2005;47(4): 44–48.

[93] Schmidt R, Chu R, Ellsworth M, Iyengar M, Porter D, Kamath V, Lehman B. Maintaining datacom rack inlet air temperatures with water cooled heat exchanger. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73468. p 663–673.

[94] Schmidt R, Cruz E, Iyengar M. Challenges of data center thermal management. IBM J Res Dev 2005;49(4/5): 709–724.

[95] Shah A, Carey V, Bash C, Patel C. Exergy based optimization strategies for multi-component data center thermal management: part I, analysis. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73137. p 205–213.

[96] Shah A, Carey V, Bash C, Patel C. Exergy based optimization strategies for multi-component data center thermal management: part II, application and validation. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS,

and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73138. p 215–224.

[97] Sharma R, Bash C, Patel C. Dimensionless parameters for evaluation of thermal design and performance of large scale data centers. Proceedings of the AIAA/ASME Joint Thermophysics and Heat Transfer Conference; June; St. Louis, MO 2002. Paper no. AIAA-2002-3091.

[98] Sharma R, Bash C, Patel C, Beitelmal M. Experimental investigation of design and performance of data centers. Proceedings of the Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM ’04); Volume 1; 2004. p 579–585.

[99] Shrivastava S, Sammakia B, Schmidt R, Iyengar M. Comparative analysis of different data center airflow management configurations. Proceedings of the ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems collocated with the ASME 2005 Heat Transfer Summer Conference. Advances in Electronic Packaging, Parts A, B, and C; July 17–22; San Francisco, CA 2005. Paper No. IPACK2005-73234. p 329–336.

[100] Shrivastava S, VanGilder J. A statistical prediction of cold aisle end airflow boundary conditions. Proceedings of the Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM ’06); May 30–June 2; San Diego, CA 2006. p 420, 9 pp.

[101] Spinazzola RS. High delta-T cooling server rack increases energy efficiency, reliability for data centers. J Assoc Energy Eng 2003;100(2):6–21.

[102] SSI. 2008. Enterprise Electronics Bay Specification 2008, version 1.0.1. Available at: https://www.ssiforum.org/index.php?option=com_content&view=article&id=5&Itemid=8. Accessed on August 21, 2014.

[103] TIA. Telecommunications infrastructure standard for data centers. Arlington: Telecommunications Industry Association; April 2005. Report nr ANSI/TIA-942.

[104] TIA. Commercial building standard for telecommunications pathways and spaces. Arlington: Telecommunications Industry Association; 2003. Report nr ANSI/TIA-569-B.

[105] VanGilder J, Lee T. A hybrid flow network-CFD method for achieving any desired flow partitioning through floor tiles of a raised floor data centers. Proceedings of the ASME 2003 International Electronic Packaging Technical Conference and Exhibition, Volume 1; July 6–11; Maui, HI 2003. Paper No. IPACK2003-35171. p 377–382.

[106] Wang D. A passive solution to a difficult data center environmental problem. Proceedings of the Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM ’04); Volume 1; 1–4 June 2004. p 586–592.

[107] Frankovsky F. 2011. Most effective computing. Available at http://www.opencompute.org/blog/more-effective-computing/. Accessed on June 19, 2014.

[108] Frachtenberg E, Lee D, Magarelli M, Mulay V, Park J. Thermal design in the open compute datacenter. Proceedings of the 13th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm); May 30–June 1; San Diego, CA 2012. p 530–538.

[109] Data centers: measuring performance. Available at https://www.facebook.com/green/app_121750054637947. Accessed on May 23, 2014.

19

DATA CENTER PROJECT MANAGEMENT AND COMMISSIONING

Lynn Brown
Encotech Engineering, Austin, TX, USA

19.1 INTRODUCTION

This chapter focuses on the processes and efforts involved with the commissioning of a facility as well as project management with an emphasis toward data center commissioning. Commissioning can be applied to any type of facility, but due to the highly complex collection of mechanical and electrical equipment and systems associated with data centers, there are differences in the approach toward the commissioning process related to a data center compared to other types of buildings.

The contents of this chapter will cover on project management with focus on commissioning. The Project Management section will address on the following:

- Project initiation
- Planning
- Execution
- Monitoring
- Closeout

The Commissioning section of this chapter will address the following:

- What is commissioning?
- Why commission a building?
- Why commission a data center?
- Selecting a commissioning firm.
- Project management and commissioning.
- Equipment and systems to be commissioned.
- Commissioning tasks.
- Leadership in Energy and Environmental Design (LEED)-required commissioning.
- Commissioning team members.
- Data center trends.

19.2 PROJECT MANAGEMENT

The commissioning team may have the best field agent in the business, but if the project management component is lacking, the success of the commissioning effort will be challenged. Due to the importance of project management, the subject of project management of the commissioning process will be discussed prior to exploring the commissioning effort.

Project management is defined as being the process of planning, organizing, motivating, and controlling resources to achieve specific goals (Fig. 19.1). The primary challenge of project management is to achieve all of the commissioning effort project goals and objectives as defined in the commissioning scope of work document while staying within the constraints of time, quality, and budget. Due to the importance of meeting the goals and objectives of the commissioning tasks defined in the scope of work document, it is vitally important to have a strong project manager with good communications skills, project management skills, and experience with managing the commissioning effort of a data center construction project.

The project manager will be the primary contact with the commissioning team and will be responsible for monitoring the progress, execution of the commissioning effort, and overall success of the project. Prior to the project manager

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

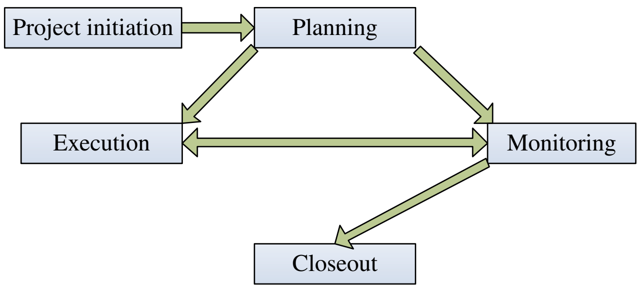


FIGURE 19.1 Project management stages.

initiation of the project management plan, the project manager will need to confirm that the contract has been fully executed and that the commissioning scope of work has been clearly defined and documented with the client.

The key to properly managing a commissioning project is to have a quality assurance–quality control plan in the form of a project management control system in place prior to initiation of the project. The project management control system should consist of the following stages: project initiation, planning, execution, monitoring, and closeout.

19.2.1 Project Initiation Stage

A data center project from a commissioning point of view is initiated when the commissioning firm and the client have agreed on the exact scope of work for the project, a contract has been fully executed, and all of the internal paperwork has been completed. It is at this point the project manager is to begin the project initiation phase of the commission effort.

The first step of the project initiation phase is for the project manager to contact the client, with the client being the person of the entity in which the contract is addressed to. The primary purpose of the meeting is for introductions and to establish the formal process involved with submission and payment of invoices. The secondary purpose of contacting the client is to confirm the point of contact with the client or client representative as well as the contact information of the construction manager/general contractor, if they are not the client.

The next step is to contact the architect or the construction manager/general contractor. The entity to contact will be dependent on if the project is in the design phase or construction phase and whether the project is a design/bid or a design/build project. If the project is in the design phase and it is a design/bid project, contact the architect. If the project is in the design phase or construction phase and is a design/bid or a design/build project, contact the construction manager/general contractor.

The purpose for contacting the architect or construction manager/general contractor is for introduction as well as asking questions and gathering the information required to begin the commissioning tasks. The questions to ask are the following:

- **Status of the project**—Knowing the status will define how quickly the first commissioning task is to be completed:
  - If in the design phase, what phase is the project current in?
  - If in the construction phase, how far along in the construction phase?
- **Request a schedule**—The schedule will assist the project manager in scheduling the commissioning tasks and inserting them into the design phase or construction phase schedule:
  - If in the design phase, request a schedule of completion of each phase and, if known, the approximate construction completion date.
- **Request access to the drawings and specifications**—These documents are utilized to conduct the following commissioning tasks: value engineering, drawing reviews, production of the prefunctional checklist (PFC), functional performance test (FPT) document, and troubleshooting.
- **Request a project directory**—The directory is to include the contact information of the design team, construction manager/general contractors, and each subcontractor installing the commissioned equipment.
- **Primary contacts**—Confirm who will be the primary contact with the client and other key players, such as the construction manager/general contractor, architectural firm, and mechanical, electrical, and plumbing engineering firm.
- **Discuss commissioning specifications**—If a project is in the design phase, bring to the architects attention; a commissioning specification will be submitted and inserted into the project specification manual. If the project is in the construction phase, ask if a commissioning specification has been written. If one has not been written, inform the construction manager/general contractor; a commission specification will be written and the necessary paperwork completed to include the commissioning specification as part of the contract documents. If one has been written, request a copy of the commissioning specification and edit as necessary to meet the agreed-upon scope of work.
- **Discuss next step**—Discuss what is the next step in the commissioning process and the deliverable that will be forthcoming based on the current schedule.

19.2.2 Planning Stage

The next step in the project management process after the project initiation phase is the planning phase. During this phase, the commissioning plan is developed. This is a very important document for it will become the roadmap for the execution of the commissioning process throughout the life of the process. The commissioning plan will either be a Design Phase Commissioning Plan or a Final Commissioning

Plan. As the saying goes, "If you do not plan, then you plan to fail." This applies to all types of projects, but due to the critical nature of a data center project, a strong project manager with good communications skills, experience, and a project management plan is invaluable.

If the commissioning process begins during the design phase, a Design Phase Commissioning Plan will be developed with a Final Commissioning Plan developed at a later date when the construction phase begins. If the commissioning work begins in the construction phase, a Final Commissioning Plan will be developed.

The table of contents of the Design Phase and Final Commissioning Plan is identical with the contents within each Plan being different. Those documents that can only be created during the construction phase will not be inserted into the Design Phase Commissioning Plan. The section to contain those documents created during the construction phase will be left blank.

When the construction phase begins, the following document will be inserted into the Design Phase Commissioning Plan and thus converted and submitted as the Final Commissioning Plan:

- Insertion of the commissioning schedule into the construction schedule in place of the Draft Commissioning Schedule of the Appendix
- Updating of the Commissioning Test Procedure Index and insertion in place of the Sample Commissioning Test Procedure Index of the Appendix
- Inserting the project-specific FPT documents for all commissioned equipment in place of the Sample PFC Document of the Appendix
- Insertion of the project-specific PFC documents for all of the commissioned equipment in place of the Sample FPT Document of the Appendix

The following is a representative of the contents of a Design Phase Commissioning Plan:

- Purpose of the Plan
- Overview of the Commissioning Process
- Specific Objectives
  - Commissioned Equipment/Systems
- Roles and responsibilities
  - Commissioning Team Members List
  - Commissioning Process General Rules
  - Commissioning Responsibility Breakdown
- Commissioning Management
  - Information Flow
  - Scheduling
  - Site Visit Protocol
  - Tracking Deliverables
  - Deficiency Reporting
  - Testing Strategy
  - Reports/Logs
  - Safety and Security
- Commissioning Process
  - Commissioning Timeline
  - Commissioning Task Overview
  - Design Phase Tasks
  - Bidding Phase Tasks
  - Construction Phase Tasks
  - Occupancy Phase Tasks
- Appendix
  - Draft Commissioning Schedule
  - Sample Commissioning Test Procedure Index
  - Sample Issue Resolution Log
  - Sample PFC document
  - Sample FPT document

All sections of the Commissioning Plan contents listed earlier were covered in the commissioning section of this chapter with the exception of the Commissioning Management section. The following will focus on the project management-related topics in the commissioning plan during the Planning Phase.

19.2.2.1 Information Flow Establishing as lines of communication between all the team members is VITAL. During the Project Initiation Phase, the primary contact of the client, construction manager/general contractor, architect, and engineering was provided (Figs. 19.2 and 19.3). During the Planning Phase, a formal line of communication for information flow is to be established between these entities. A sampling of the various lines of communication for a typical design/bid project and a design/build project is as follows.

Once the lines of communications have been established, a means of conveying the information is to be established. One such means is to set up website for the transfer of information. These information flow sites are typically set up by the architect or the construction manager/general contractor along with guidelines on how to post to the site. Some sites will have automatic notifications when new documents are posted, but a good practice once posting a document is to e-mail the parties of interest about the document being posted and attach the document to the e-mail. If it is a document of high importance, follow up the e-mail with a phone call.

19.2.3 Execution Stage

19.2.3.1 Safety and Security The safety and security of all commissioning agents on-site is to be of the upmost importance to the commissioning project management team. It is the responsibility of the commission project manager to contact the construction manager/general contractor to discuss their safety policies, safety classes/training, and procedures established

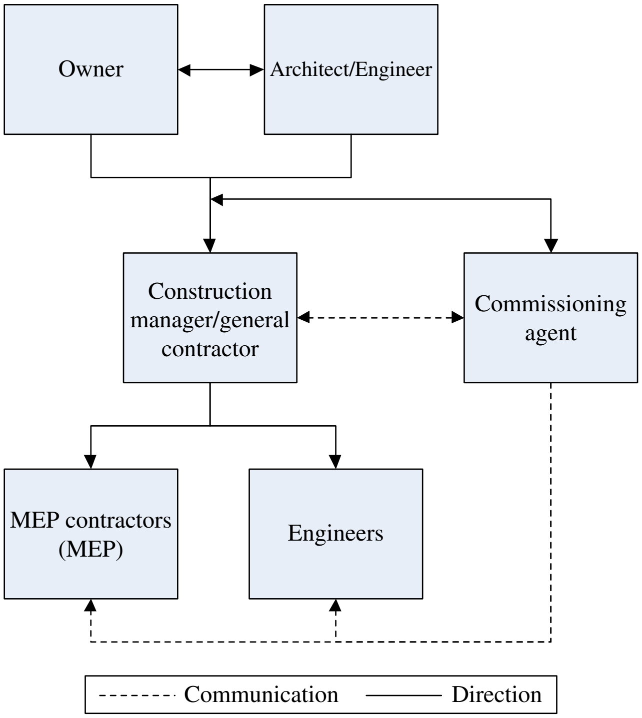


FIGURE 19.2 Lines of communication design/bid.

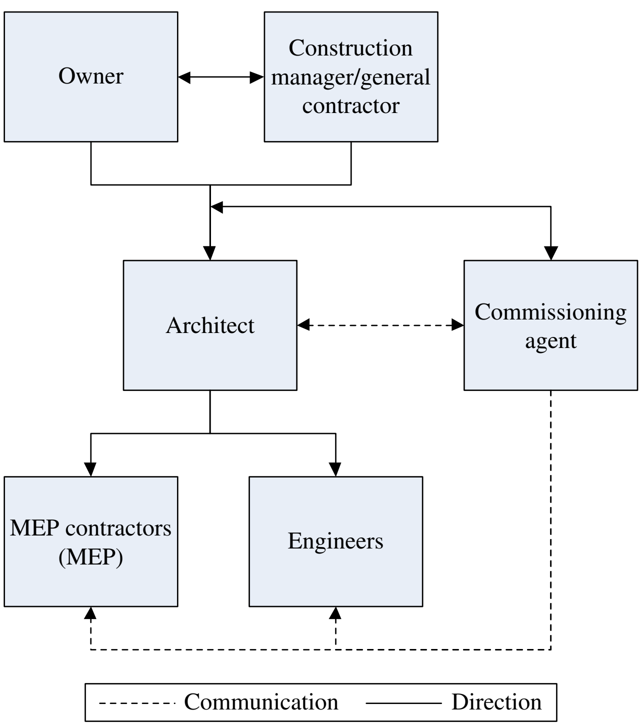


FIGURE 19.3 Lines of communication design/build.

for the project site. This not only includes the commissioning agent employees but also any subcontractors of the commissioning firm that will be visiting the site. The main points of the safety and security policy are typically the following:

- Wearing the required Personal Protective Equipment
- Adhering to the site employee background check policy
- Adhering to the site employee drug testing policy
- Properly displaying the site identification badge and parking permit
- Following the construction manager/general contractor electrical Lock Out Tag Out policy
- Utilization of the proper signage and barricades during testing
- Safeguarding of any keys or access devices given to the commissioning agents
- Participation in all mandatory safety meetings and stand-downs
- Submitting the construction manager/general contractor-required Job Hazard Analysis documentation

The intent of the commissioning project manager is to be an asset to the site safety team and prevent any unsafe conditions and recordable incidents.

19.2.3.2 Scheduling Upon receipt of either the requested design phase schedule or the construction schedule, it is of high importance for the commissioning project manager to insert all of the commissioned-related tasks and deliverables on to the schedule. If the project is in the design phase, submit the schedule back to the architect for review. If the project is in the construction phase or if the project is a design/build project, submit the construction schedule back to the construction manager/general contractor for review.

The value of the scheduling exercise is to confirm all tasks and deliverables can be completed within the design phase and construction phase timeline. Given data center projects typically have a tight and inflexible completion date, scheduling and consistent ongoing monitoring of the schedule on a frequent basis by the commission project management team and design and construction team is a very necessary step in meeting the substantial completion deadline. Typically, the review of the construction schedule is a line item on the regular schedule construction meeting conducted by the construction manager/general contractor.

19.2.3.3 Site Visit Protocol Prior to the commissioning agent setting foot on the construction site, the project manager is to establish a line of protocol with the construction manager/general contractor during each site visit. Prior to the commissioning agent visiting the site to perform a commissioning task, the project manager will confirm with the construction manager/general contractor the following:

- The reason for visiting the site is still valid.
- The date(s) to be on-site.
- The equipment and systems involved with the site visit are ready.
- The entities to assist with the commissioning process are still available.
- The person with the construction manager/general contractor team to contact upon arrival at the site.

Upon arrival on-site and before going into the construction area, the commissioning agent has to "check-in" with the construction manager/general contractor. The commissioning agent will then meet with the people and entities involved with the site visit prior to going out to the construction site to review the commissioning task to complete and everybody's role during the process. After the completion of the site visit, the commissioning agent will meet with the construction manager/general contractor to debrief the results of the site visit.

19.2.3.4 Testing Strategy The project manager is to manage the execution of the testing strategies to verify the operation of the commissioned equipment/systems conform to the owner's project requirements (OPRs) and as per the engineer's construction drawings and specifications or any other contract documents reflecting a change to the drawings and specifications (Fig. 19.4). The testing process is the heart and soul of the commissioning process and is extremely critical for data centers. The testing strategies begin upon confirmation that the commissioned equipment and systems have been installed as per the contract documents. The tests typically associated with a data center project include the following:

- Start-Up
- FPT
- Integrated System Testing (IST)

The primary objective of the test procedures is to verify that the component and system operational requirements have been achieved through a full range of operating modes and scenarios. The commissioning agent will utilize the equipment specifications, submittals, shop drawings, and control sequences of operation documents to develop the project-specific required test procedures. It should be noted it is VERY important that the testing strategies developed are project-specific and not "canned" strategies that can be pulled down off of the Internet or from another similar data center project. All data centers are very unique, and the testing procedures are to be unique and project specific as well.

The testing procedures are to include clear testing steps, expected results, and criteria for a test's passage or failure. The testing strategy process begins with verifying installation of the commissioned equipment, followed by equipment start-up and FPT and concludes with the IST. The following chart

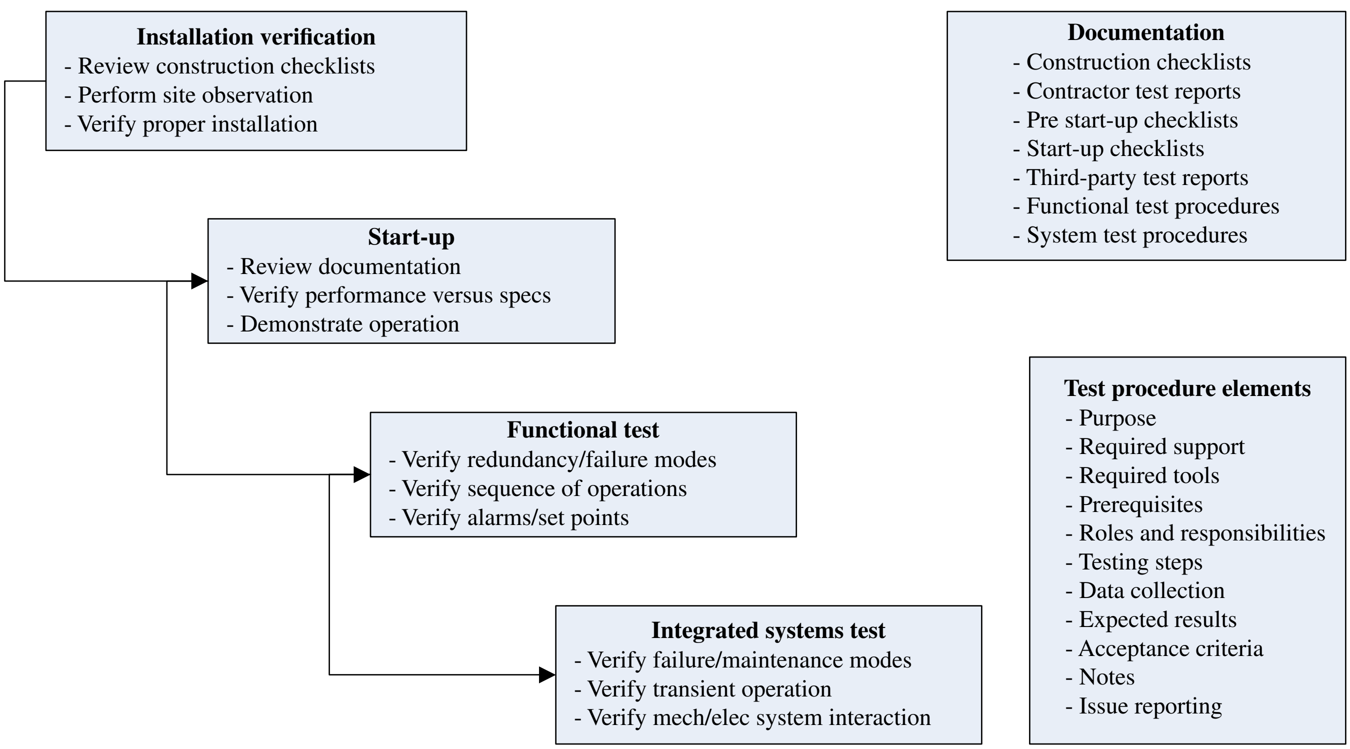


FIGURE 19.4 Commissioning testing strategy.

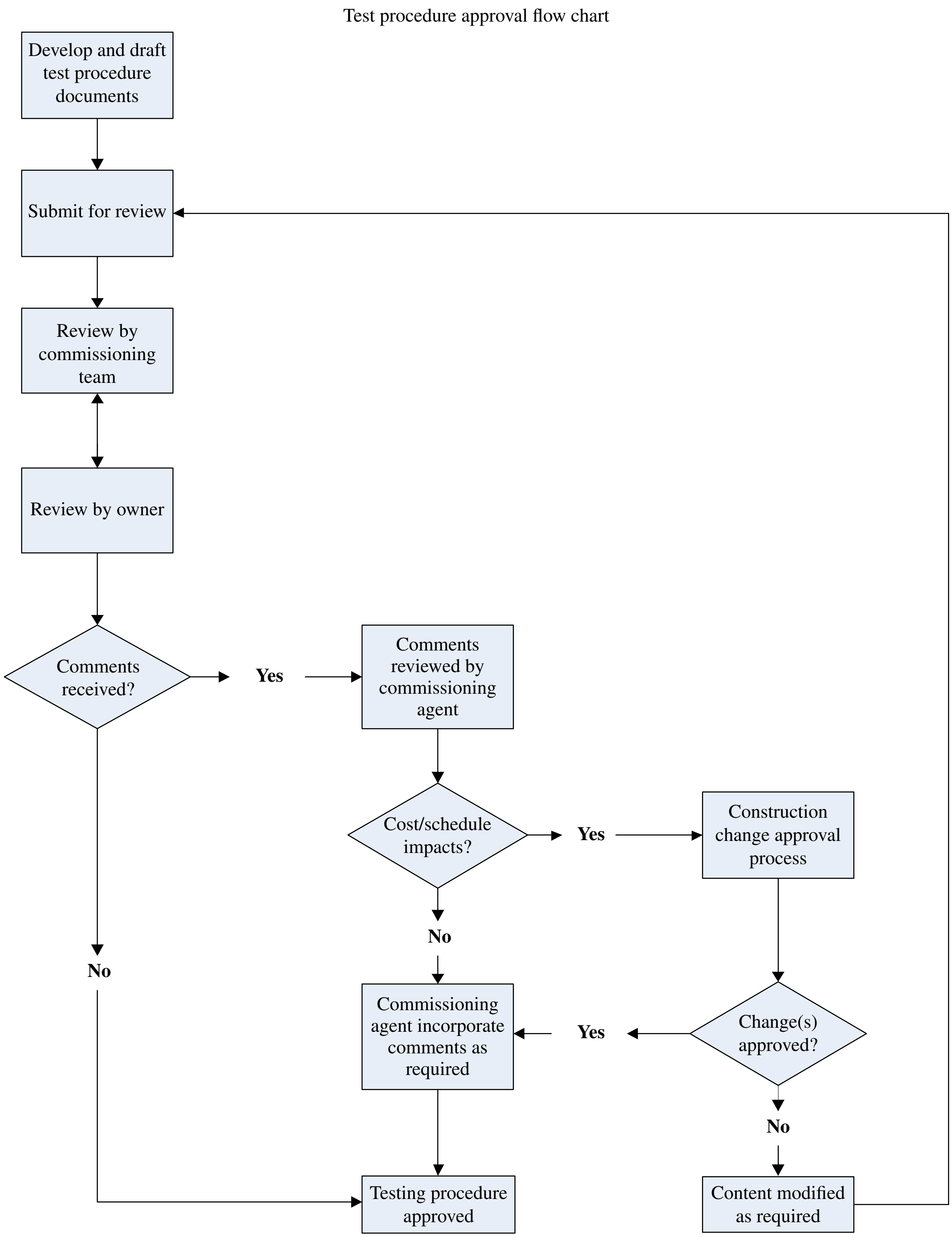


FIGURE 19.5 Test approval flow chart.

diagrams the progression of the overall commissioning testing strategy monitored by the project manager.

After the test procedure documents (Fig. 19.5) have been created, they will be submitted to the commissioning team for review and comment. Keep in mind the commissioning team is not only the commissioning firm but, as mentioned earlier, the commissioning team consists of the owner/owner's representative; design team; construction manager/general contractor; mechanical; electrical, and plumbing subcontractors; equipment vendors/manufacturer representatives; controls contractor; and test and balance contractors.

The intent of this review process is to ensure the tests are not destructive and in accordance to the design intent and owner's operational needs as applied to the commissioned equipment and systems tested. Once testing strategies are reviewed and finalized, they will be distributed to the Commissioning Team to prepare for test execution.

19.2.4 Monitoring Stage

19.2.4.1 Tracking Deliverables The timely issuance of the commissioning deliverables on the agreed-upon dates on the design or construction schedule is to be tracked by the project manager. The project manager is to create a living document that evolves over the course of the project, identifying all of the deliverable submitted, the deliverables that are in-progress, and those deliverables pending to be submitted. This document will typically be submitted on a monthly basis to the client/client representative, architect, construction manager/general contractor, and others as directed by the client/client representative.

The purposes of the deliverable tracking document are as follows:

- Keep the Owner/Client and Construction Manager informed on the status of the commissioning-related tasks completed and in-progress and all pending "to be completed" tasks.
- Used as a tool for the Cx Project Manager to keep track of the status of each project and plan for upcoming tasks.
- Used as a source of information for all members of the Commissioning Team to answer a question or confirm the status of the report.
- Used as a tool for the client during the billing process to verify if the percentage of work completed on the invoice matches the work completed to date on the deliverable tracking document.

19.2.4.2 Deficiency Reporting From a global viewpoint, the commissioning agent's primary role is to document proper installation and operation of the commissioned equipment/systems to verify conformance to the OPRs and to the engineer's construction drawings and specifications or any other contract documents reflecting a change to the drawings and specifications. The timely reporting and correcting of the observed deficiencies that are not in compliance is a very important project management role, especially on data center projects, which typically has a tight deadline.

The commissioning agent is to develop a reporting document that will list all of the deficiencies observed during the commissioning process. The deficiency document will be populated by the commissioning field agents, and the project manager will submit the findings and track the deficiencies. The deficiencies will remain on the list as open items until the issue has been corrected or proven to be correct. The deficiency will not be closed and removed from the list until the commissioning agent has field verified the item has been corrected or if proper documentation has been submitted, proving the equipment/system has been installed and/or operating as per the contract documents.

The contents of the deficiency document should not only report the observed deficiency but also provide enough information to all parties concerned with the resolution of the issue. The contents of the deficiency log at a minimum should include the following:

- Tracking number
    - This will be a unique number assigned to the deficiency for discussion purposes.
- Date entered
    - Indicate the date of the observed deficiency.
- Identify equipment/system
    - Identify the equipment/system with the deficiency and include, if provided, the identification number of mark for the equipment/system. The identifying mark or number should match what is on the construction drawings.
- Reference the source of the deviation
    - Indicate specification section, drawing number, note, detail, or any other document the observed deficiency deviated from.
- Description of the deficiency
    - Explain as clear and concise as possible the deficiency.
- Response
    - The section will be for inputting the response to the deficiency from the entity responsible for resolving the issue.
- Deficiency status
    - This will identify whether the issue is still open and requires further action to resolve or the issue is closed.

19.2.4.3 Reports/Logs The project manager is responsible for proper documentation and timely distribution of all of the reports and logs. The number and frequency of distribution of the reports and logs can vary from project to project, and there may be reports that the owner, client, or construction manager/general contractor requires during the commissioning process. In addition to any required documents from outside entities, there are at least three project management-related documents the commissioning agent's project manager should submit during the life of the project. These three documents include the following:

1. Commissioning Update Reports
   - The intent of the commissioning update report is to convey the current status of the commissioning effort. The contents of the report should include a listing of all commissioning tasks to be completed and the status of each task as being completed, in-progress, or pending, as well as a brief update of the in-progress tasks. The report is to be submitted on a monthly basis to the client, owner, and construction manager/general contractor.
2. Site Visit Report
   - Unless there is a commissioning agent on-site full time during the life of the commissioning effort, every time a commissioning agent goes to the site, a site visit report is to be written. The site visit report is to be written by the visiting commissioning agent and submitted to the project manager for review and distribution. The contents of the report are to include the date(s) of the site visit, the name of the commissioning agent visiting the site, other participants during the site visit, purpose of the site visit, general notes, and observations during the site visit. All reports are to be submitted within three working days to the construction manager/general contractor and distributed to other entities at the request of the construction manager/general contractor.
3. Deficiency Log
   - The details related to the deficiency log were discussed in Deficiency Reporting section. The deficiency log is to be submitted after each site visit as an attachment to the site visit report and also as an attachment to the Commissioning Update Reports. The deficiency log is also typically an agenda item during any regularly scheduled construction meetings and commissioning meetings.

19.2.5 Closeout Stage

At the completion of the project, closeout documentation is to be created. There are two documents that are typically produced, the Final Commissioning Report and the Systems Manual.

19.2.5.1 Final Commissioning Report The Final Commissioning Report is a document that is submitted at the completion of most all projects. This document is created by the commissioning agent and issued to the owner and construction manager/general contractor.

The contents of the Final Commissioning Report will include the following:

- Executive Summary Narrative
- Outstanding Issues
- Commissioning Activities Completed by Phase
  - Design Phase
  - Bidding Phase
  - Construction Phase
  - Acceptance Phase
  - Occupancy Phase
- Site Visit Reports
- Project Update Reports
- Appendix
  - Appendix to include all completed commissioning task documents

19.2.5.2 Systems Manual The Systems Manual is always created when the project is attempting a level of LEED certification that requires the Energy and Atmosphere (EA) Credit 3 Enhanced Commissioning (EC) requirements. The system can also be required on non-LEED projects if the owner sees the value in the Systems Manual and includes the issuance of a Systems Manual.

The purpose of the manual is to provide to the building operation staff the information needed to understand and optimally operate the commissioned systems. The Systems Manual will be developed by the commission agent and issued to the construction manager/general contractor for distribution to the owner's building operations staff. The following is a sampling of the documents that are typically found in the Systems Manual followed by the entity responsible for delivering the documents to the commission agent:

- Final version of the Basis of Design (BoD)—engineer of record
- System single-line diagrams—mechanical construction drawing 1-line diagrams
- As-built sequences of operation control drawings including original set points—controls contractor
- Operating instruction for integrated building systems—operation and maintenance (O&M) manual issued by the subcontractor.
- Recommended schedule for retesting of systems of commissioned equipment and systems, if not included in the O&M Manual—subcontractor..
- Blank FPT test forms for retesting of the commissioned equipment and systems by the certified commissioning authorities (CxA)
- Recommended schedule for recalibrating sensors and actuators—controls contractor.

19.3 COMMISSIONING

19.3.1 What Is Commissioning?

Prior to the construction of a data center or any other building, a set of construction documents are created. Construction documents consist of and encompass the preparation of construction drawings and specifications that set forth the detailed requirements for the construction of a building as well as how the various systems and equipment are to be installed and operated. The construction drawings represent the illustrative dimensions of the construction documents, while the specification represents the written. The drawings and specifications are to be complimentary of each other, with neither having precedence over the other.

The architect/engineering design team creates the construction documents. There are several reasons why construction documents are created, but the two primary reasons are twofold. One is to obtain a building permit and two to communicate as clearly as possible the design intent to meet the owner's expectations and buildings intended use. Commissioning is a quality-driven systematic process that will verify the systems and equipment being commissioned are installed and perform independently and interactively as per the construction documents in order to meet the owner's expectations and operational needs.

The commissioning effort will coordinate and implement multiple commissioning tasks primarily focusing on the installation and operation on equipment and systems to be commissioned. The commissioning firm from an overall viewpoint will complete these tasks through review of documents, site observations, and testing verification.

The proposed scope of work of the commissioning process does not replace or reduce the responsibility of system designer engineers, installing contractors, subcontractors, or suppliers in performing all aspects of work and testing in providing a finished and fully functioning product and system. The commissioning firm is not responsible for design concept, design criteria, compliance with codes, design or general construction scheduling, cost estimating, or construction management.

19.3.2 Why Commission a Building?

A very common question asked by owners when the topic on whether to utilize the services of a commissioning firm is:

> I am paying the engineers to perform construction administration services; why do I need a commissioning firm to do the same thing?

Even though there is some overlap of construction administrative (CA) services and the commissioning effort, there are substantial differences.

- The engineer's CA service scope of work only requires the engineer to be generally familiar with the installed systems, whereas the commissioning firm is heavily involved with every aspect of the installation and operation of the commissioned equipment.
- The engineer has a set number of random site visit built into their CA services, whereas the commissioning firm will have multiple site visit specifically to target the different observation and tests for each commissioned equipment and systems. Based on the standard American Institute of Architects (AIA) document, the design team is to be "generally familiar" with the installation and operation of the equipment.
- The engineers only report observed deficiencies during their site visits, whereas the commissioning firm will not only report deficiencies but witness whether each of the systems and equipment to be commissioned is installed and operates correctly.
- The engineers do not provide to the owner a "baseline" condition of the installation and operation of the commissioned equipment and systems.

Another common question from an owner is:

> We are already paying a contractor to perform start-up and confirm the equipment operates; why do we need a commissioning firm?

True, the contractor is responsible for the start-up of the commissioned equipment, but in a majority of cases, the contractor does not confirm the equipment and systems work interactively with other systems. When commissioning a building, all of the commissioned equipment will be verified as not only being installed and operating as independent units but also interactively with other equipment and systems in accordance with the contract documents.

The contractors and subcontractors, primary goal is to meet the construction completion date and within budget. In the heat of battle to complete the project on schedule, honest mistakes are made, and many of these errors are not discovered during the start-up process or during the normal course of construction. When a building is commissioned, the process is not to "catch the contractor" or point fingers but to be part of the team as another set of eyes and ears for the owner to validate that the commissioned equipment and systems are installed and operate as per construction documents.

A question that may be presented to the owner that will be a deciding factor on whether to commission a building is:

> Do you want your building to be LEED Certified?

LEED is an acronym for Leadership in Energy and Environmental Design, a sustainable rating system for buildings developed and promoted by the U.S. Green Building Council (USGBC). The USGBC is a coalition of building industry leaders who came together in 1993 to promote environmentally responsible and profitable buildings that are also healthy places to live and work. If the owner decides to pursue LEED certification, one of the requirements is for the building to be commissioned.

Based on findings from the USGBC, the following is a sample list of LEED benefits to assist the owner in deciding on whether to pursue LEED certification:

- A LEED-Certified Building is Good Business Sense.
  - Third-party commissioning is required, which verifies for the owner that the commissioned equipment and systems are installed and operating as per the contract documents.
  - Enhances building marketability.
  - Potentially protects or increases property values.
  - Promotes energy efficiency and thus reduces operating costs.
- LEED Buildings are Healthier.
  - Improved indoor air quality resulting in satisfied tenants and thus less tenant turnover.
  - A healthy building can reduce potentially liability due to poor air quality and the "sick building syndrome."
  - If owner is occupying the building, a healthy building results in increased worker satisfaction, improved morale, reduced absenteeism, and increased productivity.
- LEED Building are Environmentally Friendly.
  - A LEED building can substantially reduce or eliminate negative environmental impacts and improve existing unsustainable design, construction, and operational practices.
  - A LEED building promotes energy conservation and recycling, reduces the use of raw materials, and stops the use of toxic products.

If the owner decides to certify the building, the commissioning effort will involve, at a minimum, meeting the requirements of the Certified LEED status under the EA Prerequisite 1—Fundamental Commissioning (FC). If the owner elects to provide FC of the building to achieve a status above LEED certified, the commissioning tasks of the EA Credit 3—FC will be required. The commissioning tasks associated with the EA Prerequisite 1 and the EA Credit 3 will be discussed in Section 19.6.

The main reason why a building should be commissioned boils down to the benefits of the commissioning effort to the owner. The owner's management group basically is not primarily interested in the how's and why's of the commissioning effort but more interested in the monetary benefit of having their building commissioned. From an owner's management group viewpoint, in order for the commissioning effort to be economically justifiable, the benefit gained from the commissioning effort must be greater than the cost of the commissioning service.

19.3.2.1 Commissioning Viability = Commissioning Benefits > Commissioning Cost The following is an excerpt from the report "Building Commissioning: A Golden Opportunity for Reducing Energy Costs and Greenhouse Gas Emissions" conducted by Evan Mills of the Lawrence Berkeley National Laboratory dated July 21, 2009, for the California Energy Commission Public Interest Energy Research that provides quantifiable cost–benefit indicators related to commissioning:

> *The results are compelling. We developed an array of benchmarks for characterizing project performance and cost-effectiveness. The median normalized cost to deliver commissioning was $0.30/ft2 for existing buildings and $1.16/ft2 for new construction (or 0.4% of the overall construction cost). The commissioning projects for which data are available revealed over 10,000 energy-related problems, resulting in 16% median whole-building energy savings in existing buildings and 13% in new construction, with payback time of 1.1 years and 4.2 years, respectively. In terms of other cost-benefit indicators, median benefit-cost ratios of 4.5 and 1.1, and cash-on-cash returns of 91% and 23% were attained for existing and new buildings, respectively. High-tech buildings were particularly cost-effective, and saved higher amounts of energy due to their energy intensiveness. Projects with a comprehensive approach to commissioning attained nearly twice the overall median level of savings and five-times the savings of the least-thorough projects.*

One point that is of vital importance that cannot be quantifiable, which is not included in the study, is the potential cost saving for problems averted due deficiencies found during the commissioning effort. As an example, let us assume that during the commissioning process, a deficiency was found that would have caused a catastrophic shutdown of the building. Given downtime for a data center is "not an option," it would be very difficult to quantify the cost of a shutdown. There are numerous deficiencies that are found during the commission process that are not quantifiable, which would further improve the benefits of commissioning a building. The list of observed deficiencies is seemingly endless, and based on our findings, the most common problems are centered around:

- Controls not properly set up for equipment and systems to interact.
- Control sequence of operation for equipment is not set up properly.

- Excessive duct leakage.
- Automatic airflow control dampers do not operate properly.
- Automatic water control valves do not work properly.
- Supply, exhaust, and return air flows are not within the engineers specified requirements.

Each of the already-mentioned deficiencies correlates to a cost–benefit, and looking at energy savings alone, substantial saving can be found. Additional data from Mills' report conducted on 60 new buildings that were not commissioned and arrived at the following statistics that resulted in the heating ventilation and air-conditioning as being the number one source of complaints during the occupancy phase:

- Fifty percent of the building had heating ventilation and air-conditioning-related problems.
- Fifteen percent had missing equipment.
- Twenty-five percent of the economizers/variable frequency drives did not function.

The issues we experience during the commissioning process as well as those found by the Lawrence Berkeley National Laboratory can be greatly reduced or even eliminated further, emphasizing the importance of commissioning a building.

Saving energy is also a very important reason to commission a building. When a new building is commissioned, the commissioned equipment and systems are verified to be as per the engineer's specification for maximum energy savings. Commissioning of existing buildings also provides impressive energy saving. Based on a study conducted by Evan Mills of the Lawrence Berkley National Laboratory and published on July 21, 2009, a total of 186 buildings were commissioned. The study included government buildings, office buildings and hotels, healthcare, educational, and retail. The study indicated the energy savings on average was in the range of 10–15%.

In addition to energy saving, the following is a sampling of additional value-added benefits as a result of the commissioning process during the design, construction, acceptance, and occupancy phase:

Change orders and other claims are minimized—When drawing reviews are conducted by the commissioning agent during the design phase, potential problems can be identified and corrected "on paper" at no cost rather than during the construction phase, which helps avoid costly change orders and claims.

Fewer problems are recognized after construction—During the commissioning construction and acceptance phase, deficiencies are observed prior to the occupancy phase. By having the deficiency resolved by the contractor and verified as being closed by the commissioning agent during the construction phase, there is a reduction in contractor callbacks during the occupancy phase.

Transition turnover from contractor to building operations is shorter—By the end of the construction phase, the commissioning agent has confirmed the owner's building O&M staff been properly trained, resulting in a smooth transition from the construction phase to the occupancy phase.

Improved indoor air quality—The quality of indoor air is dependent on an acceptable level of outside air being introduced into the building. The design engineer will design a mechanical system that will operate in a manner to maintain, on a constant basis, the required amount of outdoor air to the building. The commissioning agent will verify the outside air flow requirements, as specified, are met during functional testing. The benefits of good indoor air quality are satisfied occupants and increased productivity.

Improved room temperature and humidity control—The design engineer provides a system that results in providing room temperature and humidity comfort in meeting the design intent of the space. The commissioning agent verifies the design supply air flow and humidity levels are within the engineer's stated tolerances, resulting in minimizing the number of room temperature complaints.

High-quality building—When the commissioned equipment is installed and operating as per the contract documents, the owner's occupants will occupy a high-quality building, resulting in satisfied occupants, more lease renewals, and a favorable reputation as a good place to work or visit.

Baseline data provided—At the end of the commissioning effort, documentation will be provided, indicating all the commissioned equipment are operating as per the contract documents. The form utilized during this process is the FPT checklist, and these documents will be used as the baseline data for commissioned equipment. Blank FPT will be provided to the owner for future testing for comparison of the baseline data to the current operating condition of the equipment and systems. This is a very important step because the value added by commission of a data center can quickly be eroded away if the equipment and systems are not properly maintained in their designed operating conditions. The best way to maintain the equipment and systems at their designed condition is by recommissioning the equipment utilizing the blank FPT.

In a nutshell, the primary reason for commissioning is not only to provide value to the owner but to verify the owner "gets what they paid for" and what the owner paid for is a set of contract documents indicating how the commissioning equipment is to be installed and operated. The contract documents will initially consist of the signed and sealed 100% construction drawings and specifications. Once the project is underway, contract documents will consist of all issued and confirmed Addendums, Request for information, and other contractual documents reflecting any change in the construction drawings and specifications.

When the commission agent has confirmed the commissioned equipment and systems are installed and operating as per the contract documents, the deficiencies listed earlier that are normally found and those statistics identified in the Lawrence Berkeley National Laboratory study are drastically reduced or eliminated.

19.3.3 Why Commission a Data Center

In principle, there are no fundamental differences between commissioning an office building, manufacturing facilities, healthcare facility, or a mission-critical building such as a data center. Though the principle toward commissioning from one building to another may be the same, the approach toward the commissioning process from one type of building to the next will vary. As an example, the approach toward commissioning of an office building may primarily focus on the building management system, and energy efficiency, a health care facility will focus on life safety and security, and a data center will focus primarily on availability and redundancy.

One of the unique differences between commissioning a data center and a majority of the other types of building is the high level of special consideration required toward redundancy and availability of the electrical and mechanical systems. With non-data center buildings, single-point failures are tolerable, whereas with a data center, a single-point failure is definitely NOT AN OPTION.

With non-mission-critical buildings, redundancy is treated on an "equipment" basis where there is sufficient amount of backup built into the equipment selection design to "get by" until the equipment that failed is back on line or after maintenance has been completed. As an example, if a chiller of primary chilled water pump is off-line for whatever reason, there is sufficient enough chilling or pumping capacity in the remaining chillers and pumps to maintain the cooling load of the building at a tolerable level.

With mission-critical data centers, the focus is more toward "system" failures, not "equipment" failures. If there is a "system" failure with a noncritical building, the worst scenario is the building will be out of service. For a data center, being out of service IS NOT AN OPTION!

As an example, given a chilled water system, which is designed for two of three chillers to be operational to maintain the cooling load of a data center, if there is "equipment" failure to one chiller, there will not be a problem. But if there is a "system" failure, such as losing power to two of the three electrical panels handling two of the chillers, two chillers will be down, and there will not be enough cooling capacity to the data center. Lack of adequate cooling load to a data center is also NOT AN OPTION!

As you can see, a data center MUST be reliable with an acceptable level of availability to the point that will meet the owner's requirement. It is the responsibility of the design team to provide the level of redundancy and availability to the data center, and it is the responsibility of the commissioning team to verify the redundancy and availability are in place during the design phase drawing review process and to verify the redundancy and availability during the FPT and IST process of the acceptance phase are operational as per the construction documents.

As you can see, commission of a data center is a MUST DO because of the very complex nature of the equipment and systems associated with a data center. Commissioning will not guarantee any unplanned outages or major problems because every possible conceived scenario cannot be acted out but the commissioning effort will verify that the commissioned systems will perform as per the contract documents and will work as designed based on the most likely of scenarios. One given is that if a data center is not or is poorly commissioned, there is a very high probability of an expense and disruptive downtime that could have been avoided if the data center was properly commissioned by a reputable commissioning firm.

19.3.4 Selecting a Commissioning Firm

The commissioning firm is the leader in the spearheading of the commissioning process and serves as an advocate for the owner. The selection of a commission firm is considered a necessary and vital addition to the design team and construction team throughout the life of a project. To obtain the maximum benefit from a commissioning agent, it is highly recommended to engage the serves of the commission agent at the onset of the project in the design phase. This cannot be overemphasized enough when dealing with a data center and the complex equipment and systems.

The selection process typically consists of requesting certain qualifications be met by the commissioning firm followed by an interview. The following is a sample of the qualification requirements and interview suggestions for selecting the right commissioning firm for your project.

19.3.4.1 Qualifications

Years in Business—In the past, commissioning was viewed as a luxury if the project budget allowed, but now that commissioned is viewed as adding significant value to a project, the commissioning business is rapidly expanding. Along with the growth of the commissioning industry is an influx of new firms vying to "get in the commission business" that are not qualified. When dealing with the complexities of a data center, DO NOT consider a firm that has not been in business for a short period of time.

Types of Service—When selecting a commissioning firm, select a firm that is solely in the business of providing commission services and not a firm that does commissioning "on the side" or if commissioning is not their primary line of service.

Experience—As the saying goes, there is no substitute for experience. The commission firm is to have had commissioning experience in buildings of similar size, function, and scope of your project.
References—Request references from current previous projects of similar size and scope of your project.
Independent Third-Party Firm—It is important for the commissioning firm to be independent of the engineering firm, general contractor, or subcontractor affiliated with the design and construction process.

19.3.4.2 Interview Process Even if the commissioning firm is qualified on paper, it does not necessarily mean that it is the right firm to select for your project. What sets equally qualified firms on paper apart from one another is the quality of the people working on the project. All too often, a commissioning firm will send a marketing team to the interview along with a slick presentation that does not allow you to get to know the people working on the project.

When inviting a commissioning firm to an interview, request the project manager to lead the presentation and the commission field agents' team to participate in the presentation. In order to meet the goals and objectives of the commissioning tasks defined in the scope of work document, a project manager and the commissioning field agents must be organized and have good communications skills, a working chemistry, and project management skills, which will be demonstrated during the presentation.

One final note related to selecting a commissioning firm is that do not select the commissioning firm with the lowest fee but select it based on qualification and quality of people assigned to the project. It is very expensive to repair or make unnecessary adjustments on a complex systems associated with data centers once the building is turned over to the owner. Also, if the unthinkable happens, of a shutdown due to system failure, the cost to the owner will be significant. So, it is imperative the most qualified commissioning firm with people who can communicate and be proactive during the commission process be selected to help prevent unnecessary cost once the building is in operation. A qualified commissioning team with the right personnel on the job will more than pay for itself no matter what their fee may be.

19.3.5 Equipment and Systems to Be Commissioned

When commissioning a building, not all of the equipment and systems associated with the project are commissioned. What typically dictates the equipment and systems to be commissioned is based on owner's preference, LEED requirements, type of building, or the commissioning agents' recommendations. If the commissioned equipment and systems are based on the owner's preference, type of building, or the commissioning agent's recommendation, the other factor involved is the construction budget. If the equipment and systems to be commissioned are based solely only on the LEED requirements, the following equipment and systems will be commissioned:

- The heating, ventilation, and air-conditioning (HVAC), and refrigeration systems and associated controls
- Lighting and daylighting control
- Domestic hot water systems
- Any renewable energy systems

With data centers, it is VERY important to include all mechanical equipment and systems designed to served areas sensitivity to a narrow range of temperature and humidity control. Concerning the electrical equipment and systems, it is imperative to include all emergency power equipment and systems providing redundancy and backup capabilities.

If cost was not of concern, all equipment and systems associated with the project would be commissioned. Given data centers, like all building types, have a budget to stay within, it is not possible to commission all of the equipment and systems. The following will be a representative sampling of some of the typical, major equipment and systems associated with data centers to be commissioned. Not all of the following will be applicable to all data centers but will give an order of magnitude of the major equipment to be commissioned.

19.3.5.1 Mechanical
Airside Systems—All air handling units serving temperature- and humidity-sensitive areas, including indoor floor-mounted units, suspended fan coil units, roof-mounted air handling units, packaged air handling systems, and all floor-mounted and suspended computer room units. All major floor- or roof-mounted air handling systems serving noncritical areas. Equipment in the ductwork of airside systems include constant volume and variable volume terminal boxes, duct-mounted heaters, fire/smoke dampers, and airflow measuring stations.
Refrigeration Systems—All chillers, cooling towers, and associated chilled and condensers water pumps. All air cooled condensing units serving computer room air-conditioning units.
Heating Hot Water System—Hot water boilers, expansion tanks, and all pumps associated with the heating hot water system.
Steam System—Steam boilers, pressure reducing valve stations, condensate return pumps, heat exchangers, humidifiers, and boiler feed water system.
Exhaust System—All major general exhaust fan systems, toilet exhaust fan systems, and specialty-type exhaust systems.
Control System—The building management control system controlling all of the commissioned equipment. The commissioning of the control system for a data center is a MUST!

19.3.5.2 Electrical

Emergency System—Including emergency generators, emergency generator load banks, transfer switches, and uninterruptible power supply (UPS) system. Also, the fuel oil system associated with the emergency generators, including the fuel oil storage tanks, day tanks, and fuel oil transfer pumps
Normal Power System—Including motor control center, primary switchgear, paralleling switchgear, major distribution panels, outdoor load bank, load bank switchgear, and transformers, interior lighting, and exterior lighting

19.3.5.3 Plumbing

Domestic Hot Water—Domestic hot water heaters, electric water heaters, and, if utilized, domestic hot water circulating pumps
Domestic Cold Water—If utilized, domestic water booster pumps, meters, backflow preventers, and water softening
Sanitary System—Sump pump ejectors
Storm Water System—Storm water ejector pumps
Gray Water System—Rainwater harvest tanks, rainwater filters, and expansion tanks

19.3.5.4 Building Envelope

Building Exterior—Wall, roof, and glazing
Building Components—Insulation, vapor barriers, and pressure testing

19.3.5.5 Life Safety

Systems—Stairwell pressurization and, atrium pressurization
Barriers—Fire-resistive ratings, smoke barriers, and smoke tight partitions
Rooms—Fire Command Room
Fire Alarm—Interface with the life safety Systems, fire protection system, elevators, HVAC systems, as well as the workstations, controllers, and sensing devices
Fire Protection—Fire pump, jockey pump, backflow preventer, fire department connections, standpipes, and preaction systems

19.3.6 Commissioning Tasks

The following will address the different commissioning tasks associated with each phase of a project's life. The commissioning phases are broken into the following phases:

- Design Phase
- Bidding Phase
- Construction Phase
- Acceptance Phase
- Occupancy Phase

Figure 19.6 highlights the major tasks associated with each phase of the commissioning process.

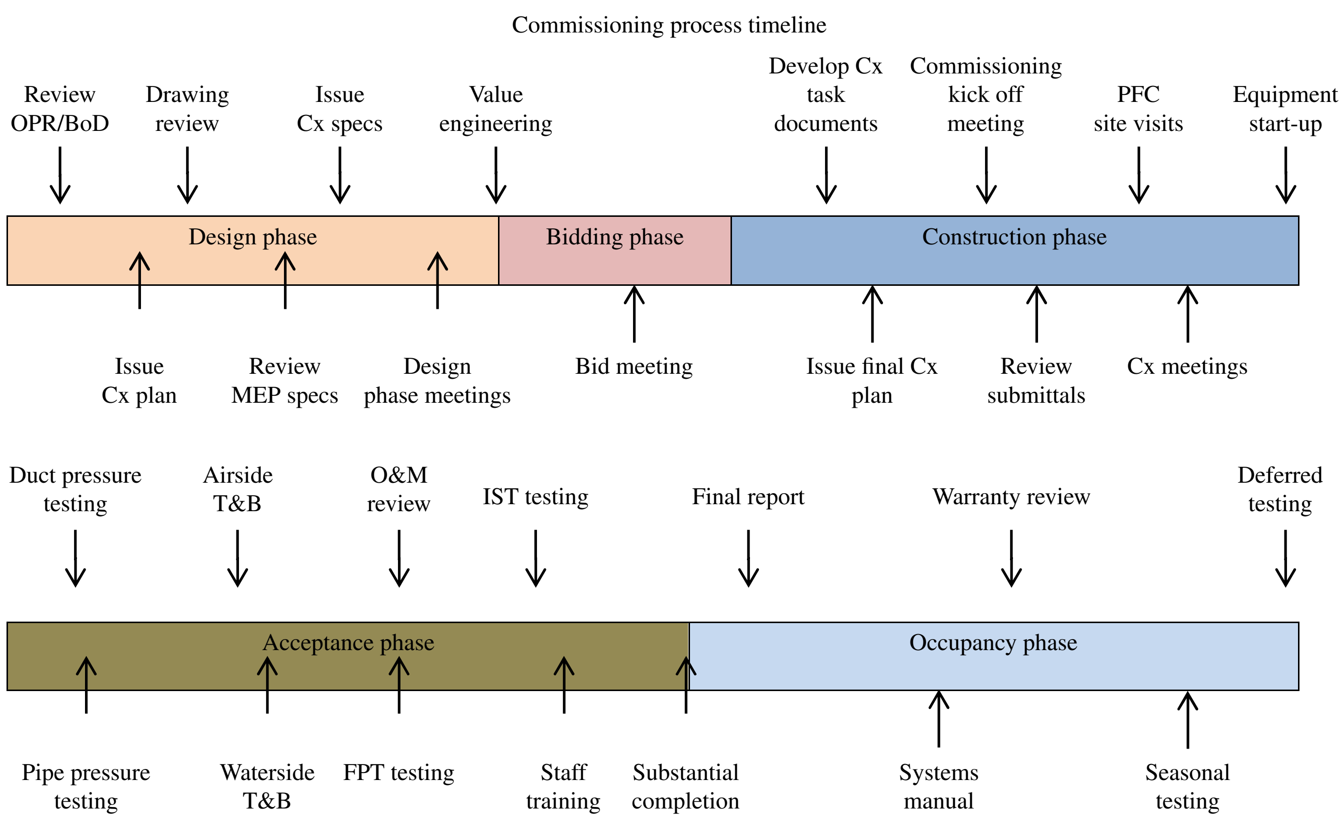


FIGURE 19.6 Commissioning timeline process.

It is highly recommended to obtain the services of a commissioning firm at the design phase of the project due to a potential cost impact on the project. By having the commissioning firm secured during the design phase, with deficiency observed or change recommended, the "cost to fix" will be low and the potential savings will be high. But if the same deficiency or recommended change occurs later on during construction or occupancy phases, the "cost to fix" will be greater and the potential savings low.

As the saying goes, "it is cheaper to correct something on paper than in the field." Given a data center is a very complex facility with complicated interaction of different systems, the cost to fix will be greater, and the potential benefits less when found later on in the project than with other major types of noncritical buildings. Figure 19.7 provides a visual representation of this viewpoint.

The following will delve into the variety of commissioning tasks throughout the life of the commissioning process. The tasks listed are not all inclusive of all of the commissioning tasks that can be completed but will be a good representation of the types of tasks associated with a data center (Fig. 19.8).

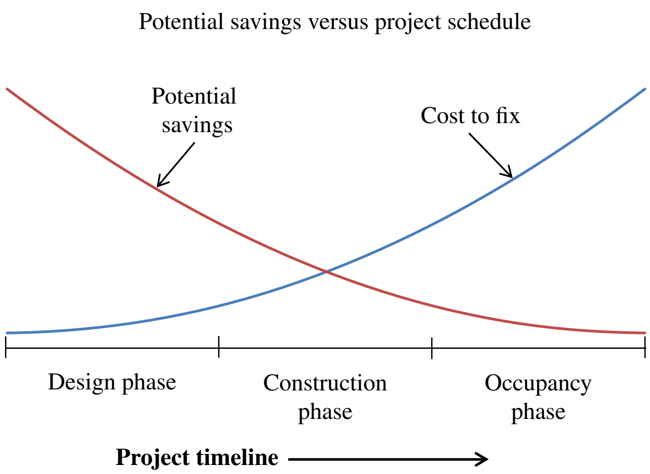


FIGURE 19.7 Potential cost/saving graph.

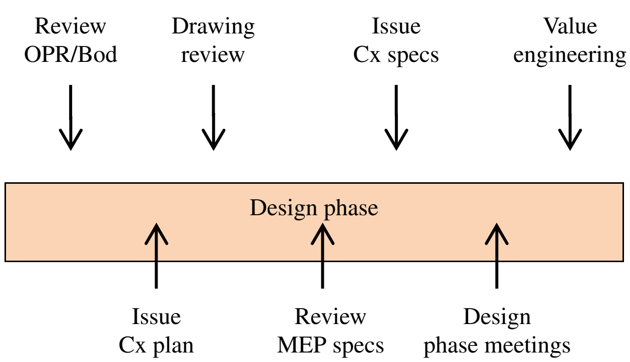


FIGURE 19.8 Design phase commissioning tasks.

19.3.7 Review of OPR and BoD Document

19.3.7.1 OPR Document The OPR is a vital document created by the owner. The OPR details the functional requirements of the project and the expectations of the building's use and operation. The criteria listed in the OPR shall be measurable, documentable, and verifiable. The document is vital to the design team for it provides all the critical information related to amount of redundancy and availability required in the design, as well as any other critical information to the operation of the building. It is recommended that the OPR address the following issues, as applicable to the project:

- **Owner and User Requirements**—Describe the primary purpose, program, and the use of the proposed project (e.g., office building with data center) and any pertinent project history. Provide any overarching goals relative to program needs, future expansion, flexibility, quality of material, and construction and operational costs.
- **Environmental and Sustainability Goals**—Describe any specific environmental or sustainability goals (e.g., LEED Certifications).
- **Energy Efficiency Goals**—Describe the overall energy efficiency goals related to local energy code or American Society of Heating, Air-Conditioning Engineers Standards or LEED. Describe any goals or requirements for building siting, landscaping, façade, fenestration, envelope, and roof features that will impact energy use.
- **Indoor Environmental Quality Requirement**—As applicable and appropriate for each program/usage area, describe the intended use, space environment requirements (including lighting, space temperature, humidity, acoustical, air quality, ventilation, and filtration criteria), desired for specific types of lighting, and accommodations for after hour use.
- **Equipment and Systems Expectations**—As applicable and appropriate, describe the required level of quality, reliability, type, automation, flexibility, and maintenance requirements for each of the systems to be commissioned. When known, provide specific energy targets, desired technologies, or preferred manufacturers for building systems.
- **Building Occupant and Operating and Maintenance Personnel Requirements**—Describe how the facility will be operated and by whom. Describe the desired level of training and orientation required for the building occupants to understand and use the building systems.

19.3.7.2 BoD Document The BoD is a document created by the design team. The purpose of the BoD is to convey to the owner and the commissioning agent the design team acknowledges the owner's requirements as stated in the OPR and will design the system to meet the requirements.

The BoD shall, at a minimum, include the following, if applicable:

- **Primary Design Assumptions**—including space use, redundancy, diversity, climatic design conditions, space zoning, occupancy, operations, and space environmental requirements
- **Standards**—including applicable codes, guidelines, regulations, and other references that will be followed
- **Narrative Descriptions**—including performance criteria for the HVAC and refrigeration systems, lighting systems, hot water systems, on-site power systems, and other systems that are to be commissioned

The commissioning agent's responsibility is to review the OPR and BoD document to verify all owner-detailed functional requirements for the project and expectations of the building's use and operations as they relate to the commissioned equipment and systems are documented in the BoD document.

The review of the OPR and BoD documents is normally included with all LEED accredited projects, but due to data centers being a mission-critical type of building, a document in the form of an OPR is valuable to the success of the project no matter if it is a LEED project or not. Equally as important is the design team's acknowledgment of the owner's expectations and design requirements by producing a document similar to the BoD.

19.3.7.3 Issuance of a Design Phase Commissioning Plan The Design Phase Commissioning Plan will be the first of two commissioning plans that will serve as the roadmap of how the commissioning process will be implemented. The second Commissioning Plan will be the Final Commissioning Plan that will be issued at the start of the construction phase and before the Commissioning Kick off Meeting.

The document will contain guidance for the participants involved, their roles and responsibilities, and direction for scheduling, implementation, testing, reporting, and documentation of the various stages of the commissioning process. This plan should be incorporated into the overall construction documents so that all parties in the construction process are informed regarding commissioning process. The Commissioning Plan will be distributed to the Commissioning Team for review and comment and will be used by the Commissioning Authority to execute the commissioning process.

A representative table of contents of a Design Phase Commissioning Plan is as follows:

- Purpose of the Plan
- Overview of the Commissioning Process
- Specific Objectives
  - Commissioned Equipment/Systems
- Roles and Responsibilities
  - Commissioning Team Members List
  - Commissioning Process General Rules
  - Commissioning Responsibility Breakdown
- Commissioning Management
  - Information Flow
  - Scheduling
  - Site Visit Protocol
  - Tracking Deliverables
  - Deficiency Reporting
  - Testing Strategy
  - Reports/Logs
  - Safety and Security
- Commissioning Process
  - Commissioning Timeline
  - Commissioning Task Overview
  - Design Phase Tasks
  - Bidding Phase Tasks
  - Construction Phase Tasks
  - Occupancy Phase Tasks
- Appendix
  - Draft Commissioning Schedule
  - Sample Commissioning Test Procedure Index
  - Sample Issue Resolution Log
  - Sample PFC document
  - Sample FPT document

19.3.8 Drawing Review

A drawing review by the commissioning firm is highly recommended, and preferably more than one review is conducted. The purpose of a drawing review is to bring to the table for discussion any changes that are necessary for constructability and to meet the design intent and OPRs. The importance of the review is to catch issues during the design phase before they are manifested during the construction phase. By catching the issues during the design phase, the cost to fix the problem is less, and the potential savings are greater than during the construction phase. Due to the complexity of data center systems, it is imperative to retain the services of a commissioning firm with extensive background in data center commissioning in order to catch issues during the design phase.

Multiple reviews are recommended during the design phase, and a review of the construction documents at the 75 and 95% issuance is the most common. At the 75% issuance, there is enough detail to catch issues before the drawings are too far along that changes found will be minimal. The 95% review has a level of completion to the point there is substantial detail and if there are issues found, there is enough time before the 100% construction documents is issued.

The focus on the drawing review will be dependent upon the phase of the construction drawings issuance and will focus in part on the following:

- Ensuring clarity, completeness, adequacy, and compliance to the OPR.
- Necessary details are provided for the development of the PFC documents for the commissioned equipment.
- Sequence of operation for the commissioned equipment is included on the drawings or in the specifications for development of the FPT.
- All commissioned equipment is scheduled.
- Clearance requirements are acceptable for the commissioned equipment to allow for maintenance and accessibility for the replacement of equipment or equipment components.
- Verify compliance to industry standard design issues.
- Verify compliance to many code issues

19.3.9 Mechanical, Electrical, and Plumbing Specification Review

Typically, the Mechanical, Electrical, and Plumbing (MEP) Specification review is conducted concurrently with the drawing review. The focus on the specification review will be dependent upon the phase of the specification issuance and will focus in part on the following:

- Verify if all OPRs related items are included in the specifications.
- Is a specification section included for each of the commissioned equipment?
- If the commissioned equipment sequence of operation is not included on the drawing, confirm the sequence of operation is included in the specifications.
- Verify if the standards and tolerances required for any of the following that is included in the commissioning scope of work are included in the specifications: duct pressure tests, water pipe pressure tests, airside testing and balancing, and waterside testing and balancing.
- Verify if there is any reference to the requirements for the following commissioning tasks:
  - Submittals related to the commissioned equipment
  - Start-up requirements for the commissioned equipment
  - Training requirements for any of the MEP equipment
- O&M manual requirements for the commissioned equipment.

19.3.10 Issuance of Commissioning Specification

A commissioning specification is vital to the success of a project and must be produced for every project and inserted into the project manual specification section of the contract documents. As a minimum, the commissioning specification should include the following:

- Responsibilities of the Commissioning Team
  - Including the architect/engineer, Construction Manager/General Contractor (CM/GC), Owner, subcontractors (mechanical, electrical, plumbing), controls contractor, test and balance contractor, LEED consultant, and commissioning authority.
- List of Equipment/Systems to be Commissioned
  - The list of equipment and systems to be commissioned will be identified in the fully executed contract with the owner.
- Execution of the Commissioning Tasks
  - All of the commissioning tasks to be executed during each phase of the commissioning process will be identified in the full executed contract. This section will identify the execution of each specific commissioning task and the involvement of each team member.
- Commissioning Documentation
  - Will identify all the documents produced by the commissioning authority during each phase of the commissioning process.

If the commissioning firm retained during the design phase of the project, the process is simple. The commissioning firm will issue a general commissioning specification to the architectural firm to be inserted into the architectural section of the specifications, and if the commissioning scope of work requires the MEP or other trades, the specifications related to these trades will also have a trade-specific commissioning specification inserted into their section.

If the services of the commissioning firm are retained after the design and bidding phase, the process is not as simple. If a commissioning specification section has been written prior to the hiring a commissioning firm, a copy of the commissioning specifications is to be submitted for review by the commission firm. The commissioning firm will review the specification to verify if the scope of work and roles and responsibilities are in agreement with the scope of work agreement between the owner and commissioning firm. If the commissioning specification has to be modified resulting in additional efforts by any of the trades, a change order will be issued to cover additional time and costs to the project. The same problem will occur if a commission specification has not been written after the design and bidding phase. The issue of not having the commission specification in the bid document is just another reason why it is important to obtain the services of a commissioning firm during the design phase and before the bidding phase.

19.3.11 Design Phase Meetings/Value Engineering

The attending of design phase meetings and participating in the value engineering process are not common commissioning tasks but occur when the budget allows. There are instances when the owner has had prior positive experience with a commissioning firm in which the owner recognized how the commissioning firm experience can provide added value during design phase meetings and value engineering. These two services can only be provided by a commissioning firm that has had experience in not only commissioning but also design experience related to the building being constructed.

19.4 BIDDING PHASE TASKS

See Figure 19.9 for more details on bidding phase tasks.

19.4.1 Attend Prebid Meeting

On very large or complex projects or when the commissioning effort is extensive, the commissioning firm may be requested to take part in the prebid meeting. The purpose of being at the meeting is to answer all questions related to the commissioning process and to answer any questions directed toward the commissioning specification (Fig. 19.10).

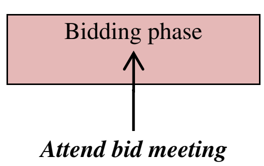


FIGURE 19.9 Bidding phase commissioning tasks.

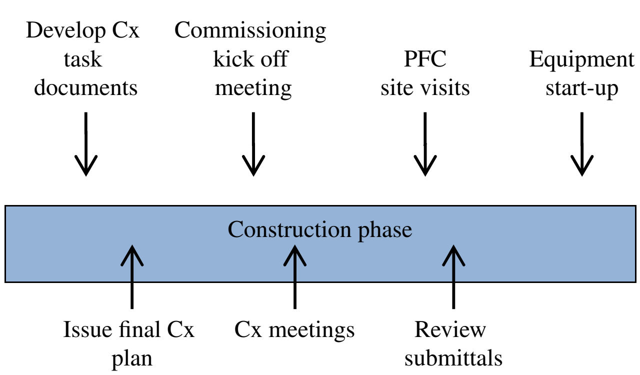


FIGURE 19.10 Construction phase commissioning tasks.

19.4.2 Final Commissioning Plan

If a Design Phase Commissioning Plan was produced during the design phase, the Final Commissioning Plan will be a slight modification to the contents of the Design Phase Commissioning Plan and a replacement of the documents in the Appendix. If a Design Phase Commissioning Plan was not created, a Final Commissioning Plan will be produced that will include the contents as described in the Design Phase Plan section earlier, along with the following modifications to the Appendix:

- Inputting an updated construction schedule
- Updating of the Commissioning Test Procedure Index
- Inserting the project-specific FPT documents for all commissioned equipment
- Insertion of the project-specific FPT documents for all of the commissioned equipment
- Inserting all Commissioning Task documents listed in the commissioning project scope of work document

19.4.3 Development of the Commissioning Task Checklist Documents

Once the 100% construction drawings and specifications are completed, the commissioning firm will be developing the project-specific checklists. The checklists to be developed will be dependent on the commissioning scope of work. A sampling of checklist documents to be created includes the following:

- PFCs
- FPT documents
- Drawing review document
- Duct pressure test review documents
- O&M review document
- OPR/BoD review document
- Pipe pressure test review document
- Start-up review document
- Submittal review document
- Test and balance review document
- Training review document

19.4.4 Commissioning Kick off Meeting

Prior to any of the commissioning field activities taking place, the commissioning agent will coordinate, schedule, and conduct a Commissioning Kick off meeting. The requested attendees at the meeting are to include, but not limited to, the following:

- Owner
- Owner's representative

- Construction manager
- Architectural/engineering (A/E) design team
- Mechanical, electrical, controls, and test and balance subcontractors
- Any third-party testing firms
- Other attendees as requested of the owner or owner's representative

The intent and purpose of the Commissioning Kick off meeting is to:

- Introduce all commissioning agent team members to owner, owner's representative, construction manager/ general contractor, and all entities involved with the construction phase of the project.
- Review of the Final Commissioning Plan to explain to the owner and the construction team the commissioning process, specific objectives of the commissioning process, roles and responsibilities of all entities involved with the commissioning process, and the commissioning administration process, commissioning scope of work, and commissioning tasks during each phase of the project and to provide a forum for the owner, owner's representative, and all team members involved in the commissioning process to ask any questions.

19.4.5 Commissioning Meetings

The commissioning agent will schedule and coordinate commissioning meetings as necessary throughout the construction and acceptance phase of the project. The attendees will include the owner/owner's representative, and construction manager/general contractor, and based on the nature of the issues to be discussed, other potential attendees will be the A/E design team and subcontractors.

The commissioning agent will issue an agenda prior to the meeting and a list of attendees required to attend. The two agenda items that will be consistent at each meeting are a review of the status of the commissioning effort and a review of all of the open deficiency items observed to date.

19.4.6 Submittal Reviews

The commissioning agent is to receive from the construction manager/general contractor a copy of the submittals for the equipment to be commissioned. The submittal will be used to aid in the development of the FPT and to verify compliance with equipment specifications for the commissioned equipment as well as verifying compliance with the OPR document.

The commissioning agent will develop a Submittal Review Document checklist identifying the engineer's specified requirements and the requirements of the OPR document. All references in the OPR and in the engineer's specifications not observed as being in the submittal will be noted on the submittal review document checklist.

19.4.7 PFC Site Visits

The objective of PFC site visits is to verify all equipment and systems to be commissioned are installed in accordance to the contract documents. The PFCs created by the commissioning firm detail how the commissioned equipment is to be installed as shown on the construction drawing and as described in the specifications.

Depending on the commissioning scope of work, the PFC documents will either be executed on-site by the commissioning agent or the subcontractor. It is highly recommended that the commissioning agent execute the PFC document due to the importance and critical nature of the equipment in a data center. If for some reason the commission budget is tight, the execution can be provided by the subcontractor. If the subcontractor executes the PFC, there should be at least a percentage of back checking by the commissioning agent to verify completeness and accuracy of the subcontractor's work.

Regardless if the commissioning agent or subcontractor visits the site, the execution of the PFC document involves going to the site with the PFC document and observing if the commissioned equipment is installed as indicated on the document. If there is an observed deviation from what is observed on-site to what is indicated on the PFC document, this deviation will be listed and documented on a deficiency list.

19.4.8 Equipment Start-Up

The engineer of record will specify which commissioned equipment will require a manufacturer's start-up. There are three tasks that can be completed. Any one the start-up tasks or all three can be completed, and the tasks to be completed will be dictated by the commissioning scope of work.

One start-up commissioning task is to review a blank copy of the start procedural document prior to initiating the start-up process. The procedural document will be provided by the equipment manufacturer and given to the commissioning agent for review. The commissioning agent will review the blank document to see if all specified requirements in the specifications are met.

The second start-up task that can be completed is to witness the start-up of the commissioned equipment. If there are multiples of the same equipment, a percentage of the equipment can be witnesses. The purpose of witnessing the test is to visually verify the test is conducted as per the manufacturer's recommendation and there were no issues related to the start-up.

The third possible start-up task is to review the final equipment start-up document after the start-up. The purpose of reviewing the final start-up document is to verify for completeness and any deviations from the manufacturer's recommended steps and the engineer's specified requirements.

19.5 ACCEPTANCE PHASE TASKS

The acceptance phase begins once the commissioned equipment is installed as per the contract documents and start-up has been successfully completed (Fig. 19.11). This phase includes the completion of additional commissioning task documents listed later, as well as performing different types of dynamic testing to the commissioned equipment and systems.

19.5.1 Duct Pressure Testing

Depending on the commissioning scope of work, there are two tasks related to duct pressure testing. One task is to visit the site during pressure testing of the ductwork and witness the duct pressure testing procedure. Typically, the engineer will specify duct pressure testing on medium- or high-pressure ductwork. Witnessing of the testing can be conducted on all of the medium- or high-pressure ductwork or only a sampling can be witnessed. The purpose of witnessing the pressure testing is to confirm the test is conducted as per the engineer's specified requirements and also the results of the tests are within the specified tolerances.

The second duct pressure testing task is to not witness the field testing but to review the final duct pressure test report. The intent of reviewing the final duct pressure test reports is to verify the engineer-specified tolerance for the testing is met.

19.5.2 Pipe Pressure Testing

Similar to the duct pressure testing, the commissioning scope of work will dictate one or both of the two tasks related to pipe pressure testing. One task is to visit the site during pressure testing of the piping and witness the pipe pressure testing procedure. The engineer will specify the piping to be pressure tested. A sampling of the piping to be pressure tested includes the following:

- Chilled water
- Condenser water
- Heating hot water
- Steam
- Steam condensate
- Domestic cold water
- Domestic hot water
- Fire sprinkler piping
- Or any specialty-type piping installed on the project

Witnessing of the testing can be conducted on all of the piping or only a sampling can be witnessed. The purpose of witnessing the pressure testing is to confirm the test is conducted as per the engineer's specified requirements and also the results of the tests are within the specified tolerances.

The second pipe pressure testing task is to not witness the field testing but to review the final pipe pressure test report. The intent of reviewing the final pipe pressure test reports is to verify the engineer-specified tolerances for the testing are met.

19.5.3 Testing and Balancing of Airside and Waterside Systems

The commission tasks for the airside and waterside testing and balancing is similar. With each system, the engineer of record will specify the method of testing and balancing to adhere to, as well as the acceptable tolerances. There are three tasks that can be completed. Any one of the start-up tasks or all three can be completed, and the tasks to be completed will be dictated by the commissioning scope of work.

One start-up commissioning task is to review a blank copy of the airside and waterside testing document prior to

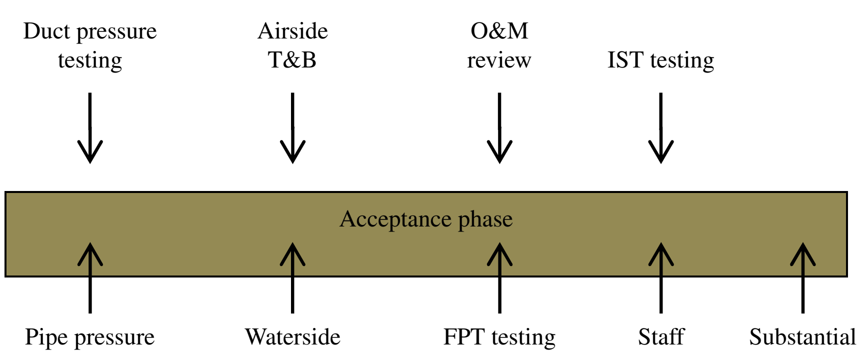


FIGURE 19.11 Acceptance phase commissioning tasks.

initiating the start-up process. The testing document will be provided by the testing and balancing firm and given to the commissioning agent for review. The commissioning agent will review the blank document to see if all specified methods and requirements in the specifications are met.

The second start-up task that can be completed is to witness testing and balancing of the airside and waterside systems. Not all of the airside and waterside testing is required to be witnessed, but a percentage of each airside and waterside system can be witnessed. The purpose of witnessing the test is to visually verify the test is conducted as per the manufacturer's recommendation and as per the engineer's specified requirements.

The third possible start-up task is to review the final airside and waterside testing and balancing documents. The purpose of reviewing the final start-up document is to verify for completeness and any deviations from the manufacturer's recommended steps.

A listing of the possible supply, return, and exhaust airside systems to perform one or more of the already-mentioned testing and balancing tasks is as follows:

- Supply and return air distribution units
- Supply air devices
- Return air devices with ducted returns
- Terminal variable air volume/constant volume terminal boxes
- Exhaust fans
- Exhaust air devices

A listing of possible waterside systems to perform one or more of the already-mentioned testing and balancing tasks is as follows:

- Chilled water systems including chillers, chilled water pumps, and all commissioned HVAC equipment with a chilled water coil
- Condensing water systems including cooling towers and condenser water pumps
- Hot water systems including hot water boilers, hot water pumps, and all commissioned heating hot water equipment with hot water coils

19.5.4 Review of Operations and Maintenance Manuals

The review of the O&M manuals is a very important task for data center projects. Due to the need for the equipment and systems to operate with minimal downtime, it is a must for the engineering and maintenance department to have a full and complete set of O&M manuals.

The commissioning agent will be producing an O&M manual review checklist document to confirm compliance with the engineer's specified requirements. As directed by the owner, other O&M requirements outside of the engineer's specifications will be included in the O&M manual checklist.

19.5.5 FPTs and IST Site Visits

The FPT and IST verification processes are two **absolute musts** when it comes to commissioning a data center. One of the main objectives during the FPT, and especially the IST process, is to confirm the two most important design intents associated with a data center: redundancy and availability.

The FPT process is a dynamic verification process testing the operation and functionality of the commissioned equipment to verify the equipment operates as an independent unit and during the IST process to verify the equipment operates interactively with other equipment and systems. Both the FPT and IST verification processes will use project-specific test procedures developed by the commissioning firm to verify the equipment operates independently and interactively in accordance with the controls contractor's sequence of operation submittal reviewed and approved by the engineer of record.

The commissioned equipment is tested under various conditions, modes of operation, and scenarios, such as design heat load, component failure, varying temperatures, fire alarm interlock, local and site power failure, and other modes necessary to verify the equipment is capable of meeting the operating requirements outlined in the BoD document and controls contractor's sequence of operation approved submittal.

The use of load banks during the FPT and IST verification process is one method of testing mechanical and electrical equipment under various operating conditions. Load banks are portable or trailer-mounted resistive or reactive "heaters" with adjustable outputs that are designed and used to create an electrical load at or near the peak design conditions for the testing of emergency generators, UPS, power distribution equipment, and computer equipment room cooling systems before the servers and switches that will ultimately occupy the space and produce the actual demand are installed.

Smaller suitcase or rack-mounted load banks are utilized during infrared scanning when the whole distribution system needs to be loaded and each panel is scanned at full load. The smaller rack-mounted load banks are also used for simulating loads in high-density areas, such as computer rooms, to verify that the air-conditioning system serving the computer room can handle the various operating conditions.

The commissioning firm will develop the FPT and IST documents in a sequentially written format based on the contract documents. These forms are used to plan, coordinate, oversee, and document the results of the execution of the FPT and IST documents. The FPT and IST verification process will be performed by the installing subcontractor or manufacturer's vendor by manual testing and manipulation of the equipment as required to verify proper operation. Monitoring the performance and analyzing the results obtained by using the building energy management systems trend log capabilities, stand-alone data loggers, and the monitoring and analysis of the electrical power quality to the commissioned equipment are additional methods of verifying the performance of the data center.

The execution of the FPT process will not begin until the International Electrical Testing Association (NETA) third-party testing activities have been successfully completed and a report has been submitted to the engineer of record and approved. The commissioning firm is not involved with the NETA process in any way but requires a copy of the NETA testing documents, as well as the megger test and torque data test reports for review. A sampling of the NETA third-party activities includes, but not limited to, primary injection testing of breakers 400 A and above, waveform captures for power quality monitoring, bus and connection resistance testing utilizing infrared scanning, and transformer turns ratio testing of the auxiliary voltage taps.

The execution of the IST process does not begin until the NETA testing and FPT processes have been successfully completed. The IST phase of testing will test all commissioned equipment and systems operating as they would during a period of normal or "real world" operations, as well as emergency scenarios. The integrated systems will be tested under various conditions, such as design heat load, component failure, maintenance modes, fire alarm, site power failure, etc. The focus on the testing will be on typical site-wide failure and maintenance sequences of operation experienced by the owner's facilities staff. Normally, these systems interactions are not covered during the FPT phase.

19.5.6 Training of Owner's Staff

The training of the owner's staff is another task that is very important with maintaining the operation of the equipment and systems at a data center after the building is turned over to the owner. It goes without saying that the more knowledge the owner's staff has in the operation of the equipment and systems, the greater the likelihood of the systems not operating properly is minimized, and if there are problems, the troubleshooting process is shortened.

The commissioning agent will be producing an Equipment Training Review checklist of all of the MEP equipment and systems specified by the engineer as requiring training. Other training will be included outside of the specifications if directed by the owner. The owner will submit to the commissioning agent the requirements of the training sessions.

The subcontractors, vendors, or manufacturer's representative will provide complete training to the owner's staff in the O&M of all equipment noted in the MEP specifications as requiring training. The construction manager and subcontractors/vendors will be responsible for developing the owner's training plan, scheduling of the training, execution of the training, and providing the necessary documents to the commissioning agent, verifying the specified training requirements are met. The commissioning agent will not be responsible for conducting the training but will be responsible for monitoring and documenting completion of the training to the building O&M staff.

For occupancy phase tasks, see Figure 19.12.

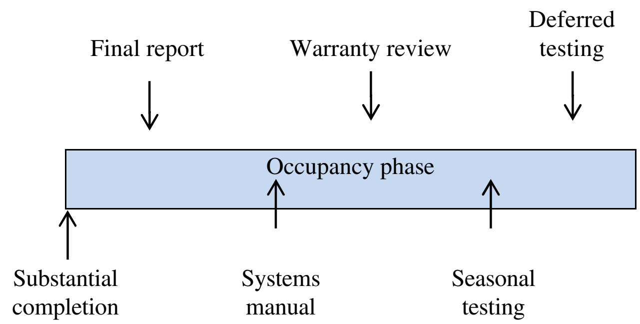


FIGURE 19.12 Occupancy phase commissioning tasks.

19.5.7 Final Commissioning Report

The final commissioning report will identify and document any deviations or variances between OPRs, BoD, contract documents, and as-built conditions. Notifications to the owner of potential shortfalls and recommendations for potential system optimization will be included in the Final Report. This report will be used to evaluate the system and serve as a future reference document for the O&M of the building's systems. The final report will also provide a benchmark for the electrical and mechanical system testing. Future tests can be compared to the initial test results.

A sampling of the contents of the Final Commissioning Report can include the following:

- Executive Summary Narrative
- Final Issues Resolution Log
- Site Visit Reports
- Project Update Reports
- Commissioning Tasks Deliverables
- OPRs/BoD Review Comments
- Drawing Review Comments
- Submittal Review Comments
- Completed PFCs
- Duct Pressure Testing Reports
- Pipe Pressure test Reports
- Start-Up Review Documents
- Test and Balance Review Documents

- Completed FPTs
- Training Documentation
- Operations and Maintenance Manual Review Comments

19.5.8 Systems Manual

The Systems Manual is another valuable document related to data centers that is a great resource for the building engineer and O&M department. The objective of the Systems Manual is to develop a document that provides future building operation staff the information needed to understand and optimally operate the commissioned equipment and systems. The Systems Manual will be produced by the commission firm and issued to the construction manager/general contractor for distribution to the owner's building operations staff.

The following will indicate the documents to be included in the systems manual and those responsible for submission of each:

- Final version of the BoD—engineer of record
- System single-line diagrams—mechanical construction drawing 1-line diagrams
- As-built sequences of operation control drawings including original set points—controls contractor
- Operating instruction for integrated building systems—O&M manual issued by the subcontractor
- Recommended schedule for retesting of systems of commissioned equipment and systems, if not included in the O&M manuals—subcontractor
- Blank FPT forms for retesting of the commissioned equipment and systems—Commissioning Agent
- Recommended schedule for recalibrating sensors and actuators—Controls Contractor

19.5.9 Near-Warranty Review/Occupancy Phase Review

The commissioning agent will coordinate and schedule the Near-Warranty End Review and Occupancy Phase Review meeting at the site approximately 10 months after the substantial completion date. The attendees at the meeting will be the commissioning agent and the building maintenance and/or engineer. It is recommended, but not required, that the construction manager/general contractor and owner's representative be present. The purpose of the meeting is to address the following:

- Any outstanding commissioning deficiency items not closed during the construction phase
- Any deficiencies that were noted by the operations staff during the warranty period
- Any reoccurring problems observed by the Building Maintenance and/or Engineering staff related to operating the facility as originally intended

If during the Warranty Meeting any deficiencies or performance issues are noted, a performance test will be conducted on the noted equipment and/or systems. The commission agent will contact the construction manager/general contractor to coordinate this activity. Tests will be executed, documented, and deficiencies corrected by appropriate contractor(s), with facilities staff and Commissioning Agent witnessing.

19.5.10 Opposed Seasonal Testing

The goal of the Opposed Seasonal Testing is to functionally test the commissioned equipment and systems during peak cooling and heating conditions. However, this is often not feasible due to the timing of the testing during the acceptance phase. If possible, this seasonal testing will be performed during the course of the commissioning process. If the seasonal conditions are not met during the regularly scheduled testing period, then it will be the owner's option to schedule this testing of the equipment and systems at a later date.

The commissioning agent will execute seasonal tests using approved test procedures and approved change management documentation. Testing shall be executed and deficiencies shall be corrected by the appropriate contractors or vendors in the presence of the Owner's facility operations staff and the commission agent. Final adjustments to equipment and systems will be reflected in updates to the O&M manuals and as-built drawings, which will be included in the Systems Manual.

19.5.11 Deferred Testing

Any FPT not completed due to building structure, required occupancy condition, or any deficiencies may be upon approval by the Owner deferred to a later date during the Occupancy Phase. These tests will be rescheduled as soon as possible. The construction manager/general contractor will coordinate this activity. Tests will be executed by appropriate contractor(s) with the deficiencies and results of the testing documented by commissioning agent.

19.6 LEED-REQUIRED COMMISSIONING TASKS

There appears to be a trend toward data centers becoming more energy conscious and sustainable, and thus, LEED certification of data centers is becoming common practice. If the client elects to pursue a LEED certification, there are certain minimum requirements of the commissioning agent. The minimum requirements include the following:

- The commissioning agent shall have documented commissioning authority in at least two building projects.
- The individual serving as the commissioning agent shall be independent of the project design and construction management, though they may be employees of the firms providing those services. The commissioning agent may be a qualified employee or consultant of the owner.
- The commissioning agent shall report results, findings, and recommendations directly to the owner.
- For projects smaller than 50,000 ft^2, the commissioning agent may include qualified persons on the design and construction team who have required experience.

If the owner is pursuing to only achieve LEED-certified status, the EA Prerequisite 1—FC tasks need to be met. If the owner desires to collect the points associated with obtaining certification above the LEED-certified level, the EA Credit 3 EC is required to be completed.

The following will highlight the tasks to be completed in order to achieve each EA Prerequisite 1—Fundamental and EA Credit 3 EC certification.

19.6.1 EA Prerequisite 1: FC

The tasks at a minimum to complete to meet the EA Prerequisite 1—FC requirements are as follows:

- Development of an OPR document
- Development of a BoD document
- Review of the OPR/BoD document
- Development of a Commissioning Plan
- Execution of the PFC
- Execution of the FPT documents
- Issuance of a Final Report

19.6.2 EA Credit 3 EC

The tasks at a minimum to complete to meet the EA Credit 3 EC requirements are as follows:

- Development of an OPR document
- Development of a BoD document
- Review of the OPR/BoD document
- Conducting a commissioning drawing design review prior to issuance of the midconstruction drawings
- Development of a Commissioning Plan
- Review of the submittals of the commissioned equipment
- Execution of the PFC documents
- Execution of the FPT documents
- Development of a Systems Manual
- Verification of the training of the owner's staff
- Issuance of a Final Report
- Review of the building operation within 10 months after substantial completion date

19.7 MINIMUM COMMISSIONING TASKS

All of the tasks listed in the Commission Tasks section are tasks that can be and have been associated with a data center project. If the commissioning budget does not allow for the commission firm to provide all of the services listed, the following will list those commissioning tasks that are highly recommended at a minimum during each commissioning phase.

19.7.1 Design Phase

- OPR/BoD review
- Drawing and specification review of the 75 and 100% construction drawings
- Issuance of a commissioning specification

19.7.2 Construction Phase

- Issuance of Final Commissioning Plan
- Commissioning Kick off Meeting
- Submittal Reviews
- Development of PFCs, FPT, and IST documents
- Witnessing start-up of commissioned equipment

19.7.3 Acceptance Phase

- Review of the final airside and waterside Test and Balance Reports
- Review of the O&M Manuals
- Site visit to verify proper installation of the commissioned equipment utilizing the PFC documents
- Site visit to verify proper operation of the commissioned equipment utilizing the FPT documents
- Site visit to verify proper system operation interactively of the commissioned equipment utilizing the IST documents
- Monitoring of the training of the building maintenance and engineering staff

19.7.4 Occupancy Phase

- Issuance of a Final Commissioning Report
- Issuance of a Systems Manual
- Near-Warranty/Occupancy Phase Review

19.8 COMMISSIONING TEAM MEMBERS

The commissioning firm selected to commission a project will more than likely consist of a project manager, lead commissioning agent, lead mechanical and electrical field agents, and several mechanical and electrical field agents. The commissioning firm will take the lead in the commissioning process and will provide the road map toward leading, planning, and coordinating a successful commission effort. The commissioning effort does not only involve the commissioning firm employees but other entities.

The obvious other commissioning team members involved with the commission of data centers other than the commission agent are as follows:

- Construction manager/general contractor
- Mechanical, electrical, and plumbing subcontractors
- Equipment vendors/manufacturer representatives
- Controls contractor
- Test and balance contractors

Other team members that are as important that at times are all too often excluded from the entire commission process are the design team (architects and engineers) and the owner or owner's representative. If the project is attempting to achieve an LEED accreditation, a LEED Consultant is another key member of the commissioning team.

The success of the commissioning effort involves a high level of and continual communication between the commissioning firm and the commissioning team. One very important item to discuss with the team members is the roles and responsibilities of each team member. The following roles and responsibilities of each team member is a sampling of the typical commissioning tasks each team member may be required to participate in and is not an all-inclusive list. The roles and responsibility will vary from project to project and will be dependent on the contractually agreed-upon commission scope of work between the commissioning firm and owner.

19.8.1 Roles and Responsibilities of the Commissioning Team

19.8.1.1 Owner/Owner's Representative It is highly recommended the owner have an actively involved employee of the organization or hire an owner's representative during the life of the project (Table 19.1). In addition, the building engineer and the head of the O&M department need to be actively involved. The commissioning agent's responsibilities are to verify all commissioned equipment and systems are installed and operate as per the contract documents, and it is the responsibility of the building engineer and maintenance department to operate and maintain the equipment once the building is turned over to the owner. Even though training is typically involved with the owner's engineering and maintenance staff, if they are involved during the commissioning process, they will already be familiar with the equipment and systems installed and can be better equipped to take over the project once the project is completed.

19.8.1.2 Construction Manager/General Contractor The construction manager/general contractor will work very closely with the commissioning agent during the entire commissioning process. Since the construction manager/

TABLE 19.1 Owner/owner's representative: Roles and responsibilities

Role	Responsibility
Write OPR	If the project is a LEED project, an OPR document is required. If the project is not a LEED project, an OPR or similar document detailing the owner's entire building and room requirements is vital in order for the design team to understand the level of redundancy, backup, and other critical data recommended in order to keep abreast of all changes to the commissioned equipment and systems
Review/comment on all Commissioning Plans	The commission agent will typically produce a draft and Final Commissioning Plan. It is imperative these documents be reviewed to confirm and understand the commissioning tasks involved and the intended plan to complete the project
Attend Commissioning Kick off meeting and other commissioning focused meetings	It is suggested the owner attend all scheduled commissioning-related meetings, but it is highly recommended the owner attend the Commissioning Kick off meeting since the commissioning plan will be reviewed and discussed in detail
Review deficiency log	It is important for the owner to review the open deficiency items to keep abreast of the deficiencies observed and also important to review the closed items to confirm they are in agreement with the outcome of the issue closed
Allow building access	In order for the commissioning efforts to be successfully completed, the owners must allow unobstructed access to all the areas and rooms to access the commissioned equipment and systems in order to perform all agreed-upon commission tasks

general contractor work is the conduit to all of the subcontractors and since the commissioning efforts heavily involve the subcontractors, the construction manager/general contractor will typically assign a point of contact the commissioning agent deals with during the life of the project (Table 19.2).

19.8.1.3 Design Team (A/E) The A/E design team is to provide all the necessary documents for the commissioning agent to develop all the commissioning-related task documents for the project (Table 19.3). Depending on the agreed-upon method of information transfer, these documents will either be directly submitted to the commissioning agent or through the construction manger/general contractor.

19.8.1.4 Subcontractors/Vendors/Manufacturer's Representatives The subcontractors are responsible for the completion of the following commissioning tasks but can rely on the assistance of the equipment vendor and/or manufacturer's representative in completing some of the tasks (Table 19.4). The following does not include the efforts required from the Controls or Test and Balance contractors.

19.8.1.5 Controls Subcontractor The controls contractor is an important entity in the commissioning process for data centers. Without proper installation and operation of the controls systems, the backup systems designed and built into the data centers will not operate and potentially result in system failures and downtime (Table 19.5).

19.8.1.6 Test and Balance Contractor The Test and Balance Contractor will be responsible for balancing the airside and waterside systems at the data center (Table 19.6).

19.8.1.7 LEED Consultant If the project is attempting to achieve one of the certification levels, a LEED Consultant is to be included with the Commissioning Team (Table 19.7).

TABLE 19.2 Construction manager/general contractor's representative: Roles and responsibilities

Role	Responsibility
Assist commissioning agent with subcontractors commissioning work	The construction manager/general contractor is to assist the commissioning agent in ensuring all subcontractors execute their commissioning responsibilities in accordance to the contract documents in a timely manner
Maintain/submit construction schedule	Submit to the commissioning agent the construction schedule and all schedule updates throughout the duration of the project
Issue Request for Information documents to the Commissioning Agent	Submit to the commissioning agent all Request for Information documents related to the commissioning equipment and systems
Review/Comment on Commissioning Plans	The commission agent will typically produce a draft and Final Commissioning Plan. It is imperative these documents be reviewed to confirm and understand the commissioning tasks involved and the intended plan to complete the project
Distribute Commissioning Plans	Distribute electronic copies of the commissioning agent-provided Commissioning Plan to the subcontractors and owner for review and use
Insert commissioning activities in construction schedule	Incorporate the commissioning activities provided by the commissioning agent into the construction schedule
Assist commissioning agent with Commissioning Kick off meeting	Attend the Commissioning Kick off meeting and coordinate with the commissioning agent the date of the Commissioning Kick off meeting with required attendees
Attend Commissioning Meetings	Attend all scheduled commissioning meetings during the duration of the construction process
Coordinate Testing dates with Commissioning Agent	Submit to the commissioning agent the dates for the execution by the subcontractor of the following tests that are applicable to the project. A sampling of tests include equipment start-up, pipe pressure testing, duct pressure testing, training sessions, airside test and balance, and waterside test and balance
Assist commissioning agent with collection of required documents from subcontractors	Gather from the subcontractors and submit to the commissioning agent all documents required to complete all applicable commissioning tasks. A sampling of documents that may be required are equipment submittals, blank start-up documents, completed start-up documents, O&M manuals, blank pipe pressure testing document, completed pipe pressure testing document, blank duct pressure testing document, completed duct pressure test document, training sign-in sheets, equipment maintenance schedule, control sequence of operations as-builts, and retesting and calibration schedule
Assist with closing of all open deficiency items and issues	Promote and assist in the timely responses and resolution by the appropriate trade to all open items on the commissioning agent's deficiency list
Assist with load bank testing	Provide the necessary resources for the operation, connection, and disconnection of load banks required for testing

TABLE 19.3 A/E design team: Roles and responsibilities

Role	Responsibility
Submit construction drawings to commissioning agent	Delivery of or access to the electronic copies of the required construction drawings to the commissioning agent to meet the needs of the commissioning process, including, but not limited to, the mechanical, electrical, and plumbing drawings
Submit project specifications to commissioning agent	Delivery of or access to the electronic copies to the commissioning agent of the mechanical, electrical, and plumbing project specifications document
Submit all contract altering documents to commissioning agent	Delivery of or access to the electronic copies to the commissioning agent of all contract altering documents including, but not limited to, any issued addendums, request for information documents, or scope of work changes of any kind through the term of the contract
Submit OPR to commissioning agent	Obtain from the owner the OPRs and submit to the commissioning agent
Submit BoD to commissioning agent	Document the design intent of mechanical, electrical, and plumbing systems in the form of the BoD document and submit to the commissioning agent
Attend Commissioning Meetings	Attend Commissioning Kick off meeting and all scheduled Commissioning Meetings during the commissioning process
Submit documents for Systems Manual to commissioning agent	Provide the following documentation to the commissioning agent required for the Systems Manual: final BoD document and single-line system diagrams of the mechanical, electrical, and plumbing systems commissioned

TABLE 19.4 Subcontractors/vendors/manufacturer's representative: Roles and responsibilities

Role	Responsibility
Review commissioning agent-produced documents	Review Commissioning specifications, Commissioning Plan, PFCs, and functional performance testing (FPT) procedures
Attend Commissioning Meetings	Attend Commissioning Kick off meeting and all scheduled Commissioning Meetings during the commissioning phase
Participate in the PFC verification process	The execution of the PFC documents is either completed by the Commissioning Agent or subcontractor. The subcontractor will execute the PFC if contractually obligated to do so.
Confirm operation of commissioned equipment	Prior to the commissioning agent visiting the site to execute the FPT process, the subcontractor is to confirm the equipment is operating as per the FPT document
Assist commissioning agent with the FPT/IST process	Assist commissioning agent during FPT and IST process by providing the necessary equipment and personnel and provide certified and calibrated instrumentation required during the testing time period
Assist in duct pressure testing	If required, provide a blank copy of the duct pressure testing document to the commissioning agent, the dates of the duct pressure testing, and a completed copy of the duct pressure testing documents for review by the commissioning agent
Assist in pipe pressure testing	Provide a blank copy of the pipe pressure testing document to the commissioning agent, the dates of the pipe pressure testing, and a completed copy of the pipe pressure testing documents for review by the commissioning agent
Participate in the equipment start-up process	Subcontractor shall execute equipment start-up per start-up plan, document results, and forward a copy of completed start-up checklists to commissioning agent for review to verify completion of start-up activities
Prepare Operations and Maintenance Manuals	Prepare and submit O&M manuals to the construction manager/general contractor. The construction manager/general contractor will forward the O&M manuals to the commissioning agent for review
Conduct Training sessions	The subcontractor is to prepare and submit a training schedule to the construction manager/general contractor. The construction manager/general contractor will forward the training schedule to the commissioning agent; coordinate the dates of each training session with the construction manager/general contractor and attendees; prepare an agenda of the topics to be discussed during the training session; prepare a sign-in sheet for each training session listing the topic, start and end time of the training session, and attendees; and, after completion of the training session, submit the sign-in sheet and agenda to the construction manager/general contractor. The construction manager/general contractor will forward the documents to the commissioning agent
Respond/close all open deficiency items	The subcontractor to respond to all open IRL items and provide a plan of action for closing all open items
Submit documents for Systems Manual to commissioning agent	Turn over the following documentation to the commissioning agent for preparation of a Systems Manual: recommended schedule/frequency for equipment maintenance requirements for all equipment to be commissioned.

TABLE 19.5 Controls subcontractors: Roles and responsibilities

Role	Responsibility
Attend Commissioning Meetings	Attend Commissioning Kick off meeting and all scheduled Commissioning Meetings during the commissioning process
Submit Controls Submittal	Prepare and submit the Controls Submittal to the construction manager/general contractor. Construction manager/general contractor will forward the submittal to the commissioning agent for review and use in the development of the functional performance testing (FPT) documents
Notify of control system readiness	The controls contractor is to declare the control system is operational as per the construction documents and the control sequence for each commissioned equipment has been tested as per the FPT checklist provided by the commissioning agent. Upon notification, the commissioning agent will coordinate a date to begin the FPT and Integrated Systems Test
Assist commissioning agent with the FPTs and IST process	Assist the commissioning agent during FPT and IST process by providing the necessary equipment and personnel and provide certified and calibrated instrumentation required
Run Trend Logs	Set up trend logs as requested by commissioning agent to substantiate proper systems operation.
Submit the required documents for Systems Manual to commissioning agent	Turn over the following documentation to the commissioning agent for preparation of a Systems Manual: as-built control sequences of operation for all equipment/systems to be commissioned and recommended schedule/frequency for recalibrating control sensors and actuators

TABLE 19.6 Test and balance subcontractors: Roles and responsibilities

Test and balance subcontractors—roles and responsibilities	
Role	**Responsibility**
Attend Commissioning Meetings	Attend Commissioning Kick off meeting and all scheduled commissioning Meetings during the commissioning process
Assist with Airside Test and Balance Process	Provide the following to the commissioning agent through the construction manager/general contractor: a blank copy of the airside Test and Balance Report, the dates for executing the airside Test and Balance Report process for witnessing of the execution of the Test and Balance Report process, and a copy of the completed airside test and Balance Report
Assist with Waterside Test and Balance Process	Provide the following to the commissioning agent through the construction manager/general contractor: a blank copy of the waterside test and Balance Report, the dates for executing the waterside Test and Balance Report process for witnessing of the execution of the Test and Balance Report process, and a copy of the completed waterside Test and Balance Report
Assist in the FPT/IST Process	Assist the controls contractor with setting up the sequence of operation for the commissioned equipment and during the FPT and Integrated Systems Test verification process

TABLE 19.7 LEED consultant: Roles and responsibilities

Role	Responsibility
LEED Facilitator	Manage the project's LEED online site
Involved with Credit Interpretation Rulings	Research Credit Interpretation Rulings and submit Credit Interpretation Rulings as required

19.9 DATA CENTER TRENDS

Data center's functionality and construction like many other types of buildings are an evolutionary process. Some of the data center trends that appear to becoming more common are as follows:

- Data centers are now going from allowing an opportunity for an orderly shutdown to being operational 100% of the time, creating extra design issues to address with backup capability, continuous power, and redundancy.
- Data centers were commonly located within an existing structure where the trend now is toward data centers being more stand-alone facilities.
- Since data centers consume a considerable amount of electricity, many are being constructed in remote locations to take advantage of low electrical rates.
- Data centers are looking to becoming more "green" in attempts to become more energy efficient and sustainable, resulting in seeing innovative use of mechanical system design, including the use of heat

wheels, economizers, and evaporative cooling, and other types of energy-saving systems and equipment are being explored.

- An increase in heat load density of servers and other equipment in data centers has been observed, causing an increase in cooling demand, resulting in cooling capacity being as critical as power availability.
- To address the additional cooling required, creative types of mechanical systems are being explored such as using ceiling distribution of air along with underfloor distribution as well as utilizing water cooled racks.
- Data centers are becoming more modular in nature in order to make quick and easier additions and modifications.
- Due to the increasing demand on data centers to be built and operational 100% of the time, there is an increase in threats due to extreme natural disasters shutting them down, causing extra efforts in planning for the extreme unusual natural events. Especially, data centers constructed along hurricane- or earthquake-prone areas.

19.10 CONCLUSION

The commissioning of data centers is an **ABSOLUTE** must. It is an absolute must due to zero tolerance to any downtime to a data center. When the services of a reputable commissioning firm with a proven track record, experience, and knowledge of successful data center commissioning are retained, the chances of any downtime and major issues with equipment and systems operation are drastically reduced. In addition, the owner will experience minimal amount of change orders, fewer or no "odds and end" problems during the occupancy phase, an efficiently operated building, a trained and knowledgeable workforce, and a high-quality building providing the comfort level meeting the design intent of all spaces within the building.

FURTHER READING

Mills E. *The Cost Effectiveness of Commercial Buildings*. Berkeley: Lawrence Berkeley National Laboratory; 2004.

Mills E. *Building Commissioning—A Golden Opportunity for Reducing Energy Costs and Greenhouse Gas Emissions*. Berkeley: Lawrence Berkeley National Laboratory; 2009.

Building Commissioning Association: http://www.bcxa.org/

PART III

DATA CENTER TECHNOLOGY

20

VIRTUALIZATION, CLOUD, SDN, AND SDDC IN DATA CENTERS*

Omar Cherkaoui[1] and Ramesh Menon[2]

[1]*University of Quebec à Montréal (UQÀM), Montreal, Quebec, Canada*

[2]*IBM Corporation, Gaithersburg, MD, USA*

20.1 INTRODUCTION

Virtualization and Cloud have significantly impacted the data center infrastructures and, in particular, the network infrastructure. There are many additional demands on today's data centers and an increasing number of devices and applications. In addition, servers and storage are becoming virtualized and Cloud computing is creating dynamic resource pools. And yet, data center cannot keep up because something keeps getting in the way: the network. The network is becoming increasingly complex to manage. Applications behave differently depending on where data is located and where it needs to flow. Depending on where it sits in the data center, it can be fast or slow.

The network in data centers has become a focal point. The current solution of adding more devices in order to scale has an impact, and the solution becomes exponentially complex. There is a need for solutions to simplify the problem while guaranteeing performance.

The network in data center is not completely virtualized, and there are still many network challenges impacting the new data centers. The first challenge lies in the ability to easily configure the data center network. The second is to facilitate network operations in a virtualized environment. The third challenge is around resource constraints, especially I/O. Choosing between different virtualization techniques is hence complex. It is important to select the technology and solution that will take into account the following criteria: overhead, I/O constraints, and flexibility in performing the required management operations (such as migration from one virtual machine (VM) to another).

Data centers are evolving at a fast pace, triggered by the venue of the mega data center infrastructure developed by Amazon, Facebook, Google, and Microsoft; the Cloud also played a significant role in the way the enterprise data centers are now built.

The low computation and storage cost resulted in the use of the same strategies and leveraging similar technologies at the network level than at the server level. Multiple network technologies, mainly based on open source approaches, gained interest from companies. Software Defined Networking (SDN)/Openflow [1, 2], Openstack [3], and OpenCompute [4] are some examples of open source approaches developed over the past years. Those network technologies coupled with the Cloud infrastructure are considered an enabler for the creation of mega data center.

The cost of the data center is well established; however, competition of cloud and data center has changed the way the data centers are being managed. An emphasis is placed on the management cost and the operations that drive the costs of data centers. An important issue is how to minimize operation costs. Companies are looking for technologies that will facilitate the migration or recovery operations and hence reduce the operation costs.

*This chapter was written by two independent teams in separate sections to give readers a complete perspective of information technologies. Sections 20.1–20.5 are written by Omar Cherkaoui. Section 20.6 is written by Hwaiyu Geng. Section 20.7 is written by Ramesh Menon. Each section represents the section author's view.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

In this chapter, we will first explain the virtualized infrastructure components present in the data center. We will then elaborate on the Cloud concepts and the related Cloud Service Offerings focusing mostly on Infrastructure as a Service (IaaS) offering. We will highlight the different issues related to building the IaaS service. We will explain the different elements in the design of a network for the new modern data center: what are the different considerations and how to dimension the network. We will learn about the management operations required for the data center network. Finally, we will present the SDN and Software-Defined Data Center (SDDC) concepts required for the network design and implementation for the new modern data centers.

20.2 VIRTUALIZATION IN DATA CENTERS

The virtualization is not a new concept for the computing environment. It was largely used in the 1960s for mainframe and forgotten for 20 years. It then resurfaced in early 2000. The paradigm for the virtualization from the mainframe to the server is the same.

Let's first define the concept of virtual servers. Server virtualization is based on the principle of operating simultaneously multiple virtual servers or VMs on a single physical server. Companies can hence operate using virtual servers instead of physical servers with the objective to improve utilization of server polling capacity and consequently reduce the investment required in physical infrastructure. Today, we are able to run between 10 and 40 virtual servers on a single physical server. In the future, it is expected to increase to hundreds of virtual servers.

With server virtualization, companies are not constrained to install a dedicated computer or server requiring a different operating systems or applications on those operating systems. Server virtualization allows running on a single server multiple operating systems and multiple applications.

Most of the virtualization technologies are based on the concept of a hypervisor. The hypervisor is an underlying kernel that allows isolation and a fair allocation of resources (I/O, CPU, Memory, etc.) between VMs. The virtualization technologies can be divided in two types: proprietary technologies and open source technologies. The open source solutions gained popularity in the past 5 years with the rise of Xen, KVM, etc. The first versions of Xen were developed at the University of Cambridge Computer Laboratory. The Xen community developed and maintained Xen as a free software, licensed under the GNU General Public License (GPLv2).

20.2.1 Benefits of Virtualization

This virtualization provides enterprises greater flexibility through rapid provisioning of the infrastructure. Improved flexibility can be achieved by the ease of deployment of the server infrastructure. The deployment has been significantly reduced from days/months to minutes in the case of the physical infrastructure. Additional flexibility comes from the ability to support legacy software as well as new OS instances on the same computer. Server virtualization hence leads to increased utilization.

Data center can also benefit from high redundancy and improved network reliability. One benefit is the dynamic fault tolerance against software failures through rapid bootstrapping or rebooting against software failures. Data centers inherit improved hardware fault tolerance through the migration of a VM to different hardware.

Another strong advantage is the fact that security is facilitated. Virtualization enables to securely separate virtual operating systems.

Virtualization also has benefits when working on development (including the development of operating systems): running the new system as a guest avoids the need to reboot the physical computer whenever a bug occurs.

Today, virtualization has moved from advanced products (such as VMWare) toward a large variety of Open Source solutions like XEN, KVM, etc.

The choice of virtualization technology is dependent on how the challenges mentioned earlier are being addressed by the proposed technologies.

20.2.2 Networking Challenges

The main challenge of virtualization comes from the network that is not easily configurable and yet not completely virtualized. Today, it is difficult to configure dynamically the network when VMs are added or removed or when there is a need to change the dimensioning of the network. The operational issue encountered is related to the fact that the solutions are not generic, which does not facilitate the implementation of data centers nor allow their expansion. The scalability of the solution remains the main problem to accelerate the creation of data centers.

Let's summarize, in this section, the main issues of the networking of the data center before it will be discussed in the further section:

- **Multitenant Support:** The data center is shared between business, administrative entities where each of them has its own QOS, management, and security requirements. Each tenant has its own set of VMs located over multiple servers. Each tenant has its own data center named **Virtual Data Center (VDC):** as an example, each tenant may require to have a full isolation of traffic for its VDC. It may also require its own Network Management to support workload mobility operation like live VM migration.
- **Separation Addressing:** Each VM has its own addresses: MAC Address and IP address. A strategic

separation needs to be done between the physical location of each VM and its logical view of VDC.

- **Topology/Hierarchy:** The network data center topology has a huge impact on the network in terms of bandwidth required between the VMs. Similar to the other network topologies (Enterprise Network, Network Provider Network), data center network topology is organized in a hierarchy. The topology of the network data center is differentiated by the ease of extensibility and scalability. The support of VM migration is another differentiator compared to the classical hierarchy network.
- **Workload Mobility:** Live migration strategies have stringent networking requirements.
- **I/O Blocking between VMs over the Blade server:** All VMs over the same server have to share the same network interface card (NIC). A VM traffic can block the traffic of the other VMs. The hypervisor requires to establish a fair allocation of network access between the VMs.

20.3 CLOUD AS AN EXTENSION OF THE DATA CENTER

Cloud can be defined as a computing infrastructure available in the network. Users have access to their information in the network. Cloud is a concept that complements the Internet by a complete availability of computing in the network. Users do not anymore need to have a computing environment. It can be available remotely.

Cloud enabled a redefinition of computing to include this new perspective: all computing resources are available remotely. Cloud computing offers three different models or technical use of Cloud Computing: IaaS, Platform as a Service (PaaS), and Software as a Service (SaaS).

20.3.1 Which Cloud (Private, Public, and Hybrid) and Which Services (IaaS, SaaS, PaaS, etc.)

Each of these models has a specific role:

- IaaS corresponds to the cloud infrastructure. It allows companies to outsource and to develop their physical infrastructure (servers, network, and storage) remotely and on demand. In IaaS, only the physical infrastructure (hardware) is dematerialized. For example, the data storage for backup purposes can be outsourced for companies that do not want to have that data internally.
- PaaS is a model that is "laying" on the IaaS. It allows a company to outsource not only the physical infrastructure but also the middleware applications, databases, data integration layer, and application development environments.
- SaaS is the final layer of cloud, the most complete, and the easiest to understand for the user. It allows the end user to access business applications hosted in a secure environment via an interface (it is sufficient to connect with his credentials via an interface). The company then uses its applications on demand, according to their actual needs.

We will focus the discussion on IaaS services as data centers are based on IaaS. When this cloud infrastructure is operated exclusively for an organization, the cloud is called "Private Cloud." For security reasons, many organizations choose to deal with their own private cloud. When an organization chooses to deal with services rendered over a network that is open for public use, the cloud is called a "Public Cloud."

Technically, there are no differences between the public and the private cloud. The public cloud may require more security features over the public network. Both of these cloud offerings use the data center technology to deliver IaaS services.

Cloud allows having access to all computing paradigms pushed to the Cloud. It first impacted the users who could have access to cloud environments to access data or applications; then it impacted companies that are interested in externalizing their entire computing infrastructure.

The benefits of the Cloud are to centralize the computing and facilitate the scalability for customers. The ability to configure the cloud in a flexible way greatly simplifies the management of the data center and hence reduces the inherent cost related to the management of the data center.

20.3.2 IaaS Benefits for the Data Center

IaaS technology provides companies greater flexibility in the use of their existing resources. IaaS facilitates the use of external resources now called the Public or Private Cloud: it also offers machines or virtual servers that are only used when you really need them. It is at this point that virtualization provides its full benefit because the company uses the resources it really needs, and thus, significant savings can be realized.

Another success factor of the Cloud is the reduction of computing costs. It enabled the deployment of Mega Data Centers.

20.3.3 IaaS Operations and Related Issues

The main challenging *IaaS* operation is the workload mobility (Live Migration), which consists of moving running VMs between physical servers for load optimization or maintenance purpose. This can be done intra or inter data center. This is a hypervisor-level operation as a service requires the provider to expose an interface in which hypervisors can

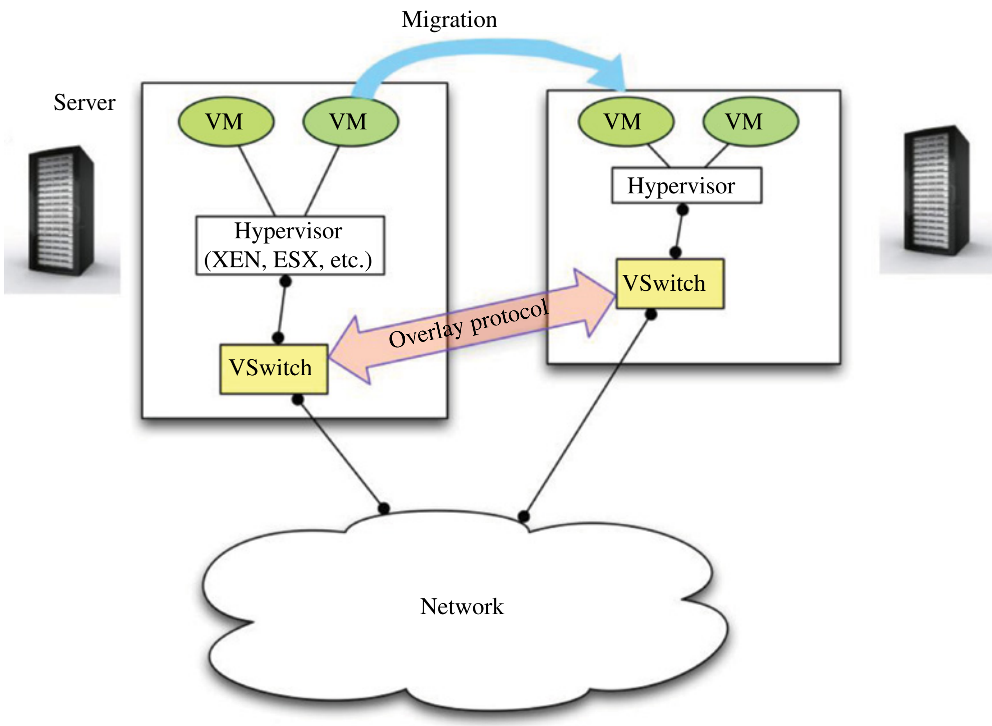


FIGURE 20.1 Migration of a VM in a data center.

transfer VMs to other hypervisors on potentially different Clouds. Figure 20.1 illustrates the basic operation of VM migration over an hypervisor. This operation needs to support a multitenancy.

To enable a VM to maintain its IP address, Live Migration is limited to a layer 2 network (Fig. 20.1). VLAN Services are not scalable to support these workload operations. The majority of solutions use an Overlay Network, which requires to run over the hypervisor a VM acting as virtual switch named Vswitch. Many overlay protocols are proposed such as VxLAN, STP, OVT, NVO3, and the Provider Backbone Bridging (PBB). PBB uses the hypervisor bridge to establish a tunnel based on the EVB/802.1Qbg Standard. This solution maps the Hypervisor VLANS into the PBB services. Other solutions such as VxLAN and NVO4 use Mac over IP and IP over IP.

On another hand, several virtual networking architectures are built to support VM Live migration. VL2 [5] and NetLord [6] create a virtual layer 2 network abstraction that can scale to hundreds of thousands of VMs, partially motivated by the perceived need for flexibility in VM migration and assignment.

20.4 NETWORKING IN DATA CENTER

In today's data centers, the network remains a fundamental challenge and solutions are required for configurable networks in data centers. This section will define the data center topologies and the evolution of network technologies.

20.4.1 Topology of the Network Data Center

The data center is composed of a few thousand servers, each containing VMs. Building the network for the data centers involves a trade-off between speed, scale, and cost. Connecting more than thousands of VMs requires multiple set of switches and routers organized in a hierarchical topology. This hierarchical topology is viewed as the number of ports west–east and north–south between the interconnected switches.

Data Center Networks are composed of three distinct tiers:

1. In the first tier, the server/storage devices are connected to the Top-of-Rack (TOR) switches. In order to reduce the cabling, each blade server has its own TOR switch where all servers of this blade are connected to the TOR switch. A set of blades is grouped via a south–north connection to the Edge Switches (ESs). A set of TOR switches is grouped together via a **Point of Distribution (POD)** where the ESs are connected through East–West links.
2. The second tier aggregates access switches to the core of the network. The ESs are interconnected to establish a POD. Those PODS are connected to the

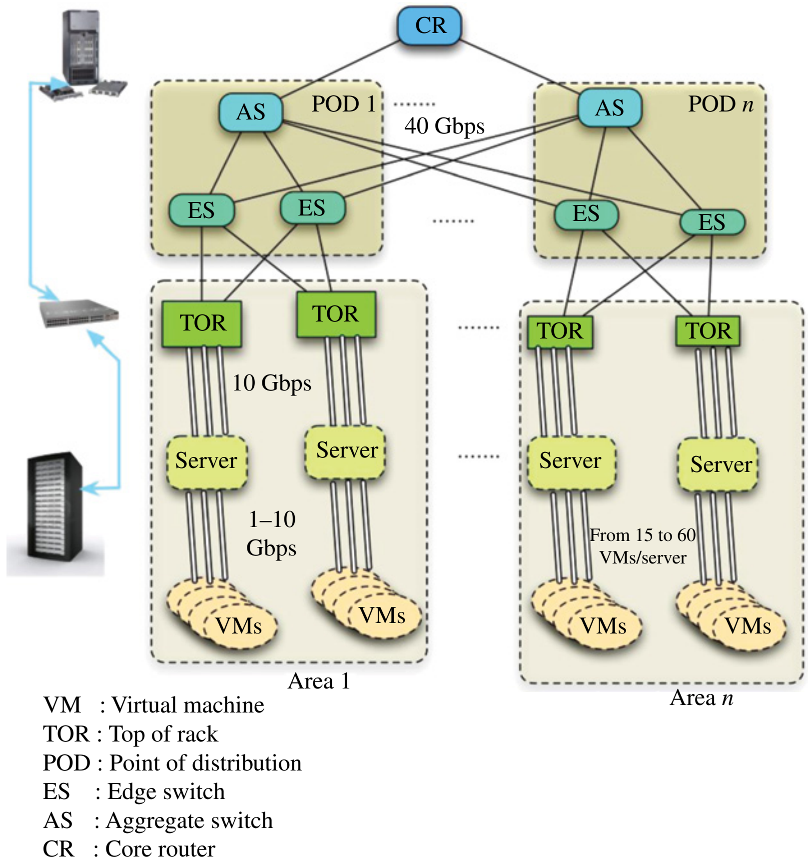


FIGURE 20.2 Multitier view of the data center network topology.

aggregate switches (ASs), which are now connected to the core switches/routers (CRs).

3. The third core network tier connects the CRs and allows communications to the outside world.

The port speed in the bottom of the hierarchy increases as we move up in the hierarchy. Server ports vary from 1 to 10 Gb. The TOR and S switches are based on layer 2 switching. AS switches are based on level 3 switching.

Figure 20.2 illustrates the interconnection hierarchy in the network data center with three layers of switches: AR, ES, and CR.

The POD is an interconnected mesh of ESs and the ASs. In the case where we do not have east–west connections between switches at the same level, the topology is a Fat Tree or Clos topology [5, 6].

20.4.2 Topology Challenges

The topology of the network data center determines the network diameter and its bisection bandwidth as well as the cost and the power consumption. The network diameter is determined by the number of switches or routers on the network section (also number of hop between servers) between two severs. The number of ports of a router or switch distinguishes low-radix routers, with a small number of ports, high-radix routers, with a large number of ports. High-radix chips divide the bandwidth into larger number of narrow ports, while low-radix chips divide the bandwidth into a smaller number of wide ports. The number of intermediate routers, in high-radix networks is greatly reduced and such networks enjoy a lower latency and reduced power consumption.

The main challenge is to give the maximum bisection bandwidth and reduce the diameter. The North–South and East–West connectivity between the PODs will have great impact on the diameter and the bisection bandwidth. The most popular approach is the Fat-Tree topology, and there are many great other good approaches proposed by manufacturers [5, 6]. Many other approaches based on Hypercube are under investigation by multiple network suppliers.

20.4.3 I/O Blocking and I/O Isolation

The number of VMs by server will continue to grow fast in the coming years, and each VM will require 40 Virtual Network Interface Card (vNIC) by server. One of the challenges today is to switch packets between the physical

NIC and the set of vNICs. The first challenge is the mapping between vNIC and the NIC Mac address.

The second and most difficult challenge is the traffic isolation between the vNIC and the NIC. As an illustration, XEN uses the Dom0 to bridge packets between the vNICS and the NIC. This can be accomplished in multiple ways: Direct I/O, Pass-through, or through the kernel of the hypervisor. Depending on the type of chassis and type of hypervisor, one of those techniques must be used to isolate the traffic between the VM. Most of the Ethernet cards support multiple MAC addresses and have embedded a small switch. Another solution for chassis approach is to push the TOR switches over the server board or to use PCI-E express as a bus between the rack servers. Another initiative from Facebook is the Open Compute where they define a new chassis by using a direct I/O to replace the TOR switch.

20.4.4 Multitenant Data Center and VDC

One of the main management issues of data center is the multitenancy. The data center is shared between multiple domains, which can be suborganization domain, applications domain, or zone domain (Fig. 20.2). Each tenant requires dedicated and isolated networking functions to manage its VDC. Each tenant has its own requirements of a level of security and privacy separating its resources from those of other tenants. For example, one tenant's traffic must never be exposed to another tenant, except through carefully controlled interfaces. The addressing schema between the tenants is an important issue: we need an address separation between tenants and an address separation between tenants and the infrastructure (Fig. 20.3).

For each tenant, VDC needs to offer similar services other than its physical counterparts. The tenant has to define its virtual network as set of services over the physical network such as L2 services (VLAN) or L3 services (VRF). Those virtual networks connect only the end station belonging to a tenant's specific virtual. As we said earlier, the main challenge is to support the workload mobility such as the VM migration in an isolated way. VLAN is the most used solution. Because the number of available VLAN is limited (in comparison to the increase of VMs), the solutions today try to extend the VLAN between the VMs. Those approaches are called overlay. The overlay principle is based on the creation of overlay networks on top of the physical network. It consists in dedicating a VM on each compute node for switching. Cisco proposed the Overlay Virtual Transport (OVT) to extend the VxLAN approach.

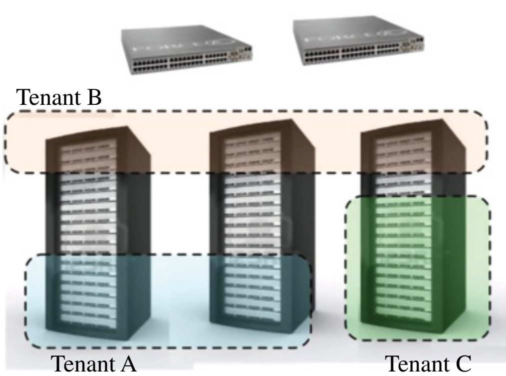


FIGURE 20.3 VDC multitenant.

20.5 SDN

SDN is a great solution for the data center Networking. It offers solutions to many of the issues presented in the last section. SDN was first proposed by Open Networking Foundation (ONF) [1]. This approach is adopted now by other forums like Openstack [3] and Opendaylight [2] (Fig. 20.4).

SDN can be defined by three principal elements:

1. A centralized control
2. A separation between the control plane and the data plane (control and forwarding functions)
3. A new interface allowing to program the behavior of the network

These three elements are mainly required for a data center.

With SDN, the switches are able to change the forwarding behavior through the control plane: a controller is used to change the forwarding behavior through a set of entry tables. SDN decouples the system that makes decisions about where traffic is sent (the control plane) from the underlying system that forwards traffic to the selected destination (the data plane). The concept is to break control over networking out of black box network switches, making it possible to write routing and switching rules in any programming language and run them on a standard server, the SDN controller. SDN simplifies networking and enables new applications, such as network virtualization, in which the control plane is separated from the data plane and implemented in a software application.

Figure 20.5 shows the elements of the SDN solution.

SDN proposed a new protocol named Openflow between the control plane and the data plane. Openflow pushes a set of entries from the controller to the switch through secure SSH channel connections. Many versions of Openflow are proposed by ONF. The first version, Version 0.9, supports only one table of flow entries. Version 1.1 supports multiple tables of flow entries. The number of match fields has now increased to 12 tuples. Unfortunately, many equipment manufacturers will not support Openflow protocol, but they are pushing their proprietary protocol. Through the Opendaylight

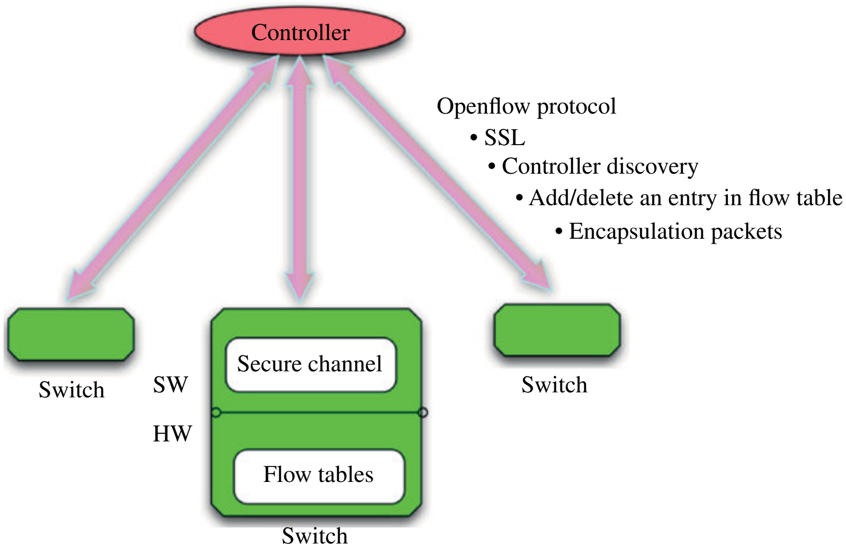


FIGURE 20.4 SDN controller approach.

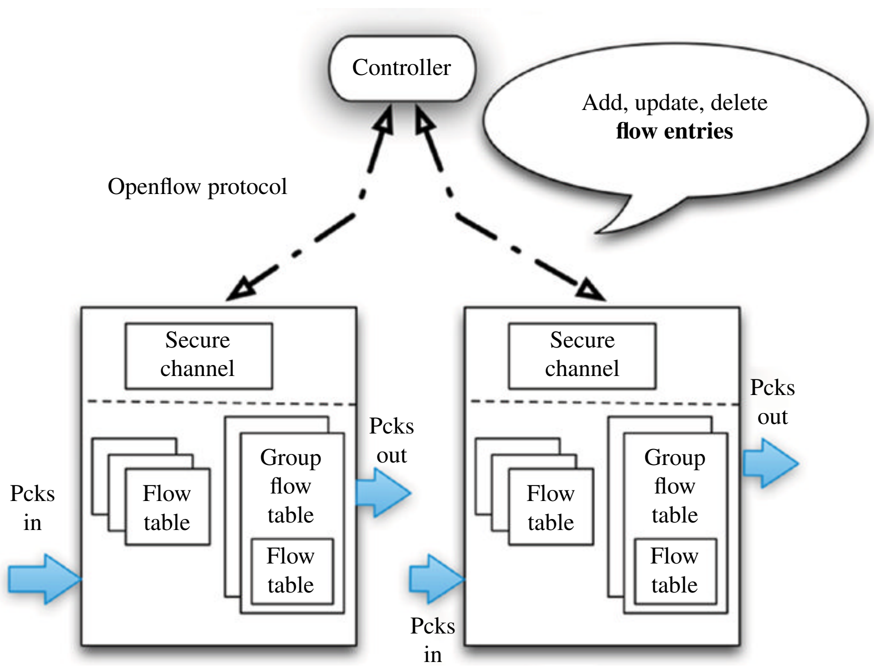


FIGURE 20.5 Openflow protocol.

forum [3], they are proposing a new API over the controller to enforce the network service over the switch.

In order to replace any manual configuration of the hardware to create and configure VMs remotely and configure firewall rules or network addresses in response, SDN offers a software approach based on automation applications to support the data center operations. SDN allows network administrators to have programmable central control of network traffic without requiring physical access to the network's hardware devices.

SDN is designed to turn networks into programmable platforms instead of collections of boxes that have to be managed on their own. While virtualized servers and storage can be reconfigured and moved around quickly from a central management point, networks still require many manual interventions.

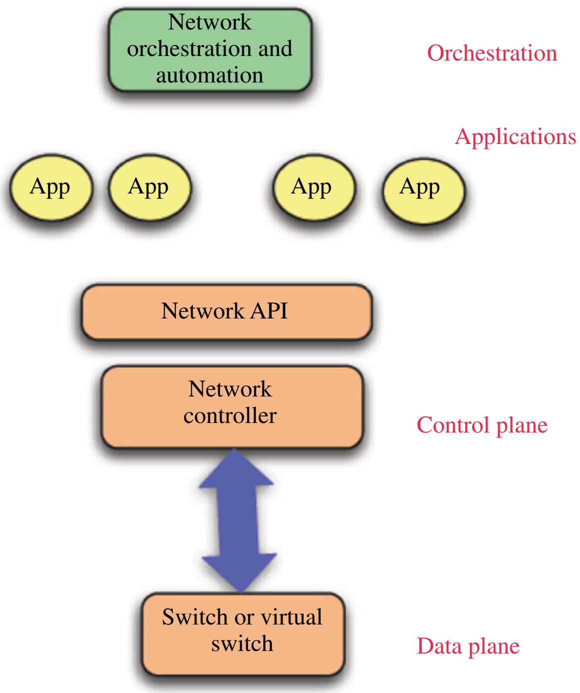


FIGURE 20.6 Opendaylight API.

SDN brings additional flexibility to the management of the data center by allowing a flexible configuration of the network. The benefits are the reduced cost of managing the network infrastructure in the data centers. SDN allows network administrators to manage network services more easily through abstraction of lower-level functionality into virtual services. We are now able to develop networking applications such as load balancing firewalling over the controller.

Figure 20.6 illustrates the network orchestration and automation at the control and data plane. At the control plane, the network controllers interact with switches or virtual switches to control their configuration dynamically.

20.6 SDDC

In a traditional data center, the infrastructure typically consists of hardware and devices that could take weeks to install. With increasing big data from mobile, social media, public, and Internet of things, they demand heavier applications for analytics, storage, and network. IT needs to move much faster to deploy increasingly larger and complex infrastructure from weeks to days or minutes. An SDDC, pioneered by VMware and recognized by industry, provides a faster, smarter, and inexpensive solution to significantly improve IT efficiency and performance.

20.6.1 Definition and Benefits

SDDC is defined as to virtualize all infrastructure components in a data center that include compute, storage, network, and security with comprehensive abstraction, pooling, and automation by software with little or no human involvement [7]. Automated software provides a framework managing logical compute, storage, network, and security services with little human involvement. SDDC provides a solution to support both legacy enterprise applications and cloud computing service [8].

All infrastructures in an SDDC can be delivered as a service leveraged as a private, hybrid, or public cloud. SDDC substantially reduces CapEx and Opex, allowing deployment of applications in minutes with automated management system. It achieves new levels of infrastructure utilization and employee productivity.

20.7 ROADMAP TO CLOUD-ENABLED DATA CENTER

IBM cloud computing reference architecture (CCRA) (Fig. 20.7) illustrates how to create both private and hybrid clouds where end users can self-provision compute, storage, and networks. Many large innovative organizations start off with an internal private cloud to augment their data center strategy.

As cloud computing evolves, so does the maturity of cloud-enabled data centers. Figure 20.8 reflects software-defined environment and cloud-enabled data center framework where the highest value comes from composable business services and APIs in cloud:

Virtualized (maturity level 1)

- Many organizations are at this level today and use virtualization to manage storage, network, or compute.
- However, at this level, there is still very little automation to support virtualization.

Deployed (maturity level 2)

- At the deployed level, the virtualization technologies are augmented with an automation layer.
- Basic management processes are established to track costs.

Optimized (maturity level 3)

- At this level, an organization incorporates additional capabilities to manage the infrastructure in order to reduce the operational costs and improves the SLAs and quality of service.

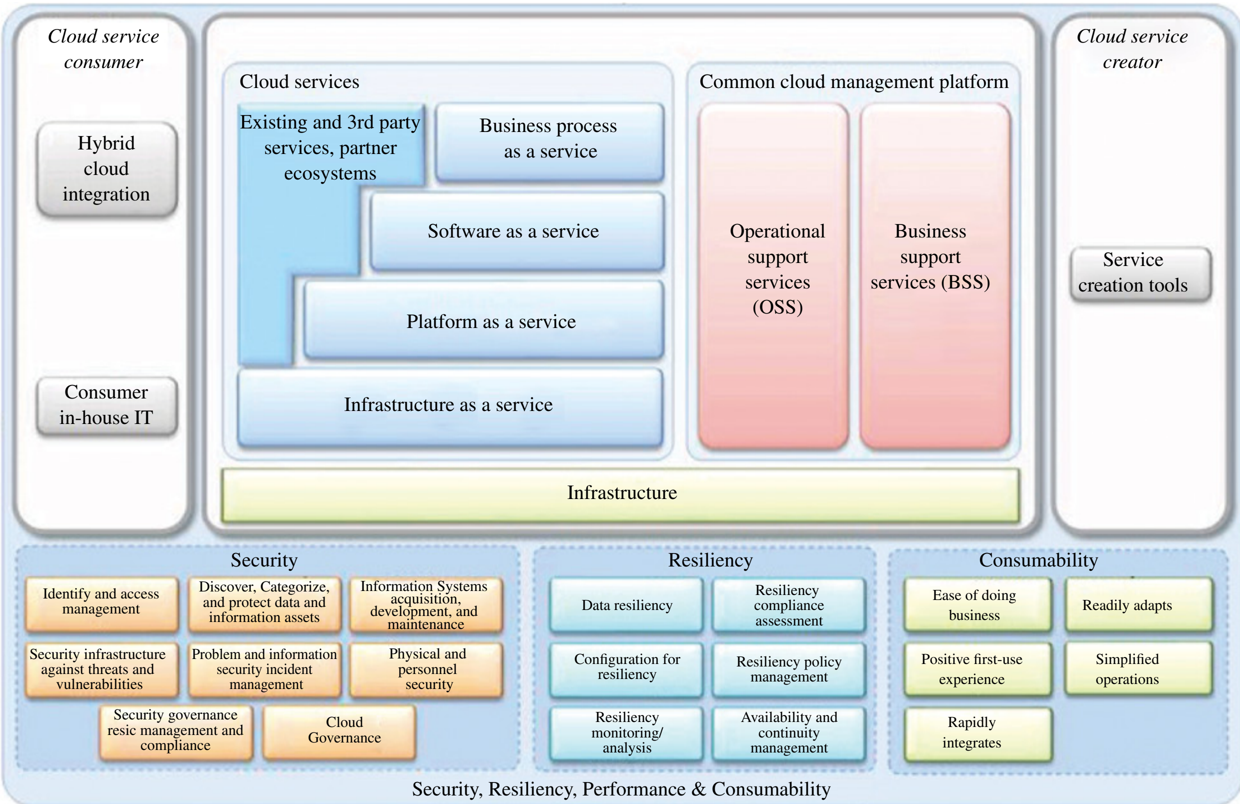


FIGURE 20.7 Cloud computing reference architecture (Courtesy of IBM Corporation).

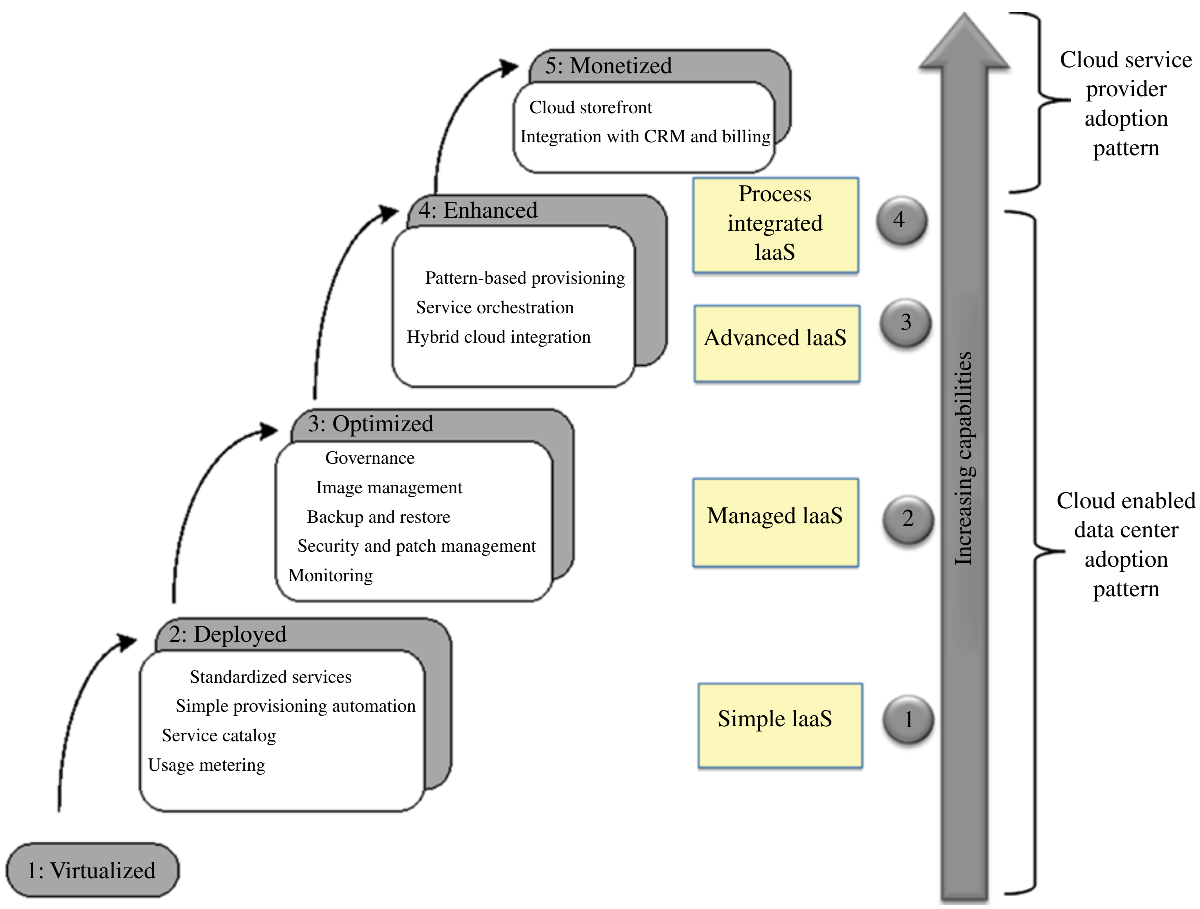


FIGURE 20.8 Maturity framework for Software-Defined Data Center (Courtesy of IBM Corporation).

Enhanced (maturity level 4)

- With level 4, the focus shifts to high-value services such as provisioning of application topologies, disaster recovery of cloud environment, and cloud-based backup services.
- Level 4 provides organizations with the ability to orchestrate provisioning of services across data centers and provisioning to off-premise public clouds to dynamically scale out and handle peak loads.

Monetized (maturity level 5)

- At this level, IT organization contributes to generate revenue by offering a utility service for obtaining compute and storage to other organizations or companies or to consumers or users.
- At this level, there are vigorous processes in place for service inception, development, offering, billing, and retirement and a greater focus on the ease of use and customizability of user interfaces.
- This level represents the natural evolution toward a Cloud Service Provider business model.

REFERENCES

[1] Open Networking Foundation (ONF). Available at https://www.opennetworking.org/. Accessed on May 20, 2014.

[2] Opendaylight. Available at http://www.opendaylight.org. Accessed on May 20, 2014.

[3] Openstack. Available at http://www.openstack.org. Accessed on May 20, 2014.

[4] OpenCompute. Available at http://www.opencompute.org. Accessed on May 20, 2014.

[5] Al-Fares M, Loukissas A, Vahdat A. A scalable, commodity data center network architecture. Comp Commun Rev 2008;38(4):63–74.

[6] Greenberg A, Hamilton JR, Jain N, Kandula S, Kim C, Lahiri P, Maltz DA, Patel P, Sengupta S. VL2: a scalable and flexible data center network. Comp Commun Rev 2009;39(4):51–62.

[7] Raghuram R. VMware software-defined data center products, vmworld; 2013.

[8] Software-Defined Data Center. Available at ttp://www.webopedia.com/TERM/S/software_defined_data_center_SDDC.html. Accessed on June 15, 2014.

FURTHER READING

Delivering on the promise of SDDC. VMware Report; 2013. Available at http://www.vmware.com/files/pdf/accelerate/VMW_13Q1_BB_SDDC_020813_FINAL_LTR.pdf. Accessed on June 15, 2014.

Fichera R, Washburn D, Chi E. The software-defined data center is the future of infrastructure architecture. Forrester Research; November 2012.

Thomas J. A road map to software-defined infrastructure. InfoWorld. Available at http://www.infoworld.com/t/data-center/road-map-software-defined-infrastructure-230713. Accessed on November 13, 2013.

21

GREEN MICROPROCESSOR AND SERVER DESIGN

Guy AlLee
Intel Corporation, Hillsboro, OR, USA

21.1 INTRODUCTION

Just as any good cooking recipe depends on the quality of the food ingredients, so too does the data center you achieve depends on the ingredients you use. Thus, this chapter concerns itself with microprocessor and server design: not so much on how to plan, design, and create them, but on how to judge and select among them as ingredients. Thus, in what follows are described, first, some guiding principles to aid your selection process, followed in detail by the prime ingredients, including the microprocessor and server system, as well as considerations with respect to storage, software, and racks.

21.1.1 Guiding Principles

Designing a data center to be green is a serious undertaking and needs to start with the end in mind. And that requires knowing the answers to the end state. What are your objectives? What constitutes “green” for you? If you are starting from an existing data center, how do you describe its current state? If not, what do you want to do differently from past data centers? This is the first of five guiding principles—start with the end in mind.

One way to approach this—and the next guiding principle—is to start with an Energy Pareto Chart (Fig. 21.1). What is the current state of energy use in your data center? Not only does this help you identify opportunities and waste, but it also helps you prioritize improvements and in what order.

The third guiding principle is to focus on productivity. Data centers currently consume about 2% of the electricity in the United States and 1.3% worldwide. Just to put that in perspective, that’s less than all six billion cell phones used on the planet. The trivial solution would be to turn off the entire data center. However, that would deny the world the benefit they provide in reducing the energy use of the other 98% of uses. In fact, the SMART 2020 report concluded that all information and communications technology (ICT) saves on the order of five times its carbon footprint than the rest of the economy. Thus, a strategy to make sure that we get the most for our energy investment is to eliminate all uses that don’t contribute to computing productivity.

So, how does one assess productivity? Ideally, you run the application and measure its results. For a green data center, that measurement will likely take the form of how much information processing you get for a given energy expenditure. If you are planning the data center or selecting hardware and software components without having them in hand for experiments with your specific applications stack, you will likely have to make do with productivity metrics that manufacturers publish. SPECpower, for example, is an industry-standard benchmark that is generally available and useful in assessing the energy efficiency of a server system.

With respect to the rest of the data center, Power Usage Effectiveness (PUE) is a good metric for helping you identify energy use that is wasted on overhead. (PUE is described in Chapter 32.) However, keep in mind that PUE is a facility-level metric and that it doesn’t differentiate between good energy use and bad energy use *within* the server. Nor does it assess the productivity of the server energy use. For that, you need to be using a productivity metric that relates key business value creation (revenue) directly to energy expenditure (costs).

The fourth guiding principle is implicit in the foregoing: measure and monitor your energy use and productivity. It is well known from ISO-9000 and Continuous Quality

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

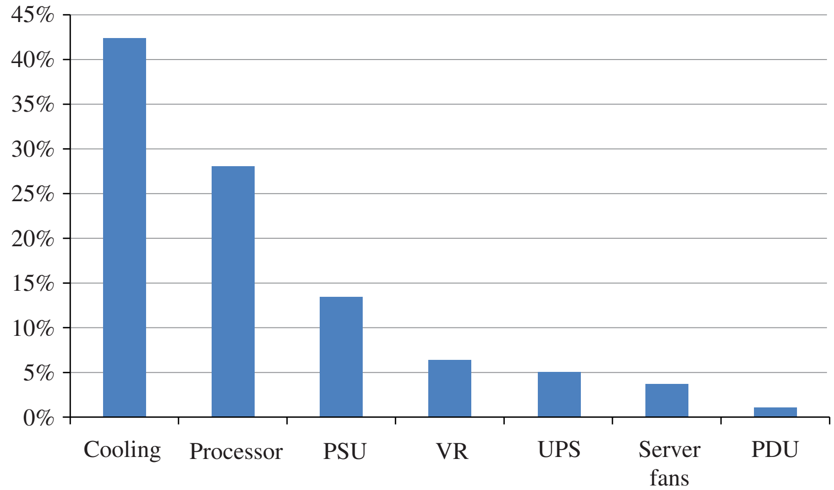


FIGURE 21.1 An Energy Pareto Chart of how the energy is used in a typical data center. Courtesy of Intel Corporation.

Improvement that "if you can't measure it, you can improve it." So put a system in place to capture your energy use—if not continuously, then at least periodically. Define the business value that your data center servers create and measure it. You undoubtedly already do; it would be rare if you didn't. All we want to do here is capture and correlate the value and the energy use.

The fifth guiding principle is to focus on optimizing energy, not just power. What's the distinction? After all energy, measured in Watt-hours, is just power, measured in Watts, integrated over time. Doesn't working in Watts to optimize power result in the same as optimizing energy? The short answer is no.

Here's an example based on current processors and servers. Take two server processors, one that runs at 15 W and one that runs at 100 W. Clearly, the 15 W processor is more energy efficient, right? Now, expand your view to the two servers, the 15 W processor in its 100 W server and the 100 W processor in its 200 W server. The 100 W server still seems more energy efficient, right? But, neither of these are energy; to get there, you have to look at the applications. And to get there, you have to look at the productivity on the application for which they are being used. Take a step back to the data center level and how you will have to provision it with these two different solutions. If the 15 W microprocessor takes four times as long to compute the application's result as the 100 W processor, each 15 W processor in each 100 W server will actually use 3.5 MW-h over the year to do the same work as the 200 W server at 1.75 MW-h. By only looking at power and only at a single component, you can save 85 W with the processor. But in so doing, you will waste 1.75 MW-h. It is only when the productivity is identical that the hours are identical—that you can take power as a shorthand for energy. Thus, focus on the energy: optimize Watt-hours first and, then Watts.

Finally, the sixth guiding principle is to upgrade older, inefficient equipment. Today's server is significantly more productive and energy efficient. Moreover, modern servers are designed with virtualization in mind so that you can run fewer of them with more jobs to keep them fully utilized and at the high end of their efficiency curves. Figure 21.2 shows a comparison between the latest generation of server and one that is 7 years old. As you can see, Moore's Law has significantly increased the productivity of the server. And, at the same time, the energy use has been cut by 75% at idle and 50% at full utilization. Because of the generational improvements in energy and productivity, today's server is 80 times more energy effective at SSJOps than a 7-year-old server. That is an 8,000% improvement at peak utilization (and even better at typical utilizations because modern servers are approaching the ideal of energy-proportional computing).

Hence, we have six guiding principles to keep in mind as we delve into the microprocessor, in particular, and server platform design, in general:

1. Start with the end in mind.
2. Use an energy Pareto.
3. Focus on productivity.
4. Measure and monitor the energy and productivity.
5. Optimize energy first and, then power.
6. Upgrade older, less efficient components.

In general, these follow the Organize, Modernize, Optimize rubric. With these in mind, it is appropriate to examine the microprocessor and server in detail.

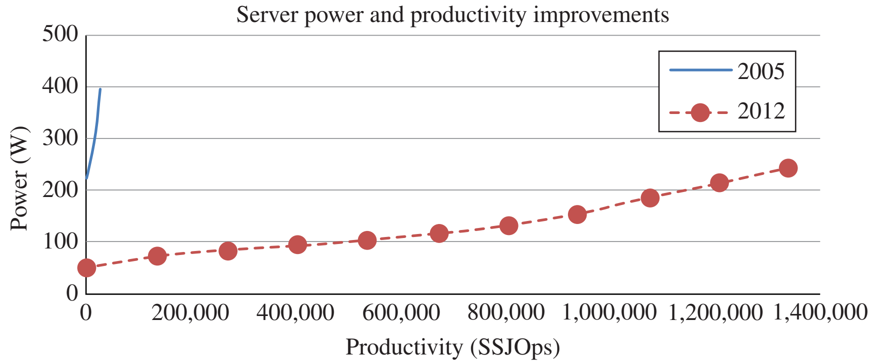


FIGURE 21.2 Productivity has increased and energy use decreased significantly from 2005 to 2012. The SPECpower metric compares server performance at idle and every 10% utilization increase. Courtesy of Intel Corporation.

21.2 MICROPROCESSOR

The microprocessor has become the engine of the information age. From the energy Pareto chart in Figure 21.1, it has the second most energy consumption in the typical data center. Thus, before we examine the server in depth, we will start with the major component of energy use in the data center—the microprocessor (Fig. 21.3).

From the perspective of creating green data centers, the microprocessor has a significant impact. From the energy Pareto, mentioned earlier, it is the second highest use and proximate cause of the first. As the industry builds more data centers with better PUE, it necessarily will succeed to the number one spot. When selecting a microprocessor, the energy use is dominated by several factors, including clock frequency, capacitance, voltage, process technology, microprocessor architectural complexity, and inherent power and thermal factors. We take each of these factors in turn in the following text.

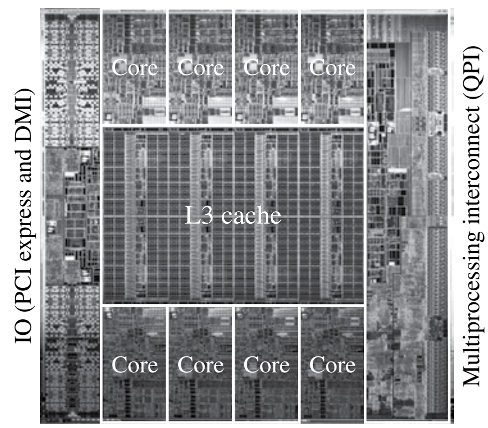


FIGURE 21.3 Photograph of Intel® Xeon® processor E5-26xx series die. The picture has been annotated to show the various major functional blocks: Cores and Uncore (everything else, including Cache, IO, multiprocessing interconnect, etc.). Courtesy of Intel Corporation.

21.2.1 Frequency, Capacitance, Voltage, and Dielectric Constant

Clock frequency (f) has historically dominated the design and selection of microprocessors. Up until the early 2000s, performance and productivity of a microprocessor was easily summarized by its clock frequency. However, the point of diminishing returns was reached at the 90 nm process node where further increases in clock frequency resulted in more incremental leakage power and less incremental performance. Clock frequency continues to be an important attribute, but it has to be considered in the context of the energy (not just power, see guiding principle #4, mentioned earlier)—how quickly can you do a computation and with how much energy? From a green perspective, this is basically productivity for your energy investment. The sooner the computation completes, the sooner you stop using energy.

The power draw of a microprocessor takes two forms, static and dynamic. Static power is the power used when no gates are being clocked (colloquially, Standby, or Sleep). Over the years, as semiconductors have been scaled to smaller sizes, the reduction in voltage used in a microprocessor has approached the fundamental limit for a MOSFET known as threshold voltage (V_T). As the MOSFET gate voltage approaches V_T, the amount of leakage current—energy loss that does nothing for computation—goes up. To the extent that you minimize leakage current, static power loss is minimal, and many of the recent chip innovations over the last 5 years have been to do just that.

Today, the predominant form of microprocessor power use is dynamic—the power that is used when you switch transistors off and on. It is the power that is wasted as you charge and discharge the capacitance in each transistor, a direct consequence of switching between on and off states. Capacitance (C) is an inherent property of an electrical device and a predominant factor in dynamic power use by a microprocessor. In fact, power has a direct relationship to the capacitance, the square of the voltage (V), and the clock frequency (f):

$$\text{Power} = \propto CV^2 f$$

Given that power is a function of the square of voltage, reducing the voltage has a big impact on reducing the power for green operation. However, there are real limits to how low you can reduce the voltage, because leakage and static power go up as you approach V_T. Conversely, as we scale down the size of microprocessor circuits, the voltage has to be reduced to use less material and waste less power. Failure to reduce the voltage as you physically scale down the size of a transistor increases the power by the cube of the scaling factor.

Today, the way out of this dilemma is to increase the capacitance while lowering the voltage. Dielectric constant (K) is the property that allows more capacitance at lower voltage in smaller physical volume. Moreover, the limit of frequency (and thus performance) is proportional to K times f:

$$\text{Performance} = \propto Kf$$

Thus, the path to optimizing performance and energy is a high-K dielectric. Since the 45 nm process node, the industry has addressed the problem by changing to high-K dielectric materials, like hafnium oxide.

Thus, to maximize performance and minimize energy, a microprocessor for a green data center needs to exhibit high frequency and use a low-K dielectric for low capacitance, and use lower voltages. For the most part, these all result from the semiconductor manufacturing process, which we examine next.

21.2.2 Process Node

As you can see in Figure 21.3, a modern server microprocessor is an incredibly complex engineering accomplishment and can have over a billion transistors in a few hundred square millimeters. In fact, over the last 40 years, the size of a transistor has been reduced in size by a factor of one million. The industry calls it Moore's Law and that inexorable progress has come from step changes in the semiconductor process about every 2 years and is effectively doubling the number of transistors each time. Each of those step changes is called a process node, which is labeled by the physical feature size of a transistor.

The process node is responsible for the physical characteristics of the transistors given earlier, for example, frequency, capacitance, voltage, and, inevitably, a microprocessor's power and performance. As of the writing of this, the industry was transitioning to the 28 nm node, with 22 nm node being used by the leading microprocessor manufacturer. Thus, a convenient rule of thumb for green microprocessors is to look for the one manufactured on the smaller process node, as it generally can do more work with less energy and fewer materials.

21.2.3 Microprocessor Architecture

Of course, what you do with those transistors has a big effect on the performance and energy. In the industry, this is referred to as the architecture. In the last 10 years, the industry has made a fundamental shift from getting more performance with higher clock frequency to creating architectures that get more work done while using less energy. This has been manifested in the shift to multithreading and multicore microprocessors. In this section, we will examine the main microprocessor architectures that you will select from, classified by the instruction set architecture, and follow up with a look at the microarchitecture, the function blocks within the microprocessor.

The main classifier of microprocessors is around their Instruction Set. While much promise accompanies the introduction of each new Instruction Set Architecture (ISA), over time, they tend to converge on the similar capabilities required by the market segments they sell into. There are four major classes of ISA that you are likely to encounter, including Complex Instruction Set Computing (CISC), Reduced instruction set computing (RISC), Very Long Instruction Word (VLIW), and The General Purpose Graphical Processing Unit (GPGPU).

CISC is the term applied to the ISA of microprocessors that evolved out of the 1970s. It represents microprocessor designs that emerged before some of the, then new, computing science had been applied to instruction set design. The designs were also incremental additions to earlier instruction sets and started before superscalar computing (executing one or more complete instructions in one clock cycle). The most familiar example would be the Intel and AMD "x86" microprocessor. Examples of these are the microprocessors such as Core, Xeon, Opteron, Atom, and Xeon Phi, which dominate the personal computer and high-volume server market segments today. The x86 microprocessors were historically driven by constantly increasing clock rate and performance as their main differentiator. However, in the last 10 years, there has been a concerted shift from clock rate increases to more performance from wider (32-bit to 64-bit), more parallel execution units in a microprocessor and lower power. This shift has resulted in significantly more performance for less energy, with the lowest performing x86, Atom, having higher performance at equivalent power to low-end RISC-based processors.

RISC was introduced in the 1980s and traded off simpler instructions for faster execution. One of the other consequences was to have fewer transistors and thus lower power than CISC implementations at the same process node.

Examples of RISC microprocessors include POWER from IBM, SPARC from Oracle/Sun, and ARM as an IP block in various SoCs. Today, RISC processors are used across the entire compute spectrum from embedded up to supercomputers. When first introduced, they had a significant speed advantage, but in the intervening decades, CISC and RISC implementations have been optimized and borrowed techniques from each other such that there is really no inherent advantage any more. More importantly, servers tend to wider 64-bit implementations, and as you increase the microprocessor to wider, multicore designs, the marginal differences between the two are converging.

VLIW was introduced in the 1990s as an alternative division of labor between hardware and software. The idea is to have the software compiler do a more in-depth analysis of the code and decide what instructions can be executed in parallel on the hardware, as well as using conditional execution as an alternative to branching. In addition, the expectation was that with simplified microprocessor hardware design, it would have less energy consumption and higher speeds. The most well-known example is the Intel Itanium, and it found success in some specific High-Performance Computing workloads and Mission-Critical applications, replacing seven of the eight mainframe architectures in the world in the early 2000s. The delays in working out the new mix of hardware and compiler complexity resulted in the products being released on older process nodes requiring more power to have competitive performance. Furthermore, extension of x86 to 64 bits has pushed VLIW to the very high end of the server market segments.

The GPGPU came about in the early 2000s from the application of Graphical Processing Units (GPUs) to traditional logic and computation. As the GPUs are originally designed for graphics, they can be convoluted to program for procedural workloads and arbitrary computations. They potentially can provide significantly more performance for a given energy expenditure. As GPGPUs are starting to implement standard IEEE floating point, they are becoming more accepted, especially as accelerators for HPC workloads. Noting that most server workloads are characterized by integer/character data and not floating point, GPGPUs are unlikely to be a universal solution in data centers.

With respect to green data centers, ISA is a loose determinant of energy use. All microprocessor architectures compete with each other and appropriate solutions and techniques from each other over time to remain competitive. As the world becomes more sensitized to the energy footprint of data centers, the competition necessarily includes energy use, with power as a proxy. Today, CISC and high-end RISC microprocessors dominate server designs. VLIW and GPGPUs are mostly applied in specialized HPC applications. Low-end RISC microprocessors are targeting server market segments, but as they boost their features and performance to be competitive (cores, threads, 64 bit, ECC, Virtualization support, I/O, memory, etc.), they necessarily increase their complexity and energy usage. It remains to be seen if there is a significant difference as the ISAs converge on the same set of requirements and features for green data centers.

21.2.4 Major Microprocessor Functional Blocks

A microprocessor is composed of various groups of circuits that provide a significant function to computation. Those groups are called Functional Blocks that are defined in the design process of functionally decomposing the requirements. They are features that can be independently developed and lend themselves to current practices in engineering high-complexity systems by breaking the systems into parallel development efforts with well-defined interfaces. At the highest level, a microprocessor can be separated into the core(s) and the uncore.

The core is the set of function blocks that have to do with program execution and includes the execution units, the registers, branch prediction, instruction and data caches, the scheduler, and pipeline management features. One can think of it as the historical microprocessor independent of the memory interface, higher-level caches, graphics, I/O, and multiprocessing features. Figure 21.4 shows the major functional blocks in a modern multicore server microprocessor. The main determinants of energy use have to do with the complexity of the core and the clock frequency and the process node on which it is implemented. How many bits wide is it? How many registers? How many executions units? Finally, at what frequency does the synchronous clock coordinate the executions among all the pipelined function blocks?

The uncore is essentially everything else within the microprocessor that provides the interfaces among multiple cores, shared cache, and the external server system

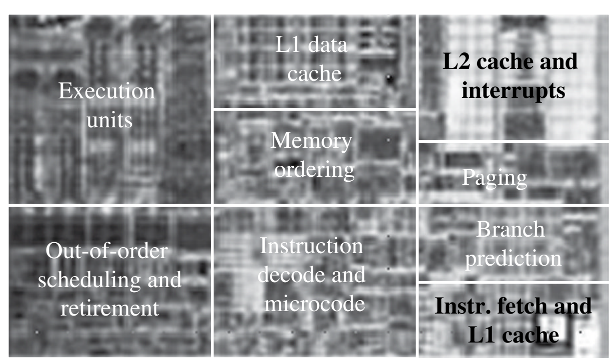


FIGURE 21.4 Photograph of a single core in the Intel® Xeon® processor E5-26xx series die, annotated with the major function blocks (one single core rotated 90° from Figure 21.3). Courtesy of Intel Corporation.

such as memory, I/O, and other microprocessors. Energy use at this level is primarily driven by integration scale—the number of items integrated. How many cores? How many memory channels and at what frequency? How big a shared cache is? How much I/O? How many multiprocessor interfaces?

The current trend in servers is to support multiple parallel executions of program workloads. That can occur at the board level, at the chip level, or within a core. At the board level, putting two or more microprocessors in the server is called multiprocessing. Today, with memory and I/O integrated into the microprocessor, the processing power, which scales with memory and I/O generally, scales with the number of microprocessors. Thus, the incremental energy for another processor with its attached memory and I/O can increase performance without substantially adding to the server board energy overhead. That is why dual-processor (DP) servers are the de facto standard for high-volume servers. Transaction-oriented workloads and HPC can benefit from even more processors in the same system.

At the chip level, putting two or more cores in the same uncore can provide more performance for workloads that don't need more memory or I/O. Server microprocessor chips with two or more cores is called multicore. Usually, the cores are identical, but a chip can also have a mix of big and little cores with different power and performance characteristics, such as the ARM Holdings' big.LITTLE architecture. One technique to match the energy use to the instantaneous computational workload demand is to turn off unneeded cores and run a single core at a higher clock speed, using, for example, Intel's Turbo mode. Conversely, if there is a mix of different cores, the high-energy-consuming cores can be turned off as the workload allows to save energy at lower utilization levels of the server (and lower performance, of course, appropriate to the lower instantaneous workload demand).

Today's cores operate significantly faster than memory or I/O, and an execution unit within a core can have to wait 500–1000 clock cycles for a memory access, burning power the whole time. One technique to improve this memory stall situation is to switch among multiple independent threads of execution when a memory stall occurs. This switching is called multithreading and has the impact of allowing a microprocessor to continue useful computation on a second parallel program thread while the first thread is waiting for a memory access to complete. Another way to address the stall is with Out-of-Order execution, which runs the next instructions that can be run while the stalled instruction is waiting and then making sure that any dependencies on that stalled instruction are accommodated at the end of the pipeline of execution. Other techniques such as branch prediction, instruction prefetch, predictive caching (and caches, in general), and load/store buffers all improve the performance at the cost of more energy used. In general, the incremental performance gain is significantly higher for the incremental energy cost.

Conversely, one of the techniques to improve performance that is wasteful of energy is speculative execution. This is the practice of executing both branches of a program before the result of the branch is known so that you get better performance. It tends to cascade in nested loops and ifs. Unfortunately, that ends up doing at least twice the work (for twice the energy) and throwing the computational results away. Ten years ago, when the industry was chasing frequency as the proxy for performance, the rule of thumb was to increase incremental power up to 3:1 for an incremental increase in performance; however, today, the industry is focused on no more than 1:1.

Thus, multithreading, out-of-order execution, branch prediction, prefetch, predictive caching are all features that a green microprocessor should exhibit, with speculative execution as a feature to avoid.

21.2.5 Virtualization, Power, and Thermals

Finally, with respect to the microprocessor, Virtualization, Power, and Thermal considerations warrant discussion. Virtualization is the ability to make the software think that there are more microprocessors than physically exist. This is accomplished by running a program called a Virtual Machine Monitor (VMM) to allocate the resources, and switch tasks, to appear to the software as if it is a machine with many more resources. There is an overhead (usually <1–5%) to running the VMM so that at the machine level, it actually takes more energy to run virtualized workloads. However, to the extent that you can turn off entire machines and consolidate all your workloads in a data center on fewer running servers (or newer, more efficient servers), the overall energy can be reduced. Green server microprocessors have specific features built into them for running virtualization with more performance and less energy.

There are power and thermal features inherent to the microprocessor at the chip level. The different functional blocks within a microprocessor can have different voltage requirements, and a server today can require tens of different regulated voltage inputs for chips on the motherboard. Voltage Regulators (VRs) are a big source of power loss within a server system (#4 in the Energy Pareto of Fig. 21.1). One of the new techniques is to have on-package or on-die voltage regulation, which is more efficient. In addition, higher efficiency spans a wider range of loads with higher-frequency digital control flattening the efficiency curve relative to the analog efficiency roll-off. Moreover, on-die VRs generally consolidate several other external VRs and provide several internal voltages at higher efficiency. This also has the effect of replacing several board-level VRs with a single VR that is more efficient and has lower cost overall.

Another feature is standardized support for changeable power states with both hardware and software hooks. The Advanced Computer Power Interface (ACPI) came about in the 1990s as a way to unify and improve platform power management. ACPI made platform power management a common feature of all platforms as a function directed by the operation system (Windows Server, Linux, etc.) rather than the previous practice of power management as a hodgepodge of platform-specific features through the platform BIOS. ACPI specifies different global and system power states, but here, we are concerned with the CPU states within them. The microprocessor CPU power states were initially defined as four different C-states: C0, operating; C1, Halt; C2, Stop Clock; and C3, Sleep. Additional gradations of sleep states have been implemented with names like Deep Sleep, Deeper Sleep, and Deep Power Down such that C-states exist up to C6 now. As multithreading and multicore solutions proliferate, distinctions are made around the thread, core, and package power states (e.g., thread-C6, core-C6, package-C6). A microprocessor does not have to implement all states, but when it does, it has to expose the interfaces and methods to drivers and the OS through the ACPI.

Many fine-grained power management techniques are applied at the function block and circuit level within the microprocessor. The system can also reduce the operating clock frequency and voltage within limited ranges to save energy at the cost of less performance, usually in a small number of discrete levels. This has occasioned a move to write the OS, drivers, and software to consolidate activities together in time and then "race to halt" to minimize the energy use until there is enough of something useful to do again. In this way, the same level of performance is maintained but stretches out the power-saving times so that the microprocessor can go into deeper energy-saving modes.

Microprocessors also have thermal features to track the temperature of the chip and even individual cores. These features are primarily to protect the chip from thermal runaway damage from self-heating by throttling (operating slower) if they get too hot. There tends to be a bit of margin engineered into platforms, so microprocessors have started using the thermal features to monitor the internal die temperatures so that it can actually take advantage of running faster during times that it is operating below the thermal limits. This leads to lower overall energy use by finishing calculations faster. The temperatures are monitored on millisecond timeframes, and the thermal implications are measured and projected over a window of a few seconds. Thus, when the package is cooler than the max power limitations, it can run faster without exceeding the thermal limits over time. In addition, it allows the microprocessor to use energy more effectively to run in full-speed bursts and complete calculations sooner without being overly restrictive because of static thermal margins designed into the platform.

Finally, a note on rated power: it used to be that the max possible power was the spec that microprocessor manufacturers published, usually called P_{max}. P_{max} was rarely ever experienced by a real system, and a lot of margin was wasted in designing a system to support that spec. It also resulted in oversizing components and, thus, wasted energy because those components were operating at lower points on their efficiency curves. Today, Thermal Design Power (TDP) is the published spec. As long as there is Thermal Throttling to protect the microprocessor in unusual and extreme events, one can safely design a system with much less margin and have it operate at this lower energy operating point.

21.3 SERVER

While the server microprocessor is a huge determinant of the energy use in a data center, in and of itself, it does nothing. It needs to be assembled onto a server motherboard and housed in a server platform to be of any practical use. Thus, the server platform is the physical unit of compute that one orders to populate a data center. In addition to the microprocessor that we have discussed in detail earlier, it contains the entire electrical, communications, mechanical, and thermal infrastructures to do computing. Major server platform components include the chassis, data storage, fans, I/O interfaces, Power Supply Unit (PSU), and the motherboard. Figure 21.5 shows these components in a 2U rack server with the cover removed.

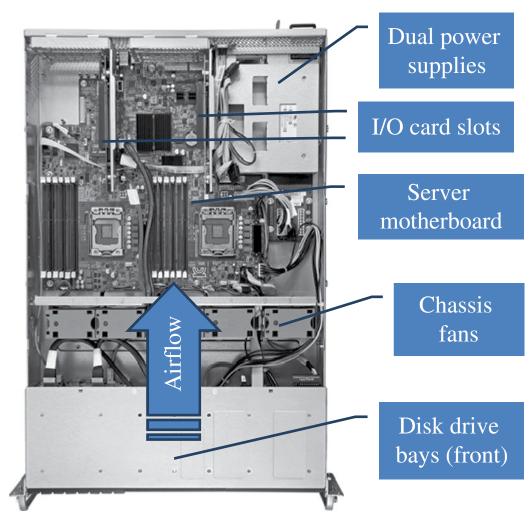


FIGURE 21.5 Typical 2U, Dual-Processor (DP) Rack Server seen from the top with the cover removed. Airflow is front to back through the rack. Note that the two microprocessors and their heat sinks are not installed, neither is the second PSU. Courtesy of Intel Corporation.

21.3.1 Chassis

Starting from the outside of the server, it is helpful to begin with a discussion of server types and form factors. There are four major classes of physical form factors for Servers: Tower/Pedestal, Rack, Blade, and Density Optimized.

A tower/pedestal form factor is basically the original Desktop PC box turned up on one side. They tend to have a lot of room for add-in cards and disk drives and, subsequently, a comparatively large PSU. Unless you are fully loading your Tower/Pedestal server up with cards and disk drives, it is likely running at the low end of the PSU rating and thus at the less efficient end of its efficiency curve. In general, the Tower/Pedestal server is used in stand-alone applications like workgroup, small business, or branch office server uses. About half of all servers are installed in unmanaged spaces (under a desk, back rooms, wiring closets); typically, they are Tower/Pedestal units. There is a potential for huge energy management gains with these platforms just by adopting best practices from commercial data centers.

Conversely, rack-mount servers are the mainstay of most data centers. They come in a variety of shapes and sizes, constrained by the internationally standardized width of 19 in. (482.6 mm). The racks are set up for vertical interval spacing on integer multiples of one rack unit (U), which is 1.75 in. (44.45 mm). Common 2-socket rack servers are 1U (“pizza box”) and 2U high and fully enclosed in metal (usually galvanized steel) for EMI compliance so that they can be sold individually. 4-socket and larger servers are sold in 3U, 4U, 5U, and even larger vertical dimensions. The smaller the rack server height, the smaller the fan diameter that will fit and, thus, the more challenging the heating and airflow issues. From a fan energy perspective, the smaller the enclosure height, the more disproportionate the energy use.

Blade servers eliminate some of the enclosure steel and interconnect cabling mess by sharing a rack-mount chassis of several U height (3–10U typically) with a mezzanine interconnect board and common networking, PSUs, and fans. Thus, each server card “blade” is just a motherboard with microprocessor, memory, chipset and chassis interconnect interface, and a faceplate that completes the chassis metal enclosure for EMI purposes. Blade servers tend to be lower-power, lower-performance solutions to provide high power density in a given rack volume. However, given the appropriate software workload, the overall energy efficiency can be optimized versus a pizza box server alternative.

Finally, Density-Optimized servers are in between a rack and blade form factor (e.g., four servers in a 1U rack form factor). They are distinguished by providing more than two microprocessor sockets per U of height.

21.3.2 Data Storage

In the end, computing is about manipulating data. It goes without saying that data needs to be stored before or after the computation and needs to persist through on/off cycles of the server. The traditional device for data storage is a disk drive, often referred to as a hard disk drive (HDD) with a mechanical rotating mechanism. The same device is also used for software program storage. Traditionally, servers have an HDD for software program storage as well as data storage. However, as the scale of data has risen, it is not uncommon for data storage and management to be a function of specialized systems within the enterprise or data center. Redundant Array of Inexpensive Disk (RAID) controllers provide for reliable data storage. Storage Array Networks (SAN) and Network-Attached Storage (NAS) provide enterprise-level data storage in lieu of local HDDs in a server.

Often, servers will have support boards for hot-plug disk drives and for RAID control of multiple drives. These are physically located with the disk drives, in or adjacent to the drive bays. They are connected to the motherboard by a SATA or SAS cable. SATA drives tend to have lower cost, lower speeds, lower capacity, and lower reliability and take less energy than SAS drives.

Recently, Solid-State Disks (SSDs) have begun to show up in servers. They typically use one-fourth the energy as an HDD with rotating media. For cost reasons, SSDs are offered in lower storage capacity ranges, although the lower energy, lower temperature, and lower latency of SSDs are accelerating their adoption. In some cases, they are being used to store application-critical data where their smaller size is an advantage; their low latency access can significantly improve the performance and reduce total energy to complete a workload. Finally, there are hybrid disk drives that use the rotating media of HDDs but buffer data in the same solid-state chips that are used in SSDs. These try to give the performance benefits of SSDs nearer the price of HDDs. Of course, it is a complex trade-off among the cost, performance, and energy factors, but SSDs and Hybrid HDDs should be considered if one is building a green data center.

21.3.3 Fans

All the energy used in a server eventually ends up as heat. Heat has to be continuously removed from an operating server to keep within the specified server operating temperature to assure reliable and nondestructive operation. Air is the medium used to remove the heat in most servers, and fans facilitate the removal. From the Server Energy Pareto, Figure 21.8, fans are a major energy use within a server. As indicated earlier, smaller U chasses limit fan size and drive up energy use. Solutions that move air at the blade chassis or rack level, rather than the server level, use less total energy.

Likewise, energy use is a power function of the fan speed. Servers that run the fan at full speed all the time waste more energy than they need to. A key to fan energy efficiency is to use sophisticated fan speed algorithms to control the speed based on system temperature and the amount of work that is being processed by the system.

21.3.4 I/O Cards

I/O cards plug into expansion slots or sockets in a server. They are used when the motherboard does not have the specific I/O interface built in. In general, I/O cards increase the energy use of the server. To the extent that one can identify an equivalent server that has the I/O on its motherboard, the energy use is typically smaller.

21.3.5 PSU

The PSU is a self-contained electronic assembly for converting the data center distribution voltage to a Direct Current (dc) voltage used by the components of the server system. All PSUs have an efficiency curve, which correlates the efficiency of the PSU at different, increasing loads from 0 to 100%. Efficiency curves are generally less at lower percentage load and highest at higher loads. The Climate Savers and 80-Plus initiatives over the last 10 years have moved the bar to require 80% efficiency from 20, 50, and 100% PSU load for servers, so you would want to makes sure that your systems have PSUs that comply.

Alternating current (ac) PSUs are always less efficient than dc PSUs because they have additional losses in the front-end rectifier and Power Factor Correction (PFC) circuits. Of course, it depends on the components and circuit design of the rest of the PSU for that to be true. In fact, one of the motivations for 380 Vdc Power Distribution in the data center is to eliminate these ac losses. (380 Vdc is covered in Chapter 28.)

For reliability reasons, some servers offer multiple redundant PSUs, called dual corded, from the point of view of data center power distribution. When you use dual-corded PSUs, each power supply is limited to no more than 50% of its full rated energy load capability (so when one PSU fails, the other can support 100%). As a consequence, each PSU is operating at a lower efficiency point on the efficiency curve. As an example, in Figure 21.6, the efficiency of the supply supporting a nominal 300 W server HW load is 89%, meaning it is drawing 337 W from the line cord. If the server is dual corded, you would have to supply 181 W to each of the two PSU line cords, or 362 W total (8% more), to support the same 300 W server load (150 W divided by 82.5% efficiency times 2). Architecting power delivery solutions into the data center that provide the same level of reliability without using dual-corded power distribution can significantly improve the energy use of your servers. Recently,

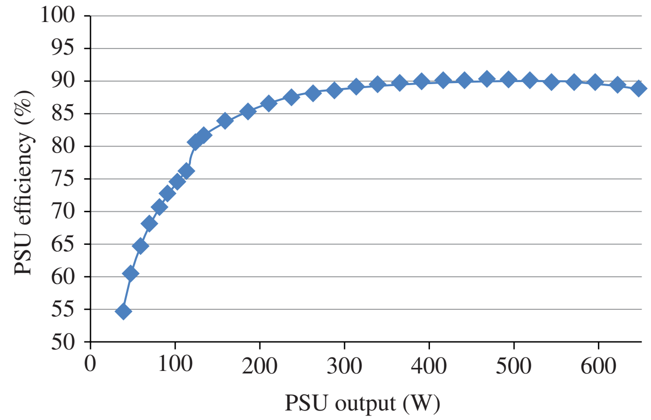


FIGURE 21.6 Power Supply Efficiency depends on the amount of power it is delivering and usually decreases rapidly at lower power output levels. Courtesy of Intel Corporation.

Cold-Redundant PSUs have appeared in the market that power only one of two redundant PSUs. They can instantly switch on the second PSU within the ride-through time of the primary PSU should it lose its input electricity. Thus, they can provide the reliability of dual corded with 5% more efficiency.

21.3.6 PMBus and Node Management

A recent feature of server power supplies is called PMBus, which provides the ability for a server PSU to communicate with the server's Baseboard Management Controller (BMC). Included in the capabilities is the actual power draw of the PSU. Thus, servers can now measure their own actual power use and PSU efficiency. Intel has provided a capability called Node Manager and released a Systems Development Kit (SDK) for the programmatic interface. Data Center Management applications can therefore measure and monitor actual power minute to minute. Data Center operators can use it to control and manage power use across the data center, including power capping, increasing computer density, dynamically balance resources, and improving business continuity.

21.4 MOTHERBOARD

The motherboard is the collection of server components that support the bulk of the electronic components in a server. It is mounted in the chassis and interconnects the integrated circuit "chips" such as the microprocessor (usually socketed), chipset and BMC, the VRs, the memory modules, I/O add-in cards, and integrated, on-motherboard I/O. Figure 21.7 shows the motherboard from Figure 21.5, out of the chassis.

Again, an Energy Pareto is a good way to look at the different components of a server. Figure 21.8 shows the major energy uses within a typical rack server without data storage. Given that the reason for a server is to compute, it should not be surprising that the microprocessor shows the highest use.

21.4.1 Chipset and VRs

The chipset is the hardware that interfaces the microprocessor on the motherboard to all the external and I/O interfaces. In general, microprocessors operate at less than 1.5 Vdc, whereas most external and I/O interfaces are in the 3.3–5 Vdc range. In addition, I/O signals—especially off-board—are generally lower frequency than on-board chip interconnects. As the mediator between the microprocessor and the external world, the chipset has energy impacts as well. With the ability to buffer I/O and directly access memory (DMA) and cache (DCA), the chipset can allow the system to go into lower energy states and work with the microprocessor to minimize the amount of energy used for a given computation. As more of the functions of the chipset migrate into the microprocessor, chipsets are following the C-state of the microprocessors, rather than independent P-states.

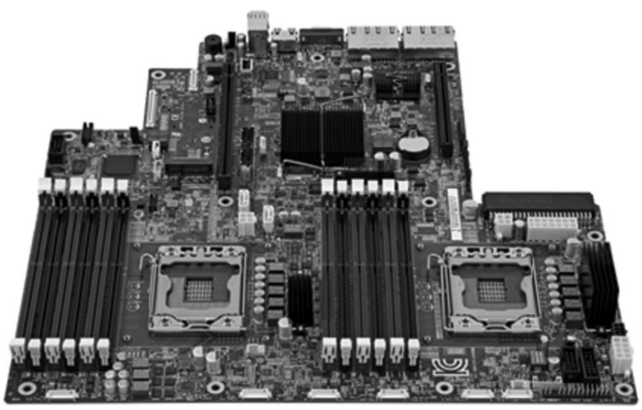


FIGURE 21.7 Intel® Xeon® E5-2400 Server Motherboard. Airflow is designed for front to rear to minimize the parts in the heat shadow of the microprocessors and memory DIMMs. Courtesy of Intel Corporation.

Likewise, the technology for on-die/on-chip VRs is migrating some of the VR functionality off the motherboard and helping the server become more energy efficient. It is not unusual for a chipset to support 10 or 20 different voltage levels for all the I/O and features that a chipset supports. Trends to on-chip/on-die VRs and the digital VR control that this technology enables are allowing for fewer VRs, higher energy efficiency, and flatter efficiency curves across wider utilization.

Thus, the important things to look for in the chipset with respect to energy are lower overall power draw, fewer voltage planes, and richer power-saving implementations (ACPI C-state/P-state support).

21.4.2 BMC

Standard high-volume servers include a controller chip that monitors the operation of a server called the BMC. The BMC allows for a data center operator to interact remotely with a server and monitor power, voltages, board and chip temperatures, fan speed, error logs, etc. The BMC is run from standby power in a server. The interaction is independent of whether the server is in the on or off state, as long as the PSU is plugged into power and the standby supply is functioning. In addition, a BMC can communicate over the server's on-board LAN interface (called in-band communications). Conversely, it can communicate over an independent LAN interface to an independent server management network (out-of-band communications). The BMC is especially important in identifying exceptional conditions on a server or in the data center in that it can provide messages and alarms when limits are exceeded. This in turn allows a data center operator to become aware of situations where energy use is out of the ordinary so that they can take action to restore the situation sooner.

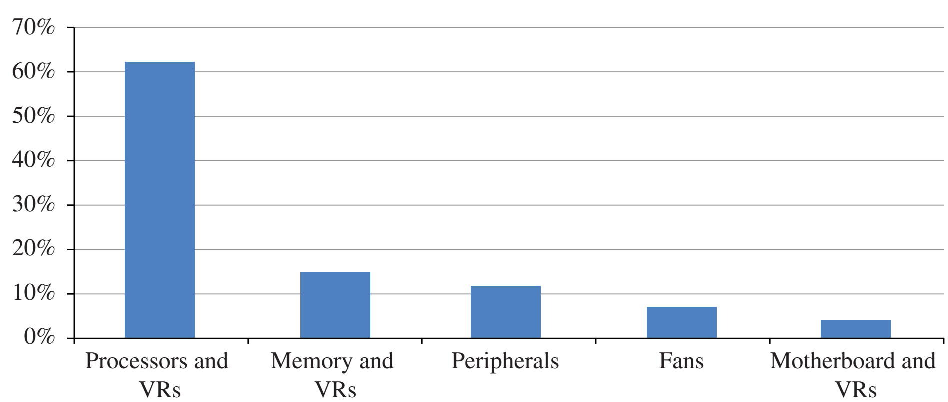


FIGURE 21.8 Energy Pareto for a typical compute server platform without integrated data storage. Courtesy of Intel Corporation.

21.4.3 Memory

After the microprocessor, memory is the next higher energy consumer, depending on configuration. Like the microprocessor, the amount of memory energy use is proportional to the clock frequency at which it operates. In addition, the number of memory channels is important. Today, system memory is Dynamic Random Access Memory (DRAM) chips, assembled as Dual In-line Memory Module (DIMM) cards. For the most part, the energy use of the memory system is a direct function of the number of total memory chips, independent of the capacity of each chip. Along the lines of memory speed, in the race-to-halt vein, faster memory results in computation completion sooner and can result in lower overall energy consumption. One has to validate it with a real software stack to know for sure. Likewise, low-power DIMMs exist, which run at lower voltages and provide slower performance. However, depending on the workload, the processor, and the proportion of time the server is at idle, it may be an appropriate trade-off that benchmarking tests can validate. In general, energy use is optimized by having faster chips, larger capacity, and fewer DIMMs in more memory channels for the amount of memory that your application needs (and no more).

21.4.4 Fans, Heat Sinks, and Heat Shadows

Next in consumption are the fans. Avoiding 1U servers allows more efficient fan diameters. In addition, most servers today vary the speed of the fans, running faster as the temperature of the server increases. Fan selection is largely at the discretion of the system manufacturer. However, the other big impact is the algorithm that the system software uses to vary the speed. Research shows that compared to traditional fan algorithms, a properly designed algorithm can save about 25% of fan energy at idle and 10% at full load. Further, heat sinks are required, but fan sinks (a fan integrated into the heat sink) generally represent an improperly designed system layout and additional energy expenditure. Motherboard layout can have a big impact on the fans, in that if components are laid out on the board in the airflow "heat shadow" of other hot components (say, the memory DIMM sockets behind the microprocessor), it can require more fan energy to remove the heat.

21.4.5 On-Board I/O (LAN, USB, and VGA)

Servers usually ship with card expansion slots for add-in cards, especially I/O. To the extent that one can define data centers that make use of the I/O integrated into the server motherboard, the additional energy to run I/O cards can be avoided. Examples of I/O that is usually provided "down" or on the motherboard are up to 4 channels of Gigabit Ethernet, USB, and SATA. SAS and RAID interfaces are sometimes available as motherboard options (rather than add-in cards that have been installed in one of the server's expansion card slots). Next, servers rarely need graphical display interfaces, so expect to use the native VGA port or even interact through the BMC over a LAN connection. Finally, small data centers have traditionally tried to multiplex several servers to a single set of operator I/O with a keyboard, video, and mouse (KVM) switch. This becomes impractical once you get to double-digit number of servers. There are solutions to provide this operator I/O multiplexing across Internet protocol (IP), which is certainly more resource efficient, as well as more energy efficient. There is no reason not to use this for small numbers of servers as well as for large ones.

Today, PCI express (PCIe) has essentially replaced PCI-X as the de facto standard for I/O card slots in servers. It certainly provides better data transfer and lower energy and has power-saving states that the platform can invoke. USB is another standard interface, usually for ad hoc or temporary connections to a server (say, a keyboard during debug). The emerging trend to keep an eye on, though, is the coming transition to optical I/O. Silicon Photonics can transmit data at 100 Gbps, with less than half the energy of today's 40 Gbps over copper. It is likely to replace both Gigabit Ethernet and Fibre Channel over time or at least the physical (PHY) layer for a number of I/O technologies as the transition occurs.

21.4.6 Server Utilization Effectiveness and Server Replacement Policy

One of the most effective things that you can do immediately at a data center level for overall energy improvement is to replace the oldest servers. Figure 21.2 shows that servers that are 7 years old have 80 times less performance per Watt than a new server.

In a recent census of a data center's server population, one Fortune 100 company found that 32% of their servers were more than 4 years old. Figure 21.9 shows the shocking result of an energy and performance analysis of those servers. Although servers more than 4 years old were only 32% of the total number of servers, they accounted for more than 60% of the energy consumption. Moreover, their productive output contributed only 4% to the total performance. Clearly, older servers use a disproportionate share of the energy to produce an abysmally small fraction of the work in this data center.

Server Utilization Effectiveness (SUE) is the metric that Intel has started using to describe this phenomenon. If Power Utilization Effectiveness (PUE) is about optimizing the data center power and cooling infrastructure, SUE is about optimizing the servers therein and over time. It is the ratio of the Server Efficiency today to the Server Efficiency currently deployed in a data center.

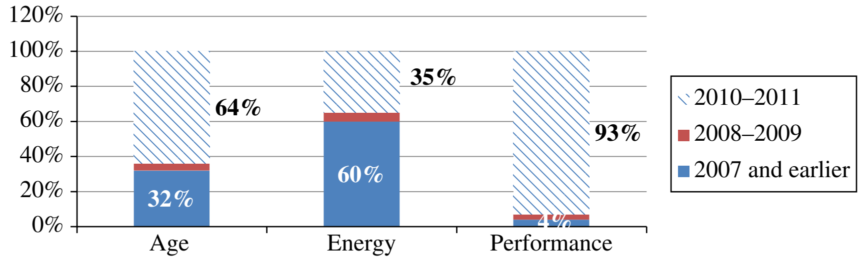


FIGURE 21.9 In a recent server census in a Fortune 100 company's data center, 32% of the servers that were more than 4 years old consumed 60% of the energy while only producing 4% of the computing output performance. Courtesy of Intel Corporation.

$$\text{SUE} = \frac{\text{Currently Available Server Efficiency}}{\text{Deployed Server Efficiency}}$$

Given the Moore's Law progress of doubling performance about every 2 years and the historical record of improving energy efficiency every few years, the compounded effect is an exponential curve that can be used to approximate the performance and energy effectiveness of aging server populations. Fitting the data empirically to SPEC performance results provides the following easy way to estimate the impact of the age of a server on the energy and performance impact it has:

$$\text{SUE} = \frac{1}{\sqrt{2}^{\text{Age[years]}}}$$

Thus, a server that is 1 year old is only 71% as effective as a new one, and a server that is 5 years old is only 18% as effective as a new one. Clearly, this is not an exact answer, but it is a heuristic that allows you to answer a class of significant data center planning questions.

SUE can also be calculated for an entire data center or organization as a weighted average of the server population. From this, one can make informed estimates of the impact of server replacement policies, for example. Take the case of a policy of replacing one-fifth of your servers each year; your SUE would be 56% over all the servers you own:

$$\text{SUE} = \frac{1}{5}\times\frac{1}{\sqrt{2}^{0}}\times\frac{1}{5}\times\frac{1}{\sqrt{2}^{1}}\times\frac{1}{5}\times\frac{1}{\sqrt{2}^{2}}\times\frac{1}{5}\times\frac{1}{\sqrt{2}^{3}} \times\frac{1}{5}\times\frac{1}{\sqrt{2}^{4}} = 56\%$$

You can also calculate the impact of cutting the budget for this year's replacement servers:

$$\text{SUE} = \frac{1}{5}\times\frac{1}{\sqrt{2}^{1}}\times\frac{1}{5}\times\frac{1}{\sqrt{2}^{2}}\times\frac{1}{5}\times\frac{1}{\sqrt{2}^{3}}\times\frac{1}{5}\times\frac{1}{\sqrt{2}^{4}} \times\frac{1}{5}\times\frac{1}{\sqrt{2}^{5}} = 40\%$$

Thus, cutting this year's spending will give you a data center that is 16% less effective without the refresh. In some sense, this is an evaluation of the lost opportunity cost; it is definitely an evaluation of the competitiveness impact to your data center. Additionally, if your organization's computing growth projections for next year are greater than 0, you have created a problem. Furthermore, for this case, if the growth projection is greater than 16%, you may have turned a short-term spending cut into an emergency project next year to build a new data center.

21.4.7 Other Considerations

There are a few more things to consider that are beyond the scope of this chapter. First, we have limited ourselves to discussing microprocessors and single servers. To some extent that is a reflection of the current market segments and how servers are sold today. However, some capabilities don't scale to this level and need to be considered at the rack level or even the data center level. Things like fans, 380 Vdc power distribution, and data center cooling alternatives should ideally be applied at a scale beyond an individual server. The Open Compute Project at Facebook is a good example of examining the problem in the whole. Other things, like how we *define* reliability, can move the set of trade-offs to a whole new efficiency level.

For example, although 380 Vdc (see Chapter 28) is primarily motivated in the data center by its reliability improvement, it can provide at least 7% better energy efficiency than new ac power distribution and 28% better in existing legacy data centers.

Likewise, the distributed nature of air movement by imbedding a large number of small fans in each server is actually less energy efficient than it needs to be. A blade chassis tries to scale this to larger physical dimensions and spread the fan energy cost over more server blades. However, a rack-level fan can be even more efficient than the current situation, especially if it is run at a higher voltage.

In addition, rack-level and data center-level alternative cooling solutions such as running the data center at 40°C (104°F) nominal, liquid cooling, or oil-immersion cooling can be more energy efficient or provide higher-density heat removal benefits, but they come with additional functional, reliability, and cost trade-offs.

Finally, reliability is traditionally defined at the HW level and separated from the software and total system availability. For example, the mega data centers like Google define reliability in terms of keeping a running version of the program going and, thus, redirect workloads around failing HW. Thus, they have no need for dual corded yet still get the reliability/availability they need with less loss in their power distribution. At first glance, replacing one server that has two redundant power supplies with two redundant servers that have two power supplies seems counterintuitive on energy. However, eliminating *all* the dual redundant power supplies in the data center results in a net energy improvement (about 5% by going from $2N$ at the PSU level to $N+1$ at the data center level).

21.5 SOFTWARE

The server platform requires software to be of any use. Having looked at the microprocessor and server hardware, it is appropriate to focus on the software. There are many levels of software and they can have an impact on the energy use of a data center. There are four levels of software that exist on a server platform: Firmware (FW), Operating System (OS), Middleware and Application Software. Together, all the software required to enable a specific functionality comprises its software stack.

At the lowest level of the software stack, FW is software that is embedded in the server platform and provides the fundamental, lowest-level interface between the hardware and the software. The Basic Input/Output System (BIOS) has been the traditional FW that ships with a motherboard to provide for all the low-level hardware interfaces and basic I/O functionality. BIOS is always platform specific and historically refers to the IBM-PC platform. The Unified Extensible Firmware Interface (UEFI) has emerged as a standard for what basic I/O features are available. It is implemented in a platform-independent way that transcends the historical limitations of the IBM-PC architecture. It is possible to boot a system to just UEFI without an OS being installed. BIOS and UEFI are typically written in Assembler and are designed to be fast and efficient because the basic I/O functions are invoked so frequently. The FW is provided as is and without alternatives, so there is virtually no opportunity to measure energy efficiency with respect to the FW, nor do anything about it as a purchaser.

The next level of software is the OS. Examples include Windows Server and Linux. In a virtualized server, there is a layer of software called a VMM that acts like the OS to the FW, and a much bigger machine to the OS (or several virtual machines to several OS instances). The OS can have a significant impact on data center energy; however, most OS selection is driven from the business need that requires a specific application, which in turn limits one to a specific OS. When a choice exists, the only way to tell which OS works best is to benchmark test cases on all alternative OS/application combinations and measure the energy consumption differences to do the same work.

Associated with the operating system and specific I/O peripherals are drivers. This is hardware- and OS-aware software that is written for the specific peripheral/OS combination. Again, benchmarking can be performed if there are alternative hardware/driver combinations or even hardware/driver/OS/application combinations. This is a significant amount of work and usually takes a backseat to just getting a capability in place to meet a business need. However, if properly planned and executed as part of a larger IT project, it *may* yield significant energy and cost savings during the planned deployment lifetime of an IT project. However, it is hard to know ahead of time if the savings exists, without prior knowledge or experience.

Middleware is the term used for software that provides services across a set of distributed systems. On a single machine, it would be considered a part or extension of the OS. Like the OS, middleware alternatives may not exist, nor be easy to correlate to the energy use of the entire software stack by alternative. Likewise, it is possible to yield energy and cost savings but hard to know without committing to a deliberate investigation.

Finally, the application software sits at the top of the stack. If you are developing the application yourself, it is a reasonable activity to instrument and measure the energy use of the application or at least provide tests during the acceptance that quantify the energy use. If you are procuring the application software, you can specify energy use as one of the attributes and use the acceptance testing as an opportunity to measure it. Of course, that begs the question of how you set the acceptance criteria. Setting the energy use requirements of application software is rarely done outside of embedded systems and real-time software. However, working with the developer or supplier, you should be able to quantify the energy cost associated with existing systems and use that as a baseline for setting acceptance requirements. Historically, job completion time has served as the proxy for energy. But actually measuring the energy use usually requires little more than a watt-hour meter (e.g., Kill A Watt, Watts Up, instrumented power strip/PDU) and a timer (perhaps supplied by the OS or embedded in the application) to put together a test case on a single server. In general, we are talking about effort that takes place before the software stack is deployed for day-to-day use. However, to be complete, one needs to do periodic audits throughout the operational deployment lifetime or even automate continuous monitoring. Software stacks can change over time, especially with updates, patches, consolidations, etc.

21.5.1 ACPI

The ACPI, introduced previously for C-States in the microprocessor section, actually encompasses the entire platform. ACPI is a software standard for energy management that defines an OS-managed power state model that is shown in the state transition diagram in Figure 21.10. It is controlled by the Operating System (OS directed). Not every state is necessarily implemented by a particular system, but those that are implemented are done in a way that exposes those features through the ACPI. The model encompasses Global states (G-States), Sleep states (S-states), Throttling states (T-States), CPU states (C-States), Performance states (P-States) in the platform, and Device states (D-States).

At the highest level are the Global States (G0–G3) and Legacy. After the OS boots, the system is in the G0, working state. The system can transition to a G1, sleeping state; a G2, soft-off state; or a G3, mechanical-off state.

Sleep states, or S-states, represent the standby states of the system when it does not have useful work to do. The S0, awake, state is part of the global G0 state. Sleep states (S1–S4) are gradations within the G1 state. S5 is part of G2, soft-off state. No program execution is performed in any of the S1–S5 states; they simply represent sleep states with successively longer latency resume times. S1 is used to suspend the system and occurs after the caches have been flushed to zero. S2 is usually not implemented in servers, although chipsets that are sold into all three market segments—server, workstation, and high-end desktop—may. S3 usually follows S1 by putting the DRAM memory into a self-refresh mode, allowing the microprocessor and/or chipset to go to a lower power state. S4, colloquially called hibernate, stores all system state on disk; it is rarely used in servers, as the wake-up latency approaches power-on times but skips the reliability, self-test steps. Finally, S5 is off and preserves no state. As far as the system is concerned, S5 makes no distinction between G2, Soft off, or G3, Mechanical Off. Wake-on-LAN (WOL) or any other hardware wake-on feature can provide for resuming a server from a sleep mode in G1 or G2 (but not G3). It really depends on what a particular system implements and which sleep modes have resume supported.

T-states and C-states have to do with the microprocessor. C-states have been covered earlier. T-states have to do with enforcing thermal limits on the microprocessor. If the system or the environment in which the server is running is unable to keep the microprocessor cool enough, the microprocessor will throttle its performance to reduce the amount of heat it is producing. Thus, unlike C-states that are about saving energy, T-states are really about protecting the microprocessor from self-heating leading to microprocessor failure. That it saves energy is a side effect of protecting the microprocessor that comes with a significant performance penalty. Thermal throttling should be avoided as it likely will require more overall energy to complete the computations when running in this mode.

P-states are for the platform and can involve the chipset and I/O in addition to the microprocessor. In general, more and deeper P-state support is desirable—at least to the extent that their use does not contribute to more overall energy to complete computations. Testing with a specific software stack on specific hardware is about the only sure way to determine this. Finally, for I/O devices like modems and disk

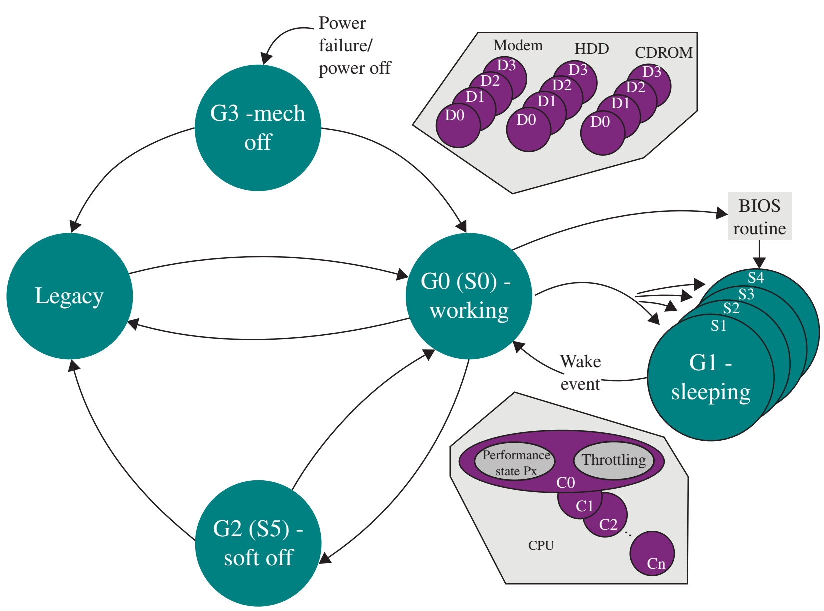


FIGURE 21.10 ACPI power states and relationship among platform states, including Global (G-states), Sleep (S-states), microprocessor Performance (P-states), CPU (C-States), and Device (D-States). Courtesy of ACPI/Intel Corporation.

drives, D-states allow for devices having power-saving modes that can be transitioned by the OS.

OS typically manage power state transitions on fairly lengthy time scales (at least for microprocessor HW) on the order of 50 ms. For microprocessors, chipsets, and other hardware in the server, there are opportunities to go into deeper sleep states and energy-saving modes in between the OS ACPI service intervals. One of the big opportunities is to consolidate activities into bursts of performance with longer periods of inactivity so that the HW can go into deeper energy-saving states yet provide the same level of performance. In the future, we are likely to see "tickless" OS that move away from OS-managed energy savings. The new approach is for a more performant, more energy-conserving model of OS-*directed*, HW-implemented energy management within the future server platform.

21.6 BENCHMARKS

Servers are used for many different purposes, and the nature of the software that is run on them, called workload, can have a major impact on the energy use. Those workloads can have a big impact on the productivity and energy use of a server. Benchmarking is a deliberate process of data collection to provide an early indication of how your data center will perform. The different techniques are a trade-off between the amount of work to accomplish and the ability to generalize the results to be predictive of your needs. Techniques include running on the actual SW stack, running a proxy workload, using the results of existing industry benchmarks, measured server power, and server power specs. Just make sure to keep the focus on the productivity and energy use.

Running the actual workload on the planned server hardware and the full software stack is the best way to benchmark a proposed installation for both productivity and energy use. Of course, this is likely the most work to accomplish. Validating a solutions stack (hardware and software) is the organizationally most mature approach.

Sometimes, planning has to proceed ahead of having the final software solution stack available to deploy and test. In this case, if you can identify a similar workload that you can run to characterize the solution stack, you can use it as a proxy workload. This is likely as much work as validating a complete solution stack but with the added benefit of getting predictive data ahead of the availability of the final solution stack.

Using industry benchmarks can be a research effort to gather data and make an approximate prediction of how a given solution stack should perform. The assumption, of course, is that you have a good enough idea of how you expect the application to perform. At least, you need to know enough to pick the appropriate benchmark to approximate your application. Keep in mind that most industry benchmarks cover either productivity or energy but rarely both. The real trick here is to select a benchmark that is close to the actual workload behavior and hardware that you expect to use. You can also run a benchmark on the target hardware ahead of having the target software application in hand. Of course, that is a more involved effort than just researching the posted benchmark results. Some industry benchmarks to consider include the following:

- SPECint_rate—is a microprocessor benchmark measuring the throughput of multicore processors on integer calculations. The results are applicable to character data manipulation workloads, as those are handled essentially as integer data.
- SPECfp_rate—is a microprocessor benchmark measuring the throughput of multicore processors on floating-point calculations. Historically, the benchmark is also a good indicator of the performance of a server's memory system.
- SPECweb—is a system benchmark measuring the productivity of a web server.
- SPECjbb—is a system benchmark measuring the productivity of server-side Java running as a three-tier client–server system. It highlights the server's microprocessor, caching and memory subsystem, and multiprocessing while deemphasizing disk and I/O.
- Linpack—is a system benchmark measuring the productivity of an HPC system.
- TPC—is a system benchmark measuring the productivity of a server on mission-critical, transaction-oriented workloads, like backend order processing. To get a competitive score requires a huge investment in a large system configuration, so smaller OEMs rarely list results with SPEC.
- Stream—is a system benchmark that characterizes the memory bandwidth of a system. It measures the productivity of memory-bound workloads and, by proxy, HPC workloads.
- GridMix—is a system benchmark that characterizes Big Data and Map/Reduce workloads.
- VMmark—is a system benchmark that characterizes the productivity of virtualized servers consolidated on a common server platform. It has three versions and can incorporate trade-offs among performance, server power, and storage power.

For the most part, you are only going to get productivity results from the benchmarks. Energy use will have to come from actual measurements or published data. The most useful information is to measure the energy directly while the server is running the application. Watt-hour meters to measure a single server can be had readily; products like the Kill

A Watt and Watts Up meters are frequently available at local home building supply stores for a quick test. Testing more accurately or testing more servers will require a real meter that is calibrated and rated for the expected power. Make sure that you get the appropriately rated unit, keeping in mind that most modern power supplies for a single server are rated over a wide range of voltages from 100 to 250 Vac. A step-up transformer or autotransformer can also be used to supply 110 Vac through a 110 Vac meter and transform the voltage to higher voltages for the server; remember to subtract the power loss of the transformer or measure at the appropriate location. Using the watt-hour function to measure energy over the duration of the application run will help you focus on energy. Remember you want to optimize total energy, first, rather than just power.

The following two tests deserve mention. Unlike the benchmarks mentioned earlier, they are not measures of productivity, but rather measurements of power at different performance levels:

1. SPECpower—is the first benchmark to assess both server system performance and power. It measures a web server workload running server-side Java (SPECjbb) at every 10% of server full load utilization.
2. SERT—not a benchmark, but rather an active mode rating tool. The proposed Standard Efficiency Rating Tool is being developed by SPEC for ENERGY STAR. While providing a power, performance, and inlet temperature assessment, it appears to ignore productivity and may exclude dc-input servers.

These are likely to become the first-level sort in the future for selecting servers. However, the real pitfall to avoid here is that they obscure energy and productivity by substituting power and performance, respectively. That may be sufficient to narrow your initial choices. But if you rely solely on these, you run the risk of erroneously assigning a higher value to a server that needs more energy to do the same amount of work.

Nameplate power rating has been used for a quick approximation. Realize that, in practice, a server is almost never run at its nameplate rating, so it will be a gross overestimate of the actual energy use (sometimes 2–3*x*). Again, multiply power by time to get energy. Better would be to find an OEM power configuration estimator on their website (the ENERGY STAR Power and Performance Data Sheet (PPDS) often lists a link). It allows one to make a much better estimate of the expected power draw of your specific server configuration knowing the installed options. Of course, this is somewhat of an overestimate because it assumes 100% server load and not the actual behavior with the target software stack installed.

Probably, one of the poorest specifications to use would be server idle power. This is a measurement of the server with a specific OS loaded but waiting for a program to be launched by the user. While the data may be easy to obtain, it is not even a real-world measurement for a useful server. The measurement that OEMs report usually is at provisioned idle with no live network connection. The simple act of putting a server on a network creates additional, not insignificant, power requirements associated with keeping the network connection alive. Moreover, this is the worst possible productivity case, as, by definition, no work is being performed; at idle, productivity is zero.

Finally, ENERGY STAR is a joint program of the U.S. Environmental Protection Agency and the U.S. Department of Energy to save money and protect the environment by recognizing products with superior energy efficiency. Servers have recently been added as a class of device that is eligible for ENERGY STAR recognition.

Servers get the ENERGY STAR label by meeting the requirements in place at the time of their certification test, usually at the product launch. The current requirements include a minimum power supply percent efficiency (in the 80s at 20, 50, and 100% of full load rating) and power factor. It also requires that the server power management features be turned on by default, as shipped. The current version allows the reporting of an active mode disclosure that includes idle and full load power measurements on a self-selected benchmark. Future versions will add a maximum idle power limit, tied to server type, PSU size, and installed options. They will also replace the active mode disclosure with the new SERT active mode benchmark. At that point, ENERGY STAR will become a rating that is relative to the other servers launched in the same year.

Thus, today, ENERGY STAR is merely a rating of the energy efficiency of a server's power supply and guarantee that it is shipped from the factory with power management features enabled. Future versions will add maximum idle power limits and active mode power and performance measurements on the SERT benchmark. An estimate of annual energy use, like other appliances have, is also to be expected, although without knowing the productivity, it will only be marginally useful. Unlike a refrigerator, TV, or clothes dryer, where you know what it does for the energy, servers have widely varying productive outputs. Thus, ENERGY STAR may be considered a necessary, but not sufficient, indicator that a particular server is a good choice for a green data center.

21.7 CONCLUSIONS

A green data center is all about getting the most compute productivity out of the least energy. Key to that is selecting an appropriate microprocessor and server with which to populate that data center. The selection involves a complex trade-off. Relying on the guiding principles to Organize, Modernize, and Optimize will help:

- Start with the end in mind.
- Use an energy Pareto.
- Focus on productivity.
- Measure and monitor the energy and productivity.
- Optimize energy first and, then power.
- Upgrade older, less efficient components.

The following general guidelines should help with respect to selecting a microprocessor for a green data center. The short answer is pick the microprocessor that makes your software stack run the best as measured by productivity and energy. Usually, this is something that is really a function of a server system, rather than a microprocessor. There are several attributes that are important to getting to the highest productivity at the lowest energy, but they never overrule actual results on your actual server hardware and application software stack. Small process node, high-K dielectric, low voltage, power-saving support, and a chip designed specifically for servers (multicore, multithreaded, 64 bit, ECC memory, virtualization support, and large cache) are probably the most important. On-die or on-chip VR is a feature to look for in the future. SPECint_rate for integer/character workloads and SPECfp_rate for HPC or floating-point workloads would be the two microprocessor benchmarks to consider. Notice that CISC, RISC, P_{max}, and TDP are all "it depends" tied to the system and workload.

The general guidelines for selecting a server system to populate your green data center are as follows. Avoid 1U form factors when air cooling; they are simply too short to accommodate efficient fans. Look for systems with advanced energy efficient fan algorithms. Use SSDs or Hybrid HDDs. With respect to memory, buy the DIMMs with higher capacity DRAMs and put fewer DIMMs in more memory channels. Use the Unregistered or Registered DIMM that gives you the lowest memory latency for your server configuration and DIMM count. Use low-power DIMMs when the SW applications stack (benchmarking) justifies it. Use on-board I/O that is fast enough for your application (upgrade the system if you must), rather than add-in card I/O. Avoid KVMs in favor of a networked console connection. Architect your data center power distribution and reliability strategy to provide the needed reliability while avoiding dual corded; consider 380 Vdc. Ensure the PSUs are greater than 80–90% efficient and have PMBus capability. Use SUE to plan your data center replacement policy. Validate and audit your servers' energy and performance over time. Remember this is a continuous improvement process, so Organize (measure), Modernize (upgrade), and Optimize (tune).

SPECpower and ENERGY STAR may be helpful in filtering the initial system choices. Other benchmarks that are representative of your workload can be used as a paper exercise in this filtering. But they are not substitutes for testing the actual software on the actual hardware and measuring energy (not just power) and performance. To the extent that you can replicate the application software stack and hardware during the selection process, the complex trade-offs can be simplified and more predictive. Benchmarking the actual application on the actual hardware (even if only at the scale of a few servers or one server) is the best way to understand the energy and performance implications. Mature IT organizations do this as a matter of course when they validate before deployment. All that may be new here is to add a focus on energy and performance as part of the application deployment, validation, and ongoing operation of your green data center.

In 2010, data centers (servers, cooling, and other electrical infrastructures) were responsible for approximately 1.3% of electrical energy consumption worldwide and approximately 2% in the United States. At the same time, the Smart 2020 report notes that ICT saves five times its carbon footprint by enabling energy efficiency in the rest of the economy. Your dedication to green data centers is laudable and gives us all confidence that armed with this advice you are going to make a significant difference for the other 98% of the economy.

FURTHER READING

80 Plus® Certified Power Supplies and Manufacturers. Available at http://www.plugloadsolutions.com/80PlusPowerSupplies.aspx. Accessed on May 22, 2014.

AlLee G, Tschudi W. Edison Redux: 380 Vdc Brings Reliability and Efficiency to Sustainable Data Centers. IEEE Power and Energy Magazine. Nov-Dec 2012; 10(6): 50–59.

Barroso LA, Hölze U. The data center as a computer: an introduction to the design of warehouse-scale machines. San Rafael: Morgan & Claypool; 2009.

Blum A. *Tubes: A Journey to the Center of the Internet*. New York: Ecco Press; 2012.

Brassard M, Ritter D, GOAL/QPC. *The Memory Jogger II: A Pocket Guide of Tools for Continuous Improvement and Effective Planning*. Methuen: GOAL/QPC; 1994.

European Union Code of Conduct for Data Centres. Available at http://iet.jrc.ec.europa.eu/energyefficiency/ict-codes-conduct/data-centres-energy-efficiency. Accessed on May 22, 2014.

Hennesey JL, Patterson DA. *Computer Architecture: A Quantitative Approach*. 5th ed. Waltham: Morgan Kaufman; 2012.

Kolinski J, Chary R, Henroid A, Press B. *Building the Power Efficient PC: A Developer's Guide to ACPI Power Management*. Hillsboro: Intel Press; 2001.

Lefurgy C, Rajamani K, Rawson F, Felter W, Kistler M, Keller T. Energy management for commercial servers. IEEE Computer, Dec 2003;36(12):39–48.

Meisner D, Gold B, Wenisch T. PowerNap: eliminating server idle power. ASPLOS'09; Washington, DC; March 2009.

Minas L, Ellison B. *Energy Efficiency for Information Technology*. Santa Clara: Intel Press; 2009.

Standard Performance Evaluation Corporation (SPEC). SPEC's Benchmarks. Available at http://www.spec.org/benchmarks.html. Accessed on May 22, 2014.

Top 500® Supercomputer Sites. The Linpack Benchmark. Available at http://www.top500.org/project/linpack/. Accessed on May 22, 2014.

US Environmental Protection Agency (EPA) ENERGY STAR. Enterprise Servers Specification Version 2.0. Available at https://www.energystar.gov/products/specs/node/142. Accessed on May 22, 2014.

Zimmer V, Rothman M, Marisetty S. *Beyond BIOS: Developing with Unified Extensible Firmware Interface*. 2nd ed. Hillsboro: Intel Press; 2011.

22

ENERGY EFFICIENCY REQUIREMENTS IN INFORMATION TECHNOLOGY EQUIPMENT DESIGN

Joe Prisco[1] and Jay Dietrich[2]

[1]*IBM Corporation, Rochester, MN, VT, USA*

[2]*IBM Corporation, Essex Junction, VT, USA*

22.1 INTRODUCTION

Energy efficiency in data centers is an important topic that is commonly discussed in the information technology (IT) and communications industry. Many of the recent data center surveys conducted by equipment manufacturers, academia, and consortia show that energy efficiency ranks high on the list of top priorities for data center operators and clients.

There are several forces driving improvement in data center energy efficiency. First and foremost, energy prices are increasing. Figure 22.1 shows data from the U.S. Energy Information Administration (EIA) [1]. The average retail price of electricity from 1960 to 2010 has grown at 0.17 nominal cents per year (not adjusted for inflation). With the demand for energy increasing to meet worldwide demand, costs are expected to continue to rise. The simplest way to reduce the utility bill and cost associated with it is to use less electricity.

Second, energy supply is increasingly at risk from short-term or long-term disruptions. These disruptions can be caused by things such as civil unrest, terrorism, politics, natural disasters (hurricanes, earthquakes), inadequate infrastructure, and accidents. Third, climate change is influenced by the emissions generated from energy use and the emissions of greenhouse gases, such as perfluorinated compounds (PFCs) and SF_6, in various manufacturing processes and systems operations. The greenhouse gas emissions trap heat in the atmosphere and cause increases in surface temperatures. Finally, governments are responding to all three of these issues by developing and implementing voluntary or regulatory-based programs for energy efficiency in a range of product types and system operations including information and communications technology (ICT) products and data centers. Governmental programs and policies such as the European Union (EU) Emissions Trading Directive, the EU Energy Efficiency Directive, the United States Environmental Protection Agency (USEPA) Mandatory Greenhouse Gas Reporting Rule, and the USEPA ENERGY STAR® program have established measures to reduce carbon emissions and reduce product and system energy use with the goal of reducing the quantity of Greenhouse Gases released into the atmosphere for each unit of GNP produced [2].

IT equipment manufacturers have traditionally focused design efforts on delivering equipment with greater computing, storage, and networking capability. Advancements in semiconductor and interconnect technologies have enabled manufacturers to significantly increase equipment performance per unit of power consumed with each generation of equipment [3]. In addition, further performance power improvements are delivered by reducing the energy consumption of various components through approaches such as Solid State Drives to replace or augment Hard Disk Drives, processors, memory, and I/O with low energy use states for periods when little or no workload is present, and the use of cache for storage systems. These innovations and system improvements have enabled the equipment to perform and manage ever more complex tasks and processes more quickly.

The demand for improvements in data center energy efficiency from data center operators, data center users, governments, and non-governmental organizations (NGOs)

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

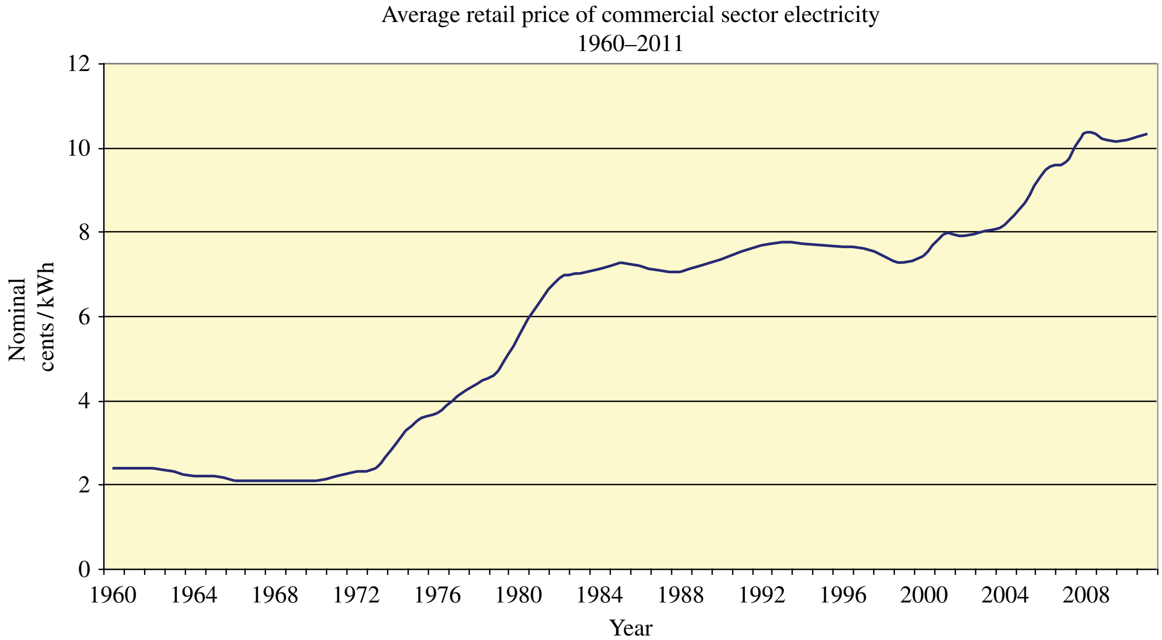


FIGURE 22.1 U.S. commercial sector electricity prices over time.

has accelerated the efforts of ICT product development teams to incorporate product features that improve the utilization of energy. The energy efficiency of IT equipment is assessed through three product capabilities: the amount of work that the equipment can deliver for each unit of energy supplied (performance/power profile), the ability to maximize the amount of work done (virtualizing workloads and maximize system utilization), and the intelligence to reduce power when workload is not present. Properly managed and balanced, these three capabilities can collectively contribute to the optimization of the IT equipment's workload delivery for a given energy use.

IT equipment that is executing a single application or workload is often only performing meaningful work for a small percentage of time—often 10% or less. The remainder of the time, the equipment sits idle, consuming 30–90% of the power and generating 30–90% of the heat that occurs when the equipment is fully loaded. In order to improve system utilization and make better use of each piece of IT equipment, IT equipment companies have developed and deployed server and storage virtualization technologies. These technologies enable the equipment to partition its resources and concurrently support multiple workloads. This increases system utilization to 20–60% or beyond and allows a single server or storage device to do the work that previously had to be managed by multiple systems. In aggregate, the energy consumed and space required by the single server running multiple applications is typically in the range of 20–80% less than a group of servers running individual applications, depending on the application and the capabilities of the original group of servers.

Even virtualized Windows and UNIX servers are idle approximately 40–80% of the time. When minimal workloads are present, it is appropriate to idle all or parts of system components to minimize energy use while maintaining the equipment in a ready state. Processor, memory, and network technologies have been developed and deployed to allow systems to reduce their power use while maintaining specified performance levels or response times when the quantity of workload is reduced. These power management capabilities typically require software enablement, but when deployed, can reduce overall data center energy use by 10–20%.

A consensus has emerged in the ICT industry and the public arena regarding the need to improve data center energy efficiency and by extension, the efficiency of ICT equipment, which has resulted in the development or proposal of voluntary programs and regulations. One of the earliest laws was passed in Japan: the Law Concerning the Rational Use of Energy (Japan Energy Law (JEL)). Established in 1994 as a result of Japan's commitment to the Kyoto Protocol, JEL established a power per performance metric for servers and watts per GB metric for storage. Each type of product has several categories that are established based on the product architecture. Each category has a weighted average target that must be met by individual products marketed in Japan. The target has been made progressively more stringent through periodic updates. The performance measurements for the best performer in a given category for a given period (the top runner) are used as the new target. While JEL measures one aspect of ICT equipment efficiency, the workload delivered per unit of energy consumed, it does not take into account power management capabilities or the ability to virtualize and achieve higher system utilizations.

In 2006, the U.S. EPA and Department of Energy (DOE) announced a partnership to improve the energy efficiency of servers and data centers. As part of this partnership, the EPA began collecting data on the energy use of data centers under the ENERGY STAR building program. The EPA published ENERGY STAR Version 1.0 Program Requirements for Computer Servers in 2009 and for Uninterruptable Power Systems (UPSs) in 2012. In 2013, EPA implemented Version 1 requirements for Storage systems and Version 2 for computer servers, as well as investigating requirements for Large Network Equipment. There has also been significant interest in this area by legislatures and governments around the world, including the EU, under the Energy-Related Products Directive, Korea, and China.

The remainder of this chapter will provide detail on the technical requirements and the evolution of the ENERGY STAR programs, as well as their impact and influence on the development of regulatory programs. The future landscape of programs and laws and their attendant requirements for workload and system utilization metrics will also be considered.

22.2 COMPUTER SERVERS

In December of 2006, the USEPA announced that they were initiating a requirements development process for Computer Servers, launching a 3-year effort to develop and publish ENERGY STAR requirements for Computer Servers. Because of the complexity of server systems, driven by the wide range in number and type of component configurations that could be created within a single model type, establishing requirements was a challenge for the ENERGY STAR program. The only product of comparable complexity for which the program had previously undertaken was computer workstations, which had significantly fewer configuration permutations in a single model and a very different power profile from a server. Nonetheless, development of a computer workstation had taken several years, foreshadowing the challenges of establishing a computer server specification.

After releasing the Specification Framework Document and holding extensive stakeholder discussions, EPA evaluated how server systems should be categorized and what energy efficiency criteria should be established. Based on the analysis of the available data, it was determined that servers were best categorized by the number of processor sockets. It was also determined that manufacturers could qualify either individual product configurations or product family. A product family was defined as a range of configurations within a given product model or machine type. Under Version 1, the product family enables a manufacturer to provide the required data on the Power Performance Data Sheet for three representative machine types or models, based on processor socket power and number of cores, to qualify all the configurations for that machine type. The manufacturer is required to certify that all configurations covered by the qualification will meet the relevant requirements. Under Version 2, the product family definition has been broadened, requiring manufacturers to provide the power use, performance, and configuration data for five defined product configurations, thus allowing the qualification of a range of processor socket power and core count within the product family. This simplifies and reduces the testing regime while providing power use and efficiency data on the full range of configurations for the product family.

EPA identified the key server capabilities that should be considered in creating criteria to differentiate energy-efficient servers: power supply efficiency, idle power, the workload capability of the server, server utilization or virtualization capability, a performance power metric, and the ability of the server to report power use and server inlet air temperature to the network. These server capabilities were explored through a set of data requests and four drafts of the ENERGY STAR Computer Server requirements before a final specification document was released in May of 2009. Each of the aforementioned attributes was analyzed in some detail before the final requirements were published.

22.2.1 Power Supply Efficiency

Power supply efficiency was generally recognized by involved stakeholders as a relevant criterion for ENERGY STAR. Losses in the power supply reduced the energy available for useful work in the server. EPA adopted the ECOVA Plug Load Solutions 80 Plus Power Supply certification program (Table 22.1)[1] to set power supply efficiency and power factor requirements. A data collection effort on power supply efficiency and power factor levels for server power supplies currently in use on the market identified that setting the power supply requirements at Silver level would drive substantial improvements in power supply efficiency of the server fleet. The 80 Plus program also had the benefit of providing an established testing procedure and certification process to simplify the execution of the power supply efficiency requirements for the ENERGY STAR requirements.

Another issue became clear as the EPA collected server energy use data. Redundant, dual power supplies, combined with the wide range of power use over the range of component combinations for a given server machine types, along with servers that typically idle 80–90% of the time, often resulted in power supplies operating at or near the 10% utilization point with accompanying low efficiencies.

As a result of this finding, EPA added an efficiency requirement for the 10% load point to drive improvements in

[1] http://www.plugloadsolutions.com/80PlusPowerSupplies.aspx

TABLE 22.1 Power Supply Efficiency Levels for 80 Plus Certification[a]

80 Plus certification	115 V internal non-redundant				230 V internal redundant			
% of rated load	10%	20%	50%	100%	10%	20%	50%	100%
80 Plus	—	80%	80%	80%	—	—	—	—
80 Plus bronze	—	82%	85%	82%	—	81%	85%	81%
80 Plus silver	—	85%	88%	85%	—	85%	89%	85%
80 Plus gold	—	87%	90%	87%	—	88%	92%	88%
80 Plus platinum	—	90%	92%	89%	—	90%	94%	91%
80 Plus titanium	—	—	—	—	90%	94%	96%	91%

[a] Courtesy of ECOVA.

power utilization during periods of no or low workload. This also highlighted to manufacturers the utility of either offering power supplies with two or three capacities for each server model to enable customers to select a power supply capacity that matched the power needs of their chosen configuration or to develop innovative ways to enable one power supply to carry the full power load while idling the redundant supply and pushing the power supply utilization point for the enabled supply into higher efficiency zones at idle and low workloads. These approaches, combined with the minimum power supply efficiency requirements, have combined to improve server power utilization in the data center. Version 2 of the requirements has increased the minimum power supply efficiency level to 80 Plus Gold.

22.2.2 Idle Power

The server utilization data, with servers sitting idle for significant periods of time, spurred an interest in setting an idle power criterion for server systems. At the time, processor systems had power management functionality that could enter low power modes if no workload was present or that could adjust the voltage and frequency of the processor or individual cores to correspond to the level of workload present in the server. Figure 22.2 shows the different power management modes and their effect on processor frequency (and by association processor power use) for an IBM Power™ processor. The Power7 processor offers four power management modes, each with its own specific power profile. Static Power Saver (SPS) reduces power use, but also can impact system performance, Dynamic Power Saver–Favor Performance (DPS-FP) enhances performance when workload is present and reduces frequency when the processor is idle, and the DPS (power) varies frequency and voltage to deliver power proportional workload processing.

The nominal mode has a consistent frequency and power use no matter how much workload is present. The DPS-FP and DPS power management modes as well as similar power management modes on ×86-based processors can reduce power at idle by up to 60% or more when compared to the power use at the maximum workload. Power management capabilities are also available for memory systems and I/O.

While the server power use data collected by EPA suggested that an idle criterion made sense for server systems, implementation was complicated by the increasing complexity and range of configurations as systems expanded from one processor socket to four processor sockets. The data analysis in Chart 1 of the document “Idle Data Analysis and Charts for the Draft 3 ENERGY STAR Computer Server Specification”[2] showed reasonable distributions for one- and two-socket systems, but large idle and maximum power increases for four-socket systems as these became more complex. The power range for four-socket systems is illustrated by the range of maximum power (*x*-axis) for system catalogued in Figure 22.3, which shows the percentage of idle power to the full power for the maximum configuration of different machine types qualified to ENERGY STAR through August 2011.

Based on the available data, EPA set an idle criterion for one and two processor socket systems and required a qualified server system to ship with power management enabled when testing the idle power used. In recognition of the complexity of the four processor systems, EPA required that qualified systems ship with power management enabled, but did not set an idle criterion for four-socket systems. As illustrated by Figure 22.3 for four-socket servers, the ratio of idle power to maximum power for ENERGY STAR qualified systems is consistently above 50%, enabling server systems with power management enabled to reduce ongoing server power use by 30% over the life of the server, assuming that a server is idle 40–80% of its operating time. Similar results can be demonstrated for one or two processor socket systems.

22.2.3 Workload Capacity of the Server

Through the power of Moore’s Law, which states that the number of transistors on a chip will double approximately every 2 years, a server system’s ability to deliver more workload for each watt of power delivered increases 20–100% with

[2] http://www.energystar.gov/ia/partners/prod_development/new_specs/downloads/servers/Draft3_Idle_Data_Analysis_Charts.pdf?c4eb-c336

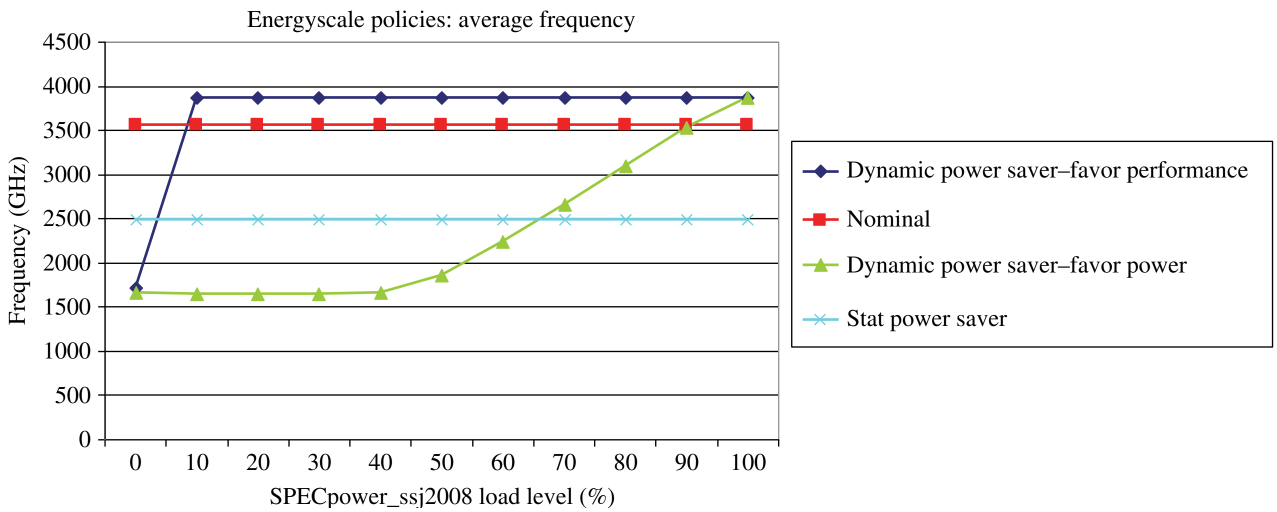


FIGURE 22.2 Processor power management function. Courtesy of IBM.

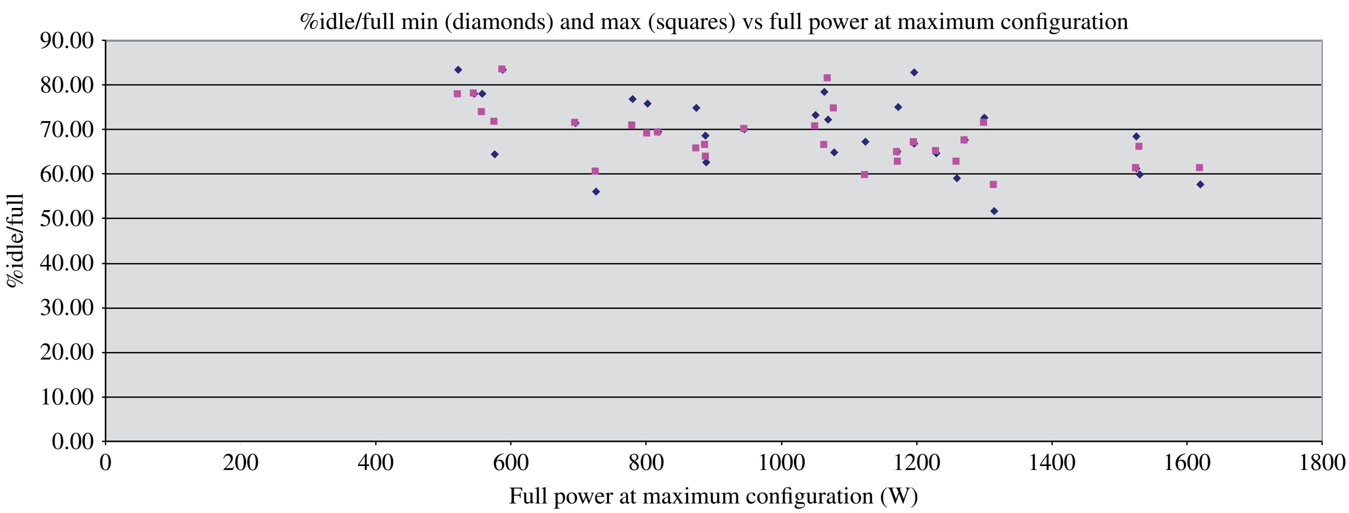


FIGURE 22.3 Idle power vs. full power for maximum configurations of Four processor socket server models. Courtesy of IBM.

each new generation of server systems. While the workload capacity is indicative of the capability to deliver workload, it is not directly indicative of the overall efficiency of the server. Movement of a single workload that utilizes an existing server 15% of the time to a next-generation server is likely to result in faster execution of the workload, but lower utilization of the new server, which does not result in any improvement in the workload delivered for each unit of energy consumed. Because of this limitation, workload capacity was not chosen as a criterion for the ENERGY STAR requirements.

22.2.4 Server Utilization, Virtualization Capability, and the Performance/Power Metric

The efficiency of a server as measured by the criterion of maximizing the workload delivered per units of energy consumed is dependent on its capacity to do work, its ability to perform multiple workloads at the same time (unless the workload is very large, requiring the full server to execute the application), and its ability to reduce power use when no workload is present. EPA, in consultation with its stakeholders, explored available metrics to assess a server's virtualization capability and its performance power characteristics.

22.2.4.1 Workload Virtualization The majority of servers have the capability to be virtualized. The extent to which they can virtualize depends on the processor, server infrastructure, and the operating system or hypervisor capabilities. The more virtualized a system becomes, the more dependent it becomes on its ability to access and control system storage, memory, and I/O. Unfortunately, there is not currently an effective, recognized metric to assess the ability of a server to virtualize and drive higher levels of utilization.

Establishing a virtualization metric for the ENERGY STAR requirements was dismissed early in the development process.

22.2.4.2 Performance/Power Benchmarks At the beginning of the ENERGY STAR requirements development process, a single performance/power metric was available: SPECPower_ssj2008.[3] While this metric provided a means to assess the performance and power characteristics of a server system from full power to idle, its effective range of coverage was primarily through two-socket systems with 8–16 GB of memory. While more heavily configured systems with four or more processor sockets can be assessed with SPECPower_ssj2008, the results have limited relevance to these larger systems. As a result, EPA chose to defer implementation of a performance/power metric for Version 1 of the Server requirements and work with industry stakeholders and Standard Performance Evaluation Corporation (SPEC) in the development of the Server Efficiency Rating Tool (SERT).[4]

The SERT metric is intended to measure and evaluate the energy efficiency of servers. Rather than focus on server performance under a specific type of workload, it tests the performance/power characteristics of the key components of the server system: processor, memory, storage, and I/O. The overall performance of a server system is a combination of the performance of the individual components. The SERT metric consists of eight worklets designed to stress and evaluate how each component affects the performance/power profile of a server system. SERT differs from a typical benchmark in that it is intended to be workload and operating system agnostic, easy to set up and use, and assess the server on its default, out-of-the-box settings rather than requiring special system tuning. The SERT design document[5] discusses the details of the worklets and the overall metric tool. The first released version of SERT reports the individual worklet scores and the power and performance measurements at the test intervals for each worklet. Currently, the intent is to report the individual worklet scores in the first production release of SERT. The development of composite rating for a server system will necessitate collecting and analyzing representative metric data sets for one, two, and four processor socket systems.

EPA chose to use Version 2 of the Computer Server Requirements to collect SERT metric data for qualifying server systems, with the intent to collect sufficient data to establish performance/power criteria for one to four-socket rack and tower servers as well as blade servers in Version 3.

[3] http://www.spec.org/power_ssj2008/
[4] http://www.spec.org/sert/
[5] http://www.spec.org/sert/docs/SERT-Design_Doc.pdf

A full assessment of the SERT metric to determine the best way to use the metric will take time. While some stakeholders may argue that progress toward more efficient server systems metric has been too slow, the current drive by stakeholders to increase the efficiency of server systems also takes time. The overall experience of the data center industry over the past 5 years would suggest that improvements in server system efficiency have shown the following results:

1. The recognition by data center operators and server manufacturers that virtualization is key to driving increased utilization of increasingly expensive server hardware. The cost of data center space and IT hardware is driving innovation that is and has the potential to dramatically improve the workload delivered for each unit of energy consumed and reduce the quantity of hardware necessary to complete a given set of activities.
2. Design cycles for server systems range from 12 to 18 months for volume servers and 18–30 months for resilient servers. Because of these design cycle times, efficiency improvements will be incremental and take time.
3. The power supply efficiency requirements have driven improvements in server power supply efficiency such that Version 2 of the Computer Server Requirements require a minimum 80 Plus Gold-level power supply.
4. The idle criteria for one and two socket servers, the requirement to ship all ENERGY STAR qualified servers with power management enabled, and the intent of both ENERGY STAR and various governmental bodies to establish performance/power criteria for servers have focused both system and component manufacturers on improving the power management and the performance/power profile of servers.

Assessing the SERT metric through evaluation of a growing data set of metric results across the range of server configurations will reinforce industry efforts to improve server efficiency.

22.2.5 Reporting of Server System Power Use and Inlet Temperature

The Computer Server Requirements require that server systems collect and report server power use and inlet temperature so that data center operators have the ability to collect and evaluate power and thermal information. While this capability has become a standard function for server systems and storage systems, which offers potentially important information to the data center operator, there is still significant work to complete with respect to

TABLE 22.2 Emerald™ Storage System Categories[8,a]

Category	Online	Near online	Removable media library	Virtual media library
Level	Online 1	Near online 1	Removable 1	Virtual 1
Consumer/Component	Online 2	Near online 2	Removable 2	Virtual 2
Low end	Online 3	Near online 3	Removable 3	Virtual 3
Mid-range	Online 4			
High end	Online 5	Near online 5	Removable 5	Virtual 5
Mainframe	Online 6	Near online 6	Removable 6	Virtual 6

[a]Courtesy of SNIA.

collecting and presenting this vast amount of data in a meaningful way.

Version 2 of the Computer Server Requirements was released in March 2013, with an effective date of December 2013. Servers qualified to the Requirements will offer data center operators improved power supply efficiency, increased power management functionality in processors, memory, and I/O, and public information on the performance/power profile of the full range of configurations available for a given machine type or mode. Integration of more efficient server systems into data centers that manage energy in a systematic way will enable improvement in the workload delivered for each unit of energy consumed in data centers.

22.3 STORAGE SYSTEMS

EPA released the Storage Specification Framework Document in June of 2009, shortly after publishing Version 1 of the Computer Server Requirements.[6] Storage systems, with their broad selection of media types, the range of media types/configurations that can be offered with each machine type, and dependency on data placement as well as software functionality to improve performance and system utilization, represented an even greater challenge than servers in establishing ENERGY STAR requirements. EPA collaborated with the Storage Networking Industry Association (SNIA) Green Storage Initiative[7] to adopt storage system categories and a metric test procedure. The requirements process took over 3 years and Version 1 was finalized in August 2013.

In order to define product categories for storage systems, EPA adopted the taxonomy from the SNIA Emerald™ Power Efficiency Measurement Specification.[8] Version 1 of the Storage System Requirements covers On-Line Categories 2 through 4, though additional categories may be specified in future versions of the Requirements (Table 22.2).

[6] http://www.energystar.gov/ia/partners/prod_development/new_specs/downloads/storage/ES_Storage_Framework.pdf?420f-b5ea
[7] http://www.snia.org/forums/green
[8] http://snia.org/sites/default/files/EmeraldMeasurementV1_0.pdf; p. 18.

EPA has determined that there are four key criteria for assessing storage system efficiency: power supply efficiency, performance/power metrics, capacity optimization methods (COMs), and reporting of storage system power use. Idle power was determined to be of limited value in assessing system energy efficiency as 70–90% of the power use of storage systems resulted from the high-density drives (HDDs), continuously spinning disks for which currently available technologies have limited, if any, power management capability. COMs were identified as providing improved energy efficiency for the data center, as they can improve system utilization and reduce the number of storage media required to execute a given workload.

22.3.1 Power Supply Efficiency

Like servers, improvements in storage system power supply efficiency increases the percentage of the line feed power used to do work. After completing a data collection exercise, EPA determined that Version 1 of the Storage System Requirements should require the 80 Plus Silver level for storage systems. Because storage systems have a more consistent power profile due to consistent activity on the controller and the HDDs, a storage system with redundant power supplies will operate at or above the 20% load point, reducing the concern with low power supply loadings in the idle mode that was identified for servers.

22.3.2 Performance/Power Metric

Because the HDDs provide a continuous power load for storage systems and the workload delivered per unit of energy consumed is highly dependent on the software, data management algorithms and available cache on the controller, EPA determined that the Version 1 Storage Requirements should implement a performance/power metric rather than an idle criterion. EPA considered available test protocols and benchmarks and chose the SNIA Emerald Power Efficiency Measurement Specification (Emerald metric) as the performance/power metric for the storage category. The Emerald metric has five workloads, hot band, random read/write and sequential read/write, and an idle measurement. The inclusion of the hot band workload is important, as data placement software functionality is becoming a key storage system capability and an expected technology direction for most storage systems over the life of Version 1. The SNIA Emerald test will provide EPA and industry stakeholders a range of data to better understand the performance/power profiles of storage systems and set meaningful performance/power criteria for Version 2 of the Requirements.

The ENERGY STAR requirements identify three types of operations for metric testing: Transaction, Streaming, and Capacity. The Requirements specify a subset of the Emerald workloads that shall be reported for each type of operation. Because of the recent release of the Emerald Metric and the lack of available test data, EPA is using Version 1 of the Storage Requirements to secure sufficient data to assess the metric results and identify the best approach to creating a single metric or set of metrics for the system qualification under Version 2 of the Storage Requirements.

The other key to the success of the storage requirements is the development of a workable product family definition. Like a server system, a given model or machine type within an On-line category will have hundreds or thousands of possible combinations of storage devices. EPA has established the framework of the product family in Version 1 of the Storage Requirements, defining three primary configurations: the Optimal Performance/Power Configuration (OPPC), which optimizes the performance/power metric for a given operation type, and Minimum and Maximum Performance/Power Configurations, defined as the configurations with a storage media count some percentage below and some percentage above the OPPC, respectively. Storage system configurations between the Minimum and Maximum configurations for a given type of operation can be qualified to the ENERGY STAR Requirements.

Manufacturers will also be allowed to report metrics for an Extended Minimum Configuration (EMC) with a lower storage media count than the Minimum Configuration. If the performance/power metrics for the EMC are within a specified percentage of the metrics for the OPPC, the storage system can be qualified to ENERGY STAR down to the EMC. Companies are allowed to identify replacement storage devices, which are comparable to the storage devices used in the qualification testing. Qualification of a replacement storage device will be accomplished by validating that specified device parameters are within defined boundaries when compared to the parameters of the storage device used to qualify the storage system.

In recognition of the fact that most customers purchase storage systems with a mix of drive types, EPA has made accommodations to enable a company to offer ENERGY STAR qualified configurations made up of a mix of storage devices. If a company chooses to qualify a model or machine type to two or three operation types, transaction, streaming, and/or capacity, the Requirements establish a methodology by which storage devices can be combined from the two or three qualified configuration groups to create a group of qualified configurations with multiple drive types. The drive types can be a mix of the drives used in the system qualification tests and qualified replacement drives. The requirements have also made provisions for testing a mixed-drive system. This enables manufacturers to broaden the storage media offerings while minimizing the number of configurations that have to be tested for qualification. It will be important to consult the final, published ENERGY STAR Storage System requirements to get the specific requirements for the testing and reporting results for a storage product family.

22.3.3 Capacity Optimization Methods

Storage system providers have developed a variety of software-based techniques to improve capacity; examples include data de-duplication, data compression, thin provisioning, and delta snapshots. These functions enable servers to store a given amount of data on a smaller number of storage devices. While they typically do not directly contribute to the energy efficiency of a device, and may increase the energy use of an installed system, they can decrease the number of storage devices needed and the energy and cooling needs of those extra devices in the data center. In recognition of the benefits that COMs bring to reducing data center energy use, EPA is requiring that an ENERGY STAR qualified storage system have a minimum number of COMs available on a qualified system.

22.3.4 Reporting of Storage System Power Use and Inlet Temperature

The Storage Requirements state that storage systems collect and report server power use. Reporting of inlet temperature is optional under Version 1 as storage systems have historically not collected inlet temperature data due to the fact that a system has many storage devices and inlet temperature points. EPA has indicated that reporting of inlet temperatures will be required under Version 2.

22.4 UNINTERRUPTABLE POWER SYSTEMS

In May of 2012, EPA released Version 1 ENERGY STAR Program Requirements for Uninterruptable Power Supplies (UPSs).[9] The requirements identified four classes of UPS

[9] http://www.energystar.gov/index.cfm?c=uninterruptible_power_supplies.pr_crit_uninterruptible_power_supplies

products covered by the requirements, one of which is Data Center UPSs intended to protect large installations of ICT equipment such as enterprise servers, networking equipment, and large storage arrays. The devices covered by the requirements include Static and Rotary UPSs with one of two output types: AC-output and DC-output UPS/rectifiers. The UPS Requirements set minimum average efficiency and power factor requirements for AC-output UPSs and DC-output UPS/rectifiers. A UPS system with capacities larger than 10,000 W, which includes metering and communication capability, receives a 1% efficiency incentive (reduced minimum efficiency) to encourage inclusion of the capability to report power use to a networked power monitoring system.

22.5 NETWORKING EQUIPMENT

In October 2012, EPA announced its intent to develop ENERGY STAR requirements for large networking equipment. In the discussion document, EPA asserted that it would be possible to reduce the energy use of network equipment by 20–50% through the adoption of efficient technologies. Given the experience with Server and Storage systems, it is likely that the requirements development process will take over years to complete.

The ENERGY STAR program has published or is developing requirements for the key components of the individual pieces of the ICT infrastructure in the data center. The requirements are intended to recognize manufacturers whose products can deliver the highest quantity of workload per unit of energy consumed through continued improvements in performance, power management, and the functionality to increase hardware utilization. The initial work on the ENERGY STAR requirements for data center IT equipment has focused on defining product families and relevant performance/power metrics as well as establishing basic requirements for product energy use characteristics: power supply efficiency criteria, idle power requirements for one- and two-socket servers, the enablement of power management capabilities on shipped products, and the collection of performance/power data to inform the development of criteria in subsequent versions of the requirements.

Currently available enterprise ICT equipment offer data center operators significant functionality to reduce data center energy use if the functionality is enabled. Implementation of power management on a server that has utilization of 30% can reduce power use by 20–40% depending on the extent of idle power savings. Utilizing virtualization capabilities on server and storage systems can enable a single machine to do the work of 6–10 current or previous generation technology machines. Improvement in ICT equipment energy efficiency and functionality to manage more workload will continue as manufacturers innovate to deliver capabilities that both improve the performance and the efficiency of the ICT equipment.

22.6 FUTURE TRENDS IN PRODUCT ENERGY EFFICIENCY REQUIREMENTS

Data center energy efficiency, as measured by the amount of work the data center delivers for each unit of energy it consumes, is influenced by all the data center systems: the IT hardware, management of IT workload placement, the facilities hardware, the management of the data center thermal profile, and management systems, which integrate some or all of these activities. As the building blocks of the data center, IT and facilities equipment will continue to be a focus of voluntary and regulatory energy efficiency programs around the globe. ENERGY STAR, in collaboration with industry stakeholders, has begun the effort to define relevant metrics to assess the energy efficiency of data center equipment. These initial efforts have made it clear that establishing energy efficiency criteria for these complex systems is a difficult undertaking. The range of system configurations, functionality, and types of unique, distinct workloads supported requires a flexible approach to assessing the energy efficiency of enterprise-level ICT equipment. The pace of technological change and innovation in the industry introduces further complexity, as standards and metrics established today may be rendered obsolete or marginalized by technology changes. Finally, the lack of general metrics, as opposed to workload specific metrics, to assess the performance/power profile of the equipment and the lack of measured system data for those metrics that do exist require a measured, incremental approach to the development of workable ICT equipment energy efficiency standards. While progress has been made, much work remains to develop standards and metrics that provide a meaningful assessment of the energy efficiency of ICT equipment.

The initial steps to manage ICT equipment efficiency have focused on simple, easily measured requirements: power supply efficiency, enablement of power management functionality, and the measurement and reporting of the maximum and idle power use of a range of configurations of a given machine type or model. Some or all of these requirement types have been implemented in various jurisdictions around the globe: Japan (the Japan Energy Law), Mexico (Energy Use Reporting Requirements), the European Union (Power Supply Efficiencies), and the United States (ENERGY STAR requirements). These requirements have created the first tier of efficiency requirements and begun the process of more complete reporting of power use and performance/power metrics across the range of system configurations available in the market.

The next step, which is expected to unfold over the next several years, is the implementation of performance/power requirements and grading systems in various jurisdictions. The ENERGY STAR program has established testing and metrics protocols to serve as a starting point for energy efficiency requirements and has established a clear path to collect performance/power data for computer servers and storage systems. The collected data will be used to establish performance/power criteria, with a top runner or leadership focus, for these product types over the next 2–5 years.

Several government entities have also declared their intent to establish performance/power regulatory requirements in this same time frame, including, but not limited to, California, China, European Union, and Korea. These entities have launched or are completing studies or regulatory development efforts. China and Korea are most advanced in these efforts. In China, the Ministry of Environmental Protection released Technical Requirements for Environmental Labeling of Products—Server,[10] which covers one to four processor socket rack and tower computer servers, blade servers, and storage servers. The program went into effect in April of 2011 and is voluntary, but will be used to inform government procurement decisions. The server requirements draw heavily from the ENERGY STAR requirements. The storage requirements include power supply efficiency, a watts/IOPs criteria, and a criteria for idle to maximum power draw. Korea has indicated its intent to establish a computer server energy efficiency grading system in 2014/2015 for one and two processor socket servers using some limited combination of SERT worklets. The European Union and the California Energy Commission (CEC) have indicated their intent to initiate a study on Computer Server energy efficiency requirements in 2013 and 2015 respectively with the intent of establishing requirements in 2015 or 2016. Ideally, these programs, and others like them in other jurisdictions, will build on the testing protocols, metrics, and measured data collected through the ENERGY STAR program. They will help standardize the testing as well as data generation and collection processes, while enabling individual governments to set metric criteria appropriate to the conditions and interests of their jurisdiction.

[10] http://english.mep.gov.cn/standards_reports/standards/others1/Technical_Requirement_Labelling/201103/t20110331_208223.htm

REFERENCES

[1] *Annual Energy Review 2011*. Washington, DC: Bernan Association, U.S. Energy Information Administration; 2012.

[2] Chennells J. Trading in carbon emissions—how to ensure compliance. Energy World 2005;330:10–11.

[3] Koomey J. *Growth in Data Center Energy Use 2005–2010*. Oakland: Analytics Press; 2011. Available at http://www.analyticspress.com/datacenters.html. Accessed on May 23, 2014.

FURTHER READING

Data Center Dynamics Data Center Efficiency. Available at http://www.datacenterdynamics.com/focus/themes/energy-efficiency. Accessed on May 23, 2014.

European Union Data Center Code of Conduct. Available at http://iet.jrc.ec.europa.eu/energyefficiency/ict-codes-conduct/data-centres-energy-efficiency. Accessed on May 23, 2014.

Green Grid Library of Resources and Tools. Available at http://www.thegreengrid.org/library-and-tools.aspx. Accessed on May 23, 2014. Offers white papers and other resources on data center energy efficiency topics.

Lawrence Berkeley National Lab (LBNL) High-Performance Buildings for the High-Tech Industry: Data Centers. Available at http://hightech.lbl.gov/datacenters. Accessed on May 23, 2014. Offers white papers and resources for assessing and improving data center efficiency.

Open Compute Project. Available at http://www.opencompute.org/. Accessed on May 23, 2014.

USEPA Energy Star. Top 12 ways to decrease the energy consumption of your data center. Available at http://www.energystar.gov/index.cfm?c=power_mgt.datacenter_efficiency. Accessed on May 23, 2014.

23

RAISED FLOOR VERSUS OVERHEAD COOLING IN DATA CENTERS

Vali Sorell

Syska Hennessy Group, Charlotte, NC, USA

23.1 INTRODUCTION

There have been two schools of thought on how to deliver air to uncontained data halls. (There are other methods too, but the two discussed here are the main ones.) On the one hand, data centers have historically been provided with upflow air distribution using perforated tiles in a raised floor environment. On the other hand, telecommunication (telecom) central offices have historically been built on concrete floors using overhead air distribution with downflow air distribution into aisles. Both methodologies have served their respective industries well for many years.

In the past several years, there has been a convergence of functions under which the differences between telecom spaces and data hall spaces have almost disappeared. The reasons for this convergence relate more to the functions of the IT equipment itself, and are beyond the scope of this chapter.

Members of these two schools have, in general, continued their accustomed practices, following the belief that their respective practice is the better means to provide cooling airflow to the IT equipment. Designers who have come to the industry after the convergence of the two types of equipment find themselves in the middle and ask the question: "Overhead or under-floor?" There is not a definitive or correct answer—both methodologies must be understood since there are advantages and disadvantages for each. These issues will be addressed more fully in this chapter.

23.2 HISTORY OF RAISED FLOOR VERSUS OVERHEAD AIR DISTRIBUTION

Once upon a time in the not-too-distant past, a mainframe computer was the size of the room that housed it. Cooling of these older larger computers was done in a more conventional manner using water circulated to heat exchangers. The heat exchangers were in turn connected at the computer and were able to directly cool processors, power supplies, and frames. Water was the primary liquid circulated from the central plant, where it was cooled to the required temperature, to the heat exchanger. As a convenience, a raised access floor was used to keep the circulating cooling liquid out of sight. Still, the piping accessories, such as valves and pumps, were within easy and convenient reach by simply lifting a tile from the raised access floor.

The predominant computer technology of the 1970s and 1980s was the relatively high-power bipolar semiconductor chips, which were liquid-cooled. The development of complementary metal oxide semiconductor (CMOS) changed the development of computers. These chips are not only lower power than the older bipolar types, but they are also air-cooled. It became a given that in the 1990s, computers shrank in size and became predominantly air-cooled [1]. Despite using less energy for a given number of compute cycles, the price of reduction in size is that more power was then placed inside the computer boxes. The net effect is that for every reduction in size, there is a greater corresponding increase in compute power as well as electrical power [2].

As the data center industry began to adjust to the air-cooled technology, it made good sense to convert the raised access floors into a means for delivering cooling air to the computers. Computer room air conditioning (CRAC) or computer room air handling (CRAH) units became a convenient way to provide cooling in small, modular units. Since many of these CRAC/CRAH units were being brought into spaces that were already in use prior to the advent of CMOS chips, they were limited to sizes that were able to fit through doorways and elevators.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

In those early years or air cooling, it was believed that computers had to be kept at very narrow temperature ranges. It was common to find temperature specifications for designs of computer rooms that required 68 ± 1°F. To make matters more complicated, it also required that these temperatures be maintained all around the compute devices since they typically had inlets on all sides. (Typically, discharge or hot air was at the top.)

Load densities began to increase, and they have continued to increase at dramatic rates. To this day, "Moore's Law," the observation that integrated circuits double in density roughly every 2 years, continues to hold. As load densities increased through the 1990s and into the first decade after 2000, removing heat from limited portions of the compute devices became unsustainable. Standards began to be developed, both by the telecommunications industry and the data center market, to define a workable protocol to allow for a more efficient and sustainable data center. Front-to-back airflow protocol was developed, and this led to the use of hot aisle/cold aisle configurations.

As attention to energy efficiency increased, it became more obvious that the better the airflow was managed within a data hall, the less the air that was needed to cool the IT equipment.

As standards were being developed, the IT equipment manufacturers also relaxed the requirement for strict temperature control. It's not that the chips themselves could eventually operate at higher temperatures. In fact, there was nothing in the technology that changed the temperature requirement at or adjacent to the surface of the chips. The change was prompted strictly because of the prevalence of the hot aisle/cold aisle configurations and the associated separation of airstreams. With less hot air to mingle into the cold air entering the IT equipment, the supply temperature could then be raised by an equivalent amount. An immediate benefit of raising the supply air temperature entering the data hall or computer rooms is that the higher temperature allows more hours of economizer (free cooling without the means of mechanical refrigeration). The better the utilization of economizers, the better the energy efficiency of the data center.

Unfortunately, the higher the supply air temperature, the more sensitive the space becomes to the effects of recirculation and bypass. This places a greater importance on the need to provide a more uniform and maintainable cold aisle temperature.

The standards and best practices that have been developed over the years recommended minimizing recirculation and bypass through use of blanking panels, grommets, air brushes, etc., all of which help improve effectiveness of air distribution.

In short, reduced recirculation and bypass not only improve energy efficiency of the data center but also allow for a more uniform thermal environment for the IT equipment.

23.3 AIR DELIVERY METHODOLOGY AS IT RELATES TO CONTAINMENT

As load densities have continued to increase, the industry has come to the realization that there are certain high densities that, when reached, cannot be maintained with conventional hot aisle/cold aisle configurations.

Full containment, which prevents any cold aisle air from bypassing the servers and entering the hot aisle, and prevents any hot aisle air from recirculating to the cold aisle, solved many problems relating to airflow management. It also created a host of new problems, some of which includes increased cost, restricted access to cabinets and data hall, more sensitive controls, and different control strategies.

Many users found that at load densities up to approximately 150 W/ft^2, uncontained environments could be maintained, and the additional cost for controls, barriers, etc. was not justifiable. There is no clear transition point, but what is accepted by the industry is that there is a higher density at which containment is a necessity.

When containment is used, airflow distribution is not an issue. All air from air handling units (AHUs) goes to the IT equipment; all IT discharge air returns to the AHUs.

The direction of flow of the supply air from the supply outlet into cold aisles becomes irrelevant. Whether it is from above (downflow), from below (upflow), or from the side (horizontal), the air gets to where it needs to go. There is no functional difference between these different containment strategies.

Many new higher density data centers have been going to contained environments. By contrast, legacy data centers and new low-to-medium density facilities do not use containment. Although there is not much data to support the claim, it is widely believed by the industry that the lower visibility, lower density, uncontained spaces comprise the majority of all kilowatts (kW) in data center and telecom central office facilities in operation today. As long as there are uncontained environments, the issue of whether to use upflow or downflow air distribution continues to be an issue.

This chapter shall deal no further with containment since all issues relating to airflow management disappear.

23.4 AIRFLOW DYNAMICS

Establishing basic best practices in airflow management is the first step in resolving mistakes and inefficiencies that commonly plague computer rooms. Several professional organizations have addressed these issues, and luckily there really are no contradictions in the recommendations put forward.

ASHRAE TC9.9 with its *Thermal Guidelines for Data Processing Environments*, first published in 2004 and revised twice since, has changed the way owners and designers approach data center design [3].

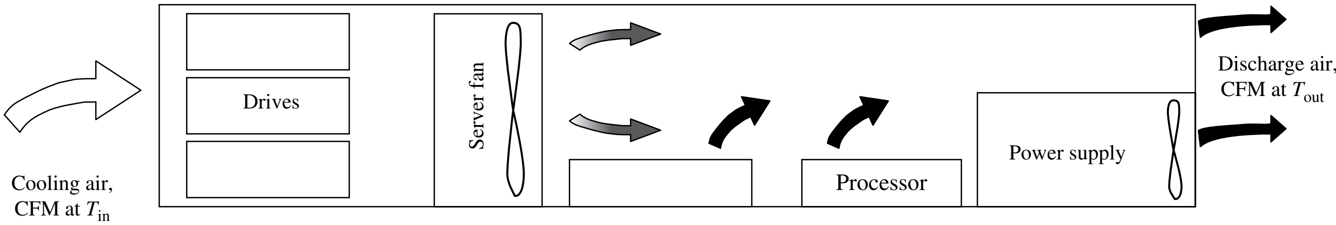


FIGURE 23.1 Representative server with passive and active airflow paths.

NEBS is another standard developed originally by Bell Labs, now managed by Telcordia. It was developed more than 30 years ago to address standards required in telcom central offices. Many of the best practices established then have found their way into the ASHRAE standards.

The basic principles relating to airflow shared by these and other standards are as follows:

1. Cabinets should be front-to-back or front-to-top-and-back airflow protocol.
2. Cabinets should be arranged to establish a hot aisle–cold aisle configuration to reduce bypass of air from cold aisle to hot aisle, or recirculation from hot aisle to cold aisle.
3. Introduce supply air into the cold aisle; remove air from the hot aisle.
4. Minimize bypass and recirculation air, to the extent possible, by using grommets, blanking panels, brushes, or any other means to close openings between servers inside cabinets, and between adjacent cabinets.

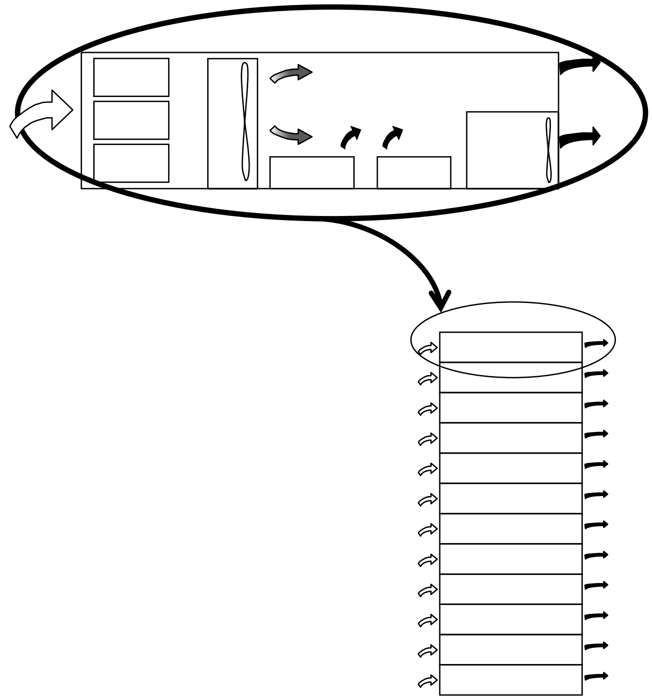


FIGURE 23.2 Representative cabinet with multiple servers.

A source of confusion for many practitioners in the industry relates to the understanding of ΔT, or temperature difference, as it relates to what happens at the servers versus what happens at the AHUs. There are actually two ΔTs to contend with, and understanding the interaction between the two is critical in understanding the airflow management concepts in uncontained environments.

Figure 23.1 shows a representative server with multiple heat-generating surfaces and passive and active airflows. A few important concepts are highlighted:

Internal components reject heat.

The airflow associated with the heat rejection at each component may be different.

Airflows may cascade, meaning that the temperature entering the first component may be lower than the one entering or passing by the second component, etc.

The logical question then follows: "Where should the temperature of the box be measured and monitored?" The correct answer is at the intake to the server box, that is, the cooling air at T_{in}. The IT manufacturer has already accounted for the fact that the second or third component will not see the same temperature as the first component.

The Thermal Envelope defined by ASHRAE's *Thermal Guidelines* book states that the INLET temperature into the server should fall within the recommended range of 64.4–80.6°F. If the temperature entering the second or third component in the direction of flow is 90°F, this doesn't matter. It's been accounted for in the design of the box and the manufacturers will stand behind that. If the outlet temperature at the back of the box is 110°F, again, that doesn't matter.

Now assume that this representative server is stacked in a cabinet with other servers, each operating at its own unique conditions. The cabinet would look something like the diagram in Figure 23.2. Each server takes the same IT equipment entering temperature from the cold aisle, T_{in}, but each may produce a different discharge temperature, T_{out}. This is because each server may be loaded to a different

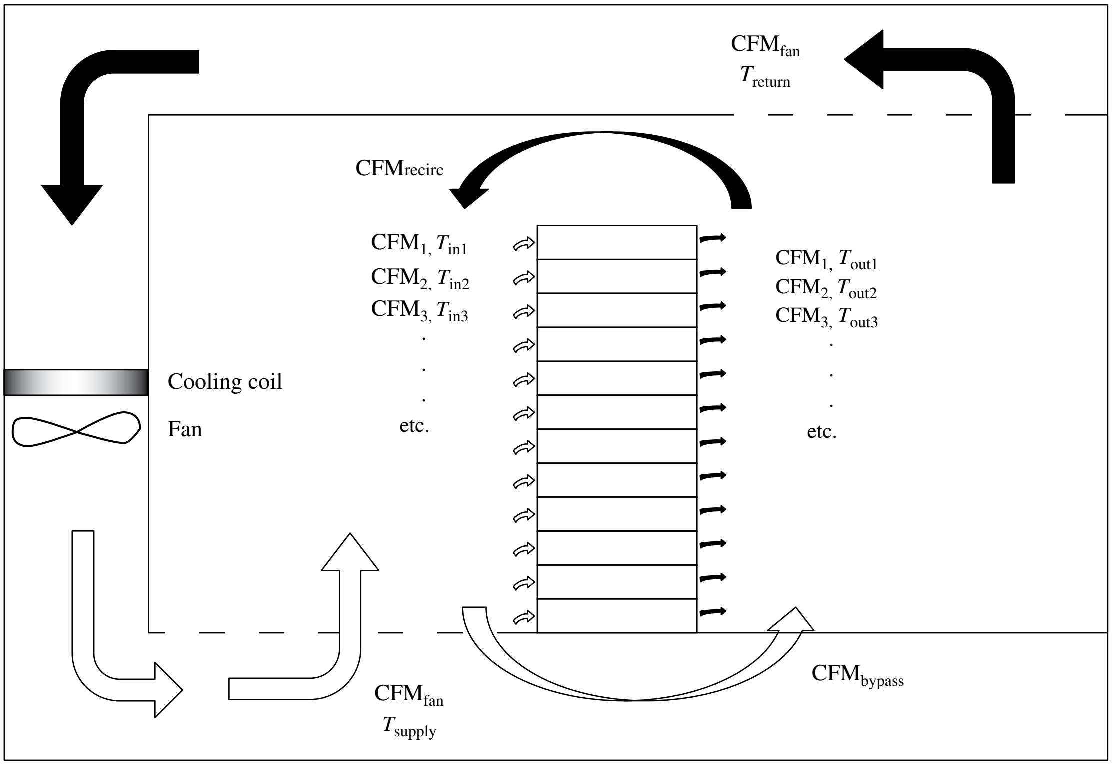


FIGURE 23.3 Representative data center.

extent, and the internal server fans may each be operating at different speeds.

The general equation for heat loss from a heat source is as follows:

$$\text{Heat Loss, } Q = 1.085 \times \left(T_{\text{out}} - T_{\text{in}}\right) \times \text{CFM}$$

An equation that can describe the heat loss for all servers in a cabinet is as follows:

$$\text{Total Heat Loss, } Q_{\text{total}} = 1.085 \times \left(T_{\text{out average}} - T_{\text{in}}\right) \times \text{CFM}_{\text{total}}$$

where $T_{\text{out average}}$ is a hypothetical average that is not easily calculated. This is because it needs to be weighted according to the airflow passing through each server, a quantity that is not usually known. Luckily, it's actually very easy to measure it—if the hot aisle condition is well-mixed, the T_{return} is actually the same as $T_{\text{out average.}}$

Now assuming that this cabinet is expanded to occupy a data center, and AHUs are added to move and cool the air, the product would look something like Figure 23.3.

This scenario is a little more complicated than what is discussed around Figure 23.2 because it introduces cold air flow into the hot aisle (bypass air), which affects the hot aisle temperature, and hot aisle air into the cold aisle (recirculation air), which affects the cold aisle temperature. Complicated as it is, the mass and heat balance must work. To simplify this process, it helps to see two different heat transfers—one representing what goes on in the room, and a second which goes to the AHU. The total airflow seen by the servers is not necessarily the same as the AHU airflow. If a cabinet server fans move 10,000 CFM of air, and the AHU moves 15,000 CFM, there has to be a net bypass flow of 5,000 CFM. If the AHU moves 6000 CFM, there has to be a net recirculation rate of 4000 CFM. There is care applied here to state "net" flow. The fact is that it is possible to have both recirculation and bypass flows at the same time and still meet the "net" balance of flows.

Because the AHU flows and the cabinet flows are not necessarily equal in uncontained environments, and because the heat transfer must be equal for both, the heat balance equation shows that the ΔT's for the AHUs cannot be the same as the ΔT's for the cabinet. The difference in flows is the net between the recirculation and bypass airflows.

These principles of airflow management must be kept in mind when considering improving how a system operates. The closer the CFM fan gets to the cabinet flow, the more efficient the system becomes, but that is true only if best effort is taken to assure that the recirculation and bypass airflows are reduced as much as possible.

23.4.1 Recirculation Airflow

When insufficient supply air is provided, hot aisle air will be pulled to make up the deficit in the cold aisle. However, recirculation airflow can also be caused even when there is sufficient supply air. If the airflow dynamics is poor and excess quantity of supply air is lost to bypass, the deficit that is lost from the cold aisle must be made up by pulling recirculation air from the hot aisle.

23.4.2 Bypass Airflow

Bypass airflow can be caused in many ways. As noted earlier, too high a velocity of supply air into the cold aisle, which is usually done in efforts to increase the airflow to the cold aisle, may actually induce the loss of more air OUT of the cold aisle. It goes without saying that gaps between cabinets, openings in floors from cable cutouts, etc. should be blocked to the extent possible. Often, a program of installing these blockages to leakage has the most direct impact in the overall performance of a data center.

It's important to carefully study every single possible source of leakage. Sometimes, the less obvious sources are the largest sources. For example, in one site, air from the raised floor was flowing into the columns through gaps between the sheet rock and the structural steel. These gaps were never noticed because they were hidden in the raised-floor environment. However, the columns were not closed above the ceiling, either, and substantial amount of air was channeled directly into the ceiling, which was used as a return air plenum. This bypass air quantity was measured at approximately 20% of the entire supply airstream.

Too much supply air increases the fraction of bypass air. Frequently, owners and operators accept large quantities of bypass air as a safety valve—it gives them comfort knowing that sufficient air is provided to the servers, even if it's at the cost of extra air being pushed through the system. As long as the environment is uncontained and the airflow delivery is effective, AHU airflow should exceed the server airflow by a small amount. When this condition is reached, the server inlet temperatures, from the lowest server in the cabinet to the highest server in the cabinet, will be as close to the supply air temperature as possible.

The best approach is to reduce the bypass fraction until "just enough" air enters the cold aisle. How much is "just enough?" The best answer to that is "just the amount it takes to keep the recirculation air quantity, CFM_{recirc}, at 0."

When CFM_{recirc} approaches 0,

$$CFM_{SUPPLY} = CFM_{IN1} + CFM_{IN2} + CFM_{IN3} + \ldots + CFM_{BY}$$

and

$$T_{IN1}, T_{IN2}, T_{IN3}, \ldots \rightarrow T_{SUPF}$$

Normally, facility operators look for the worst-case T_{in}, and they lower the T_{supply} to make sure that no T_{in} exceeds the top of the ASHRAE recommended range. Sometimes, the facility is designed for lower T_{supply} than the top of the recommended range, which in turn lowers all other T_{in} by a corresponding amount. This comes at a great cost. The lower the T_{supply}, the more refrigeration energy is used. For every degree that the T_{supply} is lowered, the chilled water supply temperature must be lowered by 1°. For each degree that the chilled water temperature is lowered, the penalty paid in consumption of energy at the chiller, which produces that chilled water, is roughly 2% [4]. (If cooling is provided by a DX system, the penalty for lowered supply air temperature is similar.)

For every degree that T_{supply} is lowered, the energy penalty in reduced hours of free cooling by use of economizers can be on the order of hundreds of hours, depending on the climate where the facility is built.

Additionally, when the T_{supply} is lowered, it increases the proportion of latent cooling that the cooling coil produces. Latent cooling does not lower the temperature of the air; it only extracts water from the air. The coil capacity is usually fixed—the more latent cooling that it is forced to provide, the more AHUs are needed to provide the sensible cooling (that part that DOES lower the temperature of the air), and this is a very wasteful approach. It also forces the issue of requiring that the facility ADD back the moisture that was already extracted in order to maintain a fixed moisture content in the data hall, and this means the addition of extra humidifiers and water consumption.

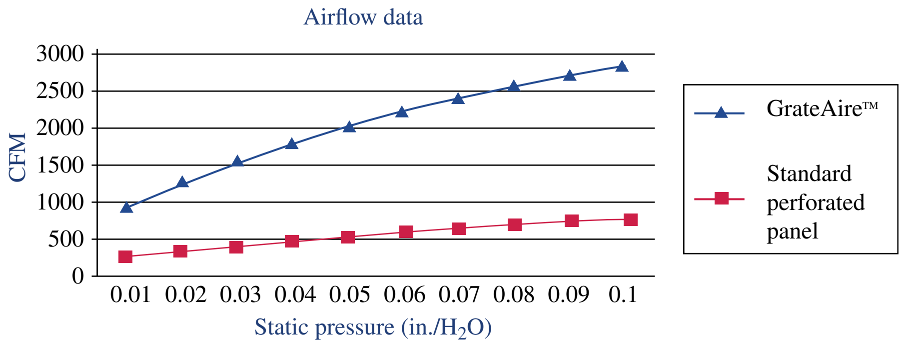
The importance of eliminating recirculation air and maintaining the T_{supply} as high as possible cannot be overstated. Reducing the bypass air quantity to save on fan energy is important also, but having a slight amount of bypass is not anywhere near as significant as having a slight amount of recirculation air.

There are strategies, both via automatic controls or via manual practices, to assure that this balance—that is, maintaining all T_{in} as close to T_{supply} as possible, using as high a T_{supply} as possible, eliminating CFM_{recirc}, and minimizing CFM_{bypass}—can be achieved. These are addressed in the sections to follow.

23.5 UNDER-FLOOR AIR DISTRIBUTION

Delivering cooling air into a data hall through a raised access floor is perhaps the most common method for providing cooling to data centers today. It can be achieved and optimized in many ways, some of which are listed in the following:

1. Modulate CRAH/CRAC/AHU airflow in unison to maintain constant under-floor pressure; move or place perforated tiles or grates to match loads.
2. Modulate CRAH/CRAC/AHU airflow in unison to satisfy the worst-case temperature sensor (i.e., the highest measured temperature) in any cold aisle.
3. Modulate CRAH/CRAC/AHU airflow in unison to maintain constant under-floor pressure; place dampered tiles, fan-assisted tiles, or variable free area tiles in specific locations to modify airflows to direct air where most needed.



GrateAire™	897	1269	1554	1794	2006	2198	2374	2537	2691	2837
Standard perforated panel	245	346	424	490	548	600	648	693	735	775

FIGURE 23.4 Performance of perforated tiles and grates. (Data from Tate Access Floors.)

There are probably as many other means to deliver cooling air to a data hall as there are designers, and it would be impractical to address each one individually within the scope of this chapter. The recommendations listed in the following address only the first method proposed earlier (i.e., modulate units in unison to maintain under-floor pressure). It is arguably the simplest and most reliable method of optimizing the airflow management strategy, but there is little doubt that it works well. The following discussion will also address some of the shortcomings of the other listed methods:

Use variable-flow CRAH/CRAC/AHU unit. In order to deliver the air into a raised access floor, the units must be downflow. Many facilities use older constant flow motors/fans, and invariably they use a significantly higher amount of energy. Converting these units to variable flow almost always provided the fastest payback on the investment. Not only is energy saved, but the airflow management strategy can be improved.

Use the ceiling as a return air plenum. This requires that the ceiling above the hot aisles be left open, and that the return air opening at the tops of the CRAH/CRAC/AHUs be ducted into the ceiling. If there is no ceiling, the performance of the space may actually be slightly better than if there were a ceiling, but it is important that the return air opening of the CRAH/CRAC/AHUs be ducted to within 2–3 ft of the top of the data hall. The taller the data hall, the better the stratification of the hot and cold airstreams, meaning that the amount of recirculation air can be reduced [5].

A floor height of 36 in. is usually ideal and can support a wide variety of load densities. Older facilities or buildings are sometimes limited to lower heights. Under those circumstances, it becomes particularly important that there be no substantial obstructions in the under-floor plenum. Major pipe runs, large conduits, cable trays, etc., if required to be placed in the under-floor plenum, should be routed away from the locations where perforated tiles or grates will be placed. They should also not be placed near the fan outlets.

Use enough pressure sensors to get a reliable measure of the under-floor pressure relative to the room pressure. Intersperse them as evenly throughout the under-floor plenum as possible. Sometimes, rules of thumb such as the "use a sensor for every 1000 ft^2" are applied, but the better the design of the under-floor plenum (in terms of few obstructions) the more uniform will be the pressure distribution, and fewer sensors can be used. Obtaining a relatively uniform pressure distribution across the entire raised access floor by following the guidelines noted is crucial in making this airflow management strategy work.

Modulate all units in unison through the VFD to maintain a constant under-floor pressure as indicated by the average of the pressure sensors.

Typical under-floor pressure set points can vary, but the following are good rules of thumb (refer to Fig. 23.4, which shows the performance of typical perforated tiles and grates as a function of under-floor pressure):

1. For perforated tiles, use approximately 0.03 in. w.g. Higher under-floor pressures can be used, and in fact should be used for contained environments; but in uncontained environments, the higher flows associated with the higher pressures cause a significant increase in the fraction of bypass airflow. The goal is to be able to slowly fill the cold aisle from below, with minimum disturbance, through the floor tiles. The more the disturbance, the higher will be the fractions of bypass and recirculation airflows. Ideally, the final level of the boundary between the cold and the warmer air above it should rest just above the tops of the cabinets. Naturally, there will be some leakage over the tops of the cabinets and out the ends of the aisle. That amount of leakage should be the extent of the bypass flow.
2. For grates, use a lower pressure such as approximately 0.015 in. There is a good reason for this: since grates have larger free areas, they discharge a larger volume of air, resulting in higher average velocities across the standard 24 in. × 24 in. tile. The higher discharge velocity tends to disturb the strategy

of slowly filling the cold aisle from below. Computational fluid dynamics (CFD) modeling and anecdotal evidence show that pressures significantly higher than 0.015 in. lead to significant losses off the tops of the cold aisle. A casual observer walking through that cold aisle will feel the airflow and will think that sufficient cooling air is provided. In fact, much of the air bypasses the servers, forcing hot aisle air to recirculate over the tops of the cabinets and into the fronts of the servers.

3. The next issue to address is how to decide whether to use grates or perforated tiles. Since higher density loads required more air, it follows that the higher density loads should use grates while the lower density loads should use perforated tiles. Although that is a true statement, it requires a significant amount of explanation because the selection of tiles is also dependent on cold aisle width and future load growth for the data hall. Placing a mixture of the two in the same field would provide a solution that is not suitable for either tile selection since the strategy requires maintaining only a single under-floor plenum pressure. That pressure can only be suitable for one type of tile. More discussion on this issue follows.

The approach of adjusting the number of tiles to match the data center load requires that the installed load be closely tracked. In fact, the load should be tracked by cold aisle. This is an important concept that must be followed for this airflow management strategy to work. Since the cabinets create the boundaries of the cold aisles, it makes sense to calculate how many tiles at a given pressure set point must be placed within that boundary; that number is completely dependent on the total load placed within that cold aisle. Use the cold aisle kilowatt load to calculate the airflow. A sample calculation is shown in Figure 23.5.

In this sample, a cold aisle is bound by 100 kW of load. For a perforated floor tile field with each tile at 400 CFM at 0.03 in. of under-floor pressure, a total of 40 tiles will be required to meet the load. Accounting for bypass of 20%, the amount of tiles should be increased to 48 tiles.

This approach is useful when considering how wide the aisles should be. The end user should have a concept of what is the maximum amount of load to be supported by a single aisle. Calculating the airflow for that aisle tells the user how many tiles are needed. For this current example, 48 tiles can fit into a cold aisle that is 4 ft wide if the aisle is 24-tiles long (i.e., ~48 cabinets in total). Fewer tiles can be installed if the load is smaller, but no more than 48 tiles may ever fit.

If the cold aisle is shorter than 24-tiles long, two approaches can be taken to allow the resultant space to address the maximum expected cold aisle load:

1. Use a 6 ft cold aisle with perforated tiles; 48 tiles would still be placed in the cold aisle, but there are three rows of tiles in the field to use, which would need to be only 16-tiles long at the least.
2. Use a 4 ft cold aisle with grates. The higher capacity tiles at 0.015 in. still allow for higher airflow rates than a standard perforated tile at 0.03 in., and only 20 tiles are needed to get the same capacity as the 48 perforated tiles; 20 tiles can easily fit within the 4 ft cold aisle.

Cold aisle kW load ⟶ Number of standard perforated tiles in cold aisle

kW load	Cooling load (BTUH)	Required tons of cooling (tons)	Required air flow (CFM)	# of standard perforated tiles
10	34,130	2.84	1,573	4
20	68,260	5.69	3,146	8
30	102,390	8.53	4,718	12
40	136,520	11.38	6,291	16
50	170,650	14.22	7,864	20
60	204,780	17.07	9,437	24
70	238,910	19.91	11,010	28
80	273,040	22.75	12,582	32
90	307,170	25.60	14,155	36
⟶ 100	341,300	28.44	15,728	40

FIGURE 23.5 Sample calculation of number of perforated tiles.

If this calculation convinces the owner to use grates, remember to readjust the tile count for all other cold aisles because they, too, should be using grates. The spreadsheet shown in Figure 23.5 is based on 400 CFM/perforated tile, but it can be adjusted easily to account for 1000 CFM/grate.

This calculation method should be used continuously as the data hall loads change. When loads are placed in an aisle, add a corresponding number of tiles. When loads are removed, remove a corresponding number of tiles.

Nothing happens at the CRAH/CRAC/AHU unless tiles are added or removed. Users often forget this—only when tiles are added can the fans ramp up in speed. As tiles are added, the under-floor pressure drops, and the fans ramp up in speed to return the under-floor pressure to the set point. Conversely, when loads decrease, removing tiles allows the fans to slow down.

To correctly select the proper types of tiles or the appropriate cold aisle widths, it may not be enough to look only at the design conditions. The owner and designer together must consult their crystal ball to address the question of what is the practical upper limit of load density that the data hall will achieve in later phased construction projects. Day 1 and design build-out calculations may indicate that the facility will perform satisfactorily with perforated tiles. However, if there is the possibility of higher density in the life of the facility, it would make sense to start off with the higher flow tile from Day 1 because converting from all-perforated tiles to all-grates while maintaining airflow in a live facility is a disruptive process.

23.5.1 Automation of Airflow Management

With the basics established for a reliable airflow management strategy using under-floor air distribution, some discussion must follow for how to automate the airflows to adapt to changing loads.

It may seem ironic that the single most important aspect of automating airflow to the data hall is a manual component. To date, there is no way to automate the installation or removal of perforated tiles or grates. In fact, if a designer wanted to automate the process of modulating the flow of air through the floor tiles, he or she could treat it in a manner analogous to a conventional ducted system serving commercial office spaces. The use of variable air volume (VAV) boxes have become practically ubiquitous throughout the HVAC marketplace, so one would ask why not add VAV boxes to data center automation. The quick reason is cost. The largest VAV boxes available today provide the equivalent airflow of two or three grate tiles. To achieve the huge flows associated with large-scale IT loads would require hundreds or thousands of boxes, and the cost of the automation would be prohibitive and the installation extremely complex.

Manufacturers have devised means to put the automation directly into the tiles or grates. These products have included fans and automatic dampers, making them behave more like VAV boxes and/or fan-powered boxes, but the cost and complexity of installing these on a large-scale data center with loads that continue to grow throughout the life of the facility makes these solutions very impractical.

But expecting a data center automation system to respond like a commercial office space misses a very critical point—data center loads do not behave like commercial offices. The commercial office space may be expected to reach 100% load in its first year of operation, and that peak would most likely occur on the day when the outdoor design conditions are reached. Hence, a commercial office space needs its full complement of VAV boxes from Day 1. The data center growth is very different. Whereas the loads may be very steady over the course of a single day or year, the loads usually continue to increase dramatically over the course of the life of the facility.

Getting back to the issue of manually adjusting tile placements—a necessity for under-floor air distribution in data centers:

1. As discussed, tiles should be managed in a manner to meet the loads per aisle. This may involve creating tables/charts to keep up-to-date information relating to the total load per cold aisle. Calculate the amount of tiles required per cold aisle based on this load. The user may choose to add a safety factor (10–20%) to assure that there is a slight amount of bypass air but no recirculation air.
2. Based on these calculations, manage the number of tiles to place per cold aisle. There is no practical way to automate this process. When loads increase, someone has to add more tiles; when loads decrease, someone has to remove tiles.

23.5.2 Control Strategies

1. Modulate all available AHUs in unison to maintain the under-floor pressure control point. Control strategies have been used to modulate each AHU individually to control to specific sensor locations, but this doesn't work because the under-floor air dynamics change drastically when the proportion of airflow changes from one AHU to another. A sensor that at one moment receives air from a unit on one side of the building may very likely receive air from another unit when a few tiles are changed or when airflow at other units changes. Modulating all units in unison provides the most stable controls.
2. All AHUs should be set to the same supply air temperature set point, and that set point should be as high as possible. (See discussions in Section 23.4.2.)
3. When more openings are added (be it removing solid tiles and/or placing perforate tiles), all fans ramp up to maintain the under-floor pressure.

4. Monitor the temperature of the air entering the tops of the cabinets. If the tops of cabinets are not meeting the temperature requirements, meaning that there is substantial amount of recirculation airflow, reset the under-floor pressure to a slightly higher value. Repeat until the temperature requirements are met.
5. A variant of the control strategy described earlier uses temperature sensors near the tops of cabinets in lieu of under-floor pressure sensors. The fans speed up or slow down to assure that all the temperature sensors meet their set point. If a temperature sensor reads too high, the fans speed up; if all temperature sensors are satisfied, the fans can slow down until a single or a predetermined fraction of sensors reaches too high a temperature. Other than the revision of using temperature sensors in lieu of pressure sensors, this control method should use the same airflow management strategies described earlier, including modifying the numbers of perforated tiles or grates based on the load per cold aisle.

23.6 OVERHEAD AIR DISTRIBUTION

The delivering of cooling air into a data hall through an overhead air distribution system, though less common than under-floor air distribution, can be equally effective. Telecom central offices have used this methodology for years, and practitioners of it have become very comfortable with it. If a raised floor exists at all, it is not used for air distribution. Optimizing airflow with overhead air distribution can be achieved using various methods.

The main difference between under-floor and overhead air distribution is that the CRAH/CRAC/AHU units are usually upflow. Manufacturers can provide upflow units that work just as well as downflow ones. However, the layout of the space in conjunction with the configuration of the units is messy because the return air openings to them must be at the low end of the units. This means that cold air from the cold aisle, which inherently wants to stay low in the space relative to the hot air, is more likely to be returned to the unit. This constitutes an increase in bypass air, which forces more air to be needed to meet the load.

There are a few approaches to deal with the bypass airflow:

1. Create a partial height barrier around the units to force air from the higher portions of the space only (i.e., the hot aisle air) to return to the units.
2. This creates a fan gallery. The issue of using modular CRAH units, as opposed to a custom AHU, becomes less of a driver. An AHU can be configured to accommodate any special arrangement, which typically allows for larger sizes of units and more customized selections.
3. The overhead approach may complicate the logistics of supplying air from above while also returning air from high in the space. A ceiling plenum does not necessarily help the situation because the overhead supply ductwork and the ceiling plenum return are both located in the same space. To accommodate this arrangement, the data hall must be tall—say at least 18–20 ft high.
4. A tall space can more easily accommodate a layering of the airflows. The supply ductwork can be placed in a layer above the aisles while the hot aisle air rises above the supply ductwork. The return ductwork or fan inlets can capture the return air from that high point and return the air back into the system to be cooled again.

The following recommendations comprise one strategy for accomplishing overhead air distribution in data halls. Others may also work, but this recommended approach is a good starting point:

1. Use VFD-controlled CRAH/AHU units. Whereas downflow units are recommended to be controlled to maintain a constant under-floor pressure, overhead systems should be maintained to control to a constant duct static pressure. This is analogous to the control strategy utilized by many commercial VAV systems. As a means to achieve this pressure-based control, a main supply duct should be used as a manifold to receive the airflow from multiple units.
2. The manifold duct should be sized to accommodate the maximum amount of flow expected through it when the load reaches 100% of design. The previous section dealt extensively with how to design an under-floor air distribution plenum. There is not a significant reason to go extensively into the design of overhead ductwork for this section because this ductwork is not any different than what would be designed for commercial office spaces. The HVAC industry already has this type of design well-documented.
3. Branch ducts should be used to deliver air to the cold aisles. One branch duct per cold aisle—this is necessary so that the instantaneous quantity of required air to the cold aisle can be controlled from a single control damper.
4. Temperature-based control by cold aisle is the appropriate methodology for overhead air distribution. Place a representative number of temperature sensors at high points of cabinets; use the worst-case measured temperature in that cold aisle to modulate the branch damper to assure that the temperature set point is satisfied.
5. Since the airflows to each cold aisle usually far exceed the sizes of most commercial VAV boxes, dampers

must be used in lieu of VAV boxes. The dampers may act like VAV boxes to some extent, but there is not a practical way to use flow-measuring devices upstream of the dampers. Consequently, these devices behave like pressure-dependent VAV boxes.

6. As a result, the manifold supply duct must be kept at a constant pressure by modulating all the fans in unison. To assure that the pressure across the entire length of the manifold duct is relatively uniform, that duct should be as straight and simple as possible, and it should be generously oversized.
7. The presence of the dampers in the branch ducts enables a feature that is not available in the under-floor air distribution model. This feature is the ability to modulate airflow into the cold aisle based on the load in that aisle. This modulation is based on the temperature readings within the aisle. This means that the load need not be tracked manually by cold aisle, as it must be done for the under-floor air distribution model. Additionally, since the branch damper can vary the flow of air through the openings in the duct by cold aisle, there is no manual adjustment of the delivery system required.
8. Section 23.5.1 describes how an under-floor system can be controlled through automation; yet, despite the automation, a manual component—that is, monitoring the load, adjusting the numbers of tiles accordingly—is required to make the system work. The overhead system does not have this limitation. Once the system is established, the automation can control the air flow into the cold aisle through a wide range of loads within the cold aisle with no human intervention. It is a much simpler system to maintain.

23.6.1 Effects of Aisle Width, Placement

1. Since the branch ducts providing air to a single cold aisle can adjust the airflow to the aisle, the width of the aisle is not as critical a component as it is for the under-floor air distribution model. Provided that the aisle is selected at a width appropriate for the maximum velocities expected in that aisle or any aisle in the data hall, all aisles can be designed to the same width.
2. The under-floor system has issues relating to the discharge velocity. What is an appropriate discharge velocity for an overhead system? There is not a single correct answer. The most prudent approach would be to model the airflow dynamics in an aisle at the design flow using CFD. The designer should select the initial velocity out of the supply duct as low as practical (meaning the opening is as large as can be accommodated for a reasonable cost and layout), and assuring that the velocity of the airflow that rides UP the front of the servers is not too extreme to prevent it from entering the fronts of the servers. Often, the discharge velocity can be anywhere between 500 and 1000 ft/min. Once that discharge air hits the floor, disperses, and follows the contours of the fronts of the cabinets back up to the tops, this velocity can be greatly reduced. If the velocity into the aisle is too large, the air entering the cold aisle can be induced to flow right out of the cold aisle again—analogous to trying to fill a glass of water under a faucet at full flow. CFD modeling should be used extensively to optimize this flow to "as small a bypass airflow as possible."
3. The opposite problem of too low a velocity is actually not a problem. With the proposed arrangement, the situation is self-correcting. When the air does not reach the tops of the server cabinets, the temperature sensors indicate they are not satisfied, thereby driving the damper to a more open position, placing more air into the cold aisle.
4. These design issues may appear to be too much trouble to make the overhead system a worthwhile approach. No doubt, it does require more advanced planning. But once the system is set up properly and the initial velocities are determined to a workable solution, the advantage of full automation (no human intervention) makes this an easier system to maintain over the life of the facility. Additionally, as the system grows, no human intervention is required to make the appropriate adjustments.

23.6.2 Automation of Airflow Management

Much has already been stated about the automation for an overhead system. These are the basic requirements to make it work:

1. Temperature sensors must be placed at the high points of the cabinets.
2. The controls system must monitor the multiple sensors within the cold aisle and drive the branch damper to open more or close more to maintain the temperature set point. (Anecdotal information indicates that sometime using the average of the temperatures rather than the worst-case temperature in the cold aisle gives more stable control of airflow into that cold aisle.)
3. A pressure sensor should be placed in the duct manifold. The AHUs shall all be modulated in unison to maintain a constant manifold pressure.
4. As more branch dampers open, the duct pressure will drop. The system shall respond by speeding up the fans to return the duct pressure to the set point. The same approach shall be taken in reverse when the dampers close.

5. The duct static pressure can be reset to assure that the branch dampers are almost fully open. This would optimize the fan energy usage, and can be built into the automation system.

23.7 CONCLUSION

Overhead and under-floor air distribution, the two main methods of delivering cooling air to uncontained data halls, each has advantages and disadvantages.

Under-floor air distribution has greater flexibility in placing air distribution outlets where they're needed since perforated tiles or grates can easily be relocated to deliver cooling air to where the load occurs in the greatest concentrations. Additionally, the placement of these tiles, as well as the placement of the rows of cabinets, does not need to be determined from Day 1. If greater density is needed after Day 1, the next rows of cabinets can be placed on the floor with a wider cold aisle width, thereby allowing more air to be delivered to higher density cold aisle.

The main complication found with under-floor air distribution is that the selection of the tiles and the velocity of the air through those tiles are critical. Too high a velocity means that the air will not be captured by the IT equipment, thereby encouraging a larger fraction of bypass air. Too low a velocity means that the cold air will not reach the top servers in the cabinets, thereby encouraging a larger fraction of recirculation air. This balance of airflows—in essence filling the cold aisle from below—is a complicated issue that often requires sophisticated CFD modeling to resolve.

Overhead air distribution is less flexible in the planning phase because it usually requires that the overhead ductwork be placed above the data hall space, and sometimes this occurs before the cabinet layouts are fully determined. The cold aisle and hot aisle spacing must be determined by Day 1 so that no construction projects (i.e., installing ductwork) need to occur over a live data center.

There is a big upside to overhead air distribution. The benefit is that the system can be automated to adjust for varying loads by cold aisle. As loads change and as the facility grows, the system can adjust without any human intervention. An under-floor air distribution system requires that the loads be monitored closely by cold aisle, and that the tiles be moved accordingly. Without the tiles being moved, the system cannot respond with an increase or decrease in fan speed.

Many CFD models have shown that the cold air delivered from overhead fills the cold aisle from the top down, then follows the contours of the fronts of the IT cabinets right back up to the tops of the cabinets. There is usually less of a concern about reaching the tops of the cabinets because the extra velocity associated with the discharge from above seems to provide enough inertia to allow the air to wash the entire fronts of the cabinets.

Additionally, because of this more uniform distribution, the delivery of high-density cooling air to the cold aisle is less sensitive to aisle width. A width can be selected for all aisles within a data hall, and the same width can be used throughout, regardless of the load density at each aisle.

The issue of containment has only been addressed on a cursory basis in this chapter, because once containment is implemented, the method of delivery becomes irrelevant as does the velocity from the point of discharge. High-density data centers should use containment. For lower-density data centers, there may be cost or logistical reasons as to why containment is not used. But once it is decided to NOT use containment, these differences between overhead and under-floor air distribution must be considered and analyzed for applicability.

REFERENCES

[1] Beaty D, Schmidt R. Back to the future: liquid cooling data center considerations. ASHRAE J 2004;46(12).

[2] ASHRAE. *Thermal Guidelines for Data Processing Environments.* 3rd ed. Atlanta: ASHRAE; 2012.

[3] ASHRAE. *Considerations in Data Center Energy.* 2nd ed. Atlanta: ASHRAE; 2009.

[4] Moss D. Under-floor Pressure Control: A Superior Method of Controlling Data Center Cooling. ASHRAE Transactions, Vol. 118 Issue 1, p3; Atlanta; ASHRAE; 2012.

[5] Sorell V, Khankari K, Abougabal Y, Gandhi V, Watve A. An Analysis of the Effects of Ceiling Height on Air Distribution in Data Centers. ASHRAE Transactions CH-06-9-2; Atlanta: ASHRAE; 2006.

FURTHER READING

Schmidt R, Iyengar M. Comparison between underfloor supply and overhead supply ventilation designs for data center high-density clusters. ASHRAE Trans 2007;113(1):115–125

Sorell V, Escalante S, Yang J. Comparison of overhead and underfloor air delivery systems in a data center environment using CFD modeling. ASHRAE Trans 2005;111(2):756–764.

Herrlin M, Belady C. Gravity-assisted air mixing in data centers and how it affects the rack cooling effectiveness. Proceedings of the Tenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronics Systems (ITHERM '06); May 30-June 2; San Diego, CA 2006. p 438, 5 pp.

Mulay V, Karajgikar S, Iyengar M, Agonafer D, Schmidt R. Computational study of hybrid cooling solution for thermal management of data centers. Proceedings of the ASME 2007 InterPACK Conference collocated with the ASME/JSME 2007 Thermal Engineering Heat Transfer Summer Conference, Volume 1; July 8–12; Vancouver, British Columbia, Canada; 2007. Paper No. IPACK2007-33000. p 723–731.

24

HOT AISLE VERSUS COLD AISLE CONTAINMENT

DAVE MOODY
Schneider Electric ITB, O'Fallon, MO, USA

24.1 EXECUTIVE SUMMARY

Hot Aisle Containment system (HACS) or Cold Aisle Containment system (CACS) can be installed in new or existing data centers, in conjunction with internal or external cooling systems, and either on raised-floor systems or hard-floor (overhead) systems. Passive ducts directly from racks (chimney) are considered a special case of HACS.

Both Hot Aisle Containment (HAC) and Cold Aisle Containment (CAC) are capable of their primary function, which is separation of the cool supply air and hot return airflow streams in a data center.

Both HAC and CAC may provide a benefit to some degree as follows:

a. Increase cooling system coil capacity, by affecting Return Air Temperature (RAT)
b. Increase cooling system coil efficiency, by affecting RAT
c. Decrease cooling system fan power
d. Provide predictable and reliable IT equipment inlet air temperatures
e. Improve redundancy in row-based cooling system by extending the sphere of influence for the cooling units

HACS and CACS also impact data center operations as follows:

f. Affect personnel working conditions inside the data center or inside the containment system
g. Impact ride-through time for cooling system and data center temperatures in the event of cooling system failure or cooling system power loss
h. Impact conditions for peripheral equipment in the data center outside any of the HAC or CAC zone(s)
i. Impact economizer operation time periods during cooler outside ambient temperatures

24.2 CONTAINMENT: THE AIRFLOW ARCHITECTURE MODELS

See Figure 24.1 for traditional data center design row-based IT loads and Figure 24.2 for side view of CAC and HAC, with raised floor.

24.2.1 HAC

HACS deployment may be based on two different airflow architecture models, in accordance with descriptions as provided by Schneider Electric White Paper #55 [1].

24.2.1.1 Fully Ducted Return: Flooded Supply In this architecture model (Fig. 24.3), hot return air as exhausted by the row-based IT equipment is directed into a closed return air system and passes from the outlet of the IT equipment air exhaust system through ductwork or a physical structure to the inlet of the cooling system. The physical containment structure assists in preventing any hot air from mixing with the room air or IT equipment supply air in the room and/or cold aisles. The system may include either internal cooling units placed in the rows with the IT equipment racks and close-coupled by way of the containment common hot aisle volume to the outlets of the IT equipment racks or an external cooling system comprised of cooling unit(s) that shares their inlets with the fully ducted

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

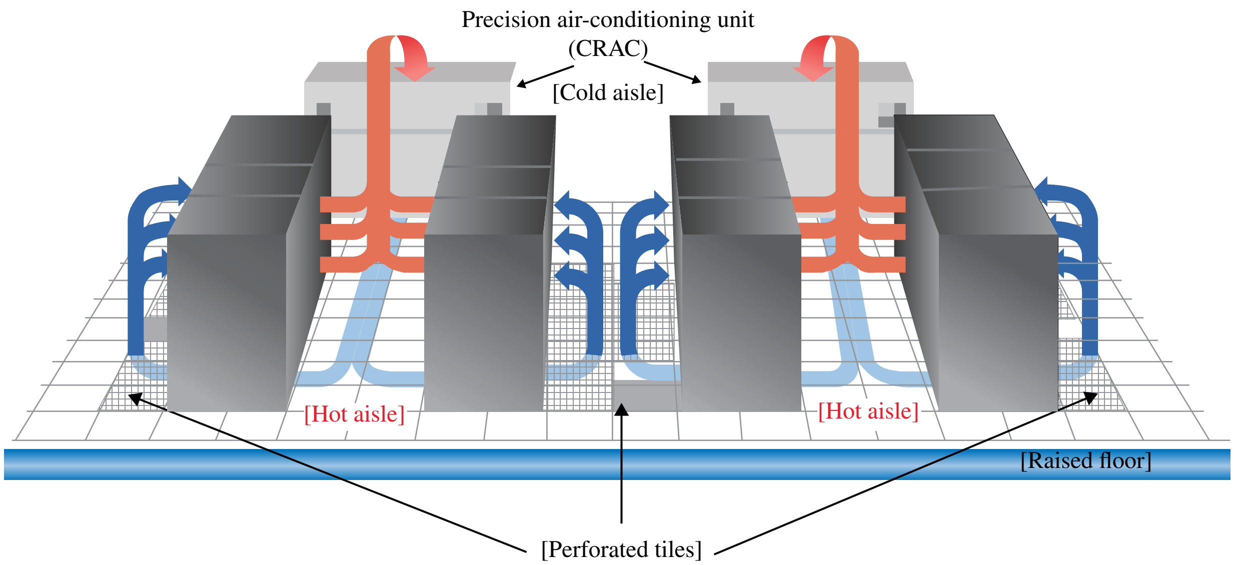


FIGURE 24.1 Hot aisle/cold aisle configurations with raised floor. Courtesy of Emerson Electric Corp.

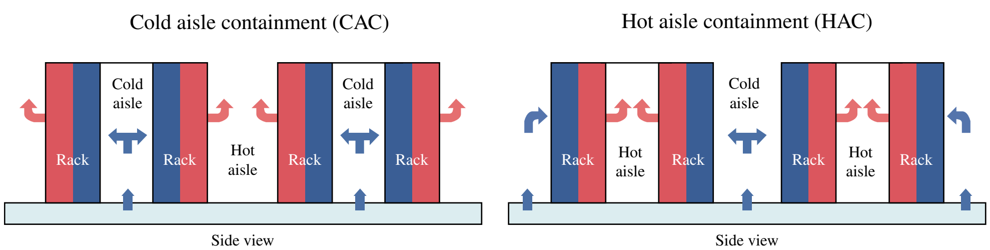


FIGURE 24.2 Side views of CAC and HAC, with raised floor. Courtesy of Emerson Electric Corp.

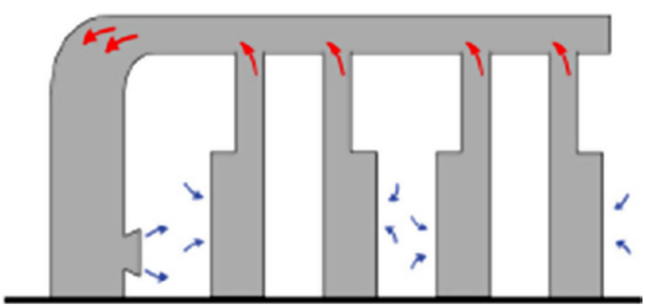


FIGURE 24.3 Fully ducted return—flooded supply. Courtesy of Schneider Electric.

return system but is located outside the rows and containment system, typically at the perimeter of the data center room, or in a location completely outside the data center room.

Cooling system total airflow matches or exceeds IT equipment total airflow in the return duct system, so that all of the IT equipment exhaust air is captured by the cooling return air system, and there is no exit or leakage of IT equipment hot exhaust air from the return system into the room.

Supply air is provided from the outlet(s) of the cooling units to the room without the use of any supply air ducting or distribution system in the room. Cooling system total airflow matches or exceeds IT equipment total inlet airflow requirement. Once the proper amount of supply air from the cooling units enters the room and is available to the IT equipment inlets, the IT equipment fans automatically draw from the room air the required airflow per IT equipment unit or server. The room air is the supply air for the IT equipment.

24.2.1.2 Fully Ducted Return: Locally Ducted Supply In this architecture model, hot IT equipment exhaust air is managed in the same manner as in the Fully Ducted Return—Flooded Supply architecture. Cooling system total airflow matches or exceeds IT equipment total airflow requirement (Fig. 24.4).

Supply air is provided from the outlet(s) of the cooling units to the room by means of a supply air ducting or distribution system in the room. Once the proper amount of supply air from the cooling units enters the room, some means of ducting the air closer to the IT equipment inlets is used, typically a raised-floor delivery system with perforated or partially open floor tiles to provide supply air into the cold aisle(s), or ducting overhead to the room space above the cold aisles, where supply air is available to mix with the room air and eventually be drawn into the IT equipment. The IT

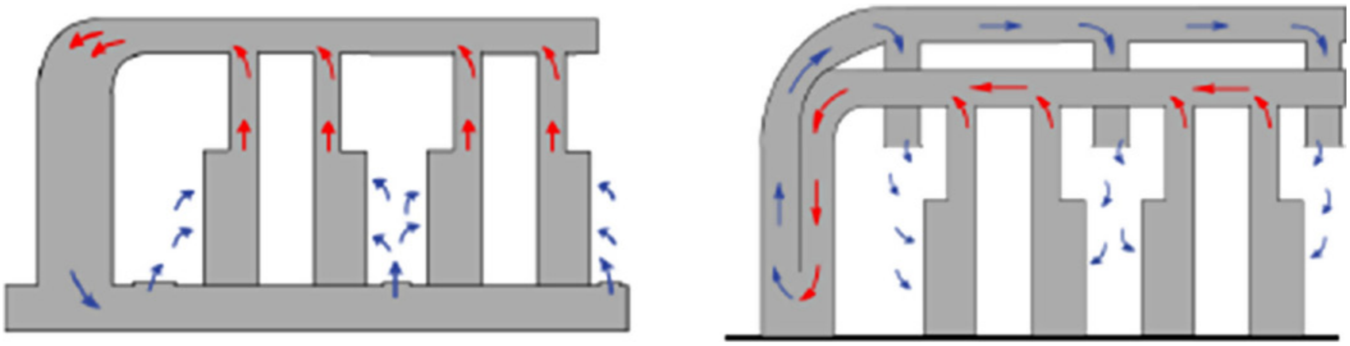


FIGURE 24.4 Fully ducted return—locally ducted supply (Left, Raised floor; Right, Overhead). Courtesy of Schneider Electric.

equipment fans automatically draw from the cold aisle air the required airflow per IT unit. The mixture of room and cold aisle air is the supply air for the IT equipment.

This case more often applies in an existing data center with a previously installed raised-floor or overhead local supply duct system, where HAC as a Fully Ducted Return system is retrofit for efficiency and performance gains. In a new data center, with Fully Ducted Return architecture, there is no benefit to installing any supply duct system.

Some industry documents imply that with the Flooded Supply architecture, the cold air distribution to the servers is open and exposed to disturbances from the room and has a higher risk of not providing the server with its required input temperature. Such concerns are not valid with a properly designed and managed Flooded Supply system, in conjunction with a Fully Ducted Return system, where the IT equipment inlet temperatures are constantly monitored and cooling system performance for coil heat removal and airflow is adjusted to maintain acceptable IT equipment temperatures throughout the data center.

It is a caution to data center managers to deploy any containment system from vendors who do not coordinate and provide for monitoring and adjusting the cooling system coil(s) heat removal and airflow performance based on real-time data from the management system common to both.

24.2.2 CAC

CAC deployment may be based on two different airflow architecture models, in accordance with descriptions as provided by Schneider Electric White Paper #55 [1].

24.2.2.1 Flooded Return: Fully Ducted Supply In this architecture model (Fig. 24.5), cool IT equipment supply air as provided by the cooling equipment is directed into a closed supply air system and passes from the outlet of the cooling unit system through ductwork or a physical structure to the inlet of the IT equipment. The physical containment structure assists in preventing any cool supply air from mixing with the room air or return air in the room and/or hot aisles, making it available to the inlets of the IT equipment or servers. The system may include either internal cooling units placed in the rows with the IT racks and close-coupled to the inlets of the IT equipment racks or an external cooling system comprising cooling unit(s) that shares their outlets with the fully ducted supply system but is located outside the rows and containment system typically at the perimeter of the data center room, or in a location completely outside the data center room.

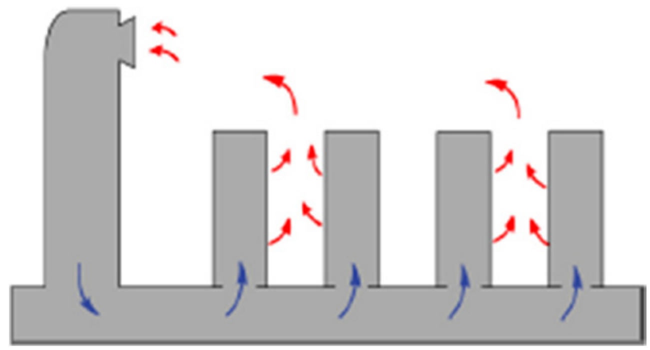


FIGURE 24.5 Flooded return—fully ducted supply. Courtesy of Schneider Electric.

Cooling system total airflow matches or exceeds IT equipment total airflow requirement plus any leakage in the supply duct system, so that all of the IT inlet air is provided by the cooling air system. Any excess of supply airflow greater than that required by all the IT equipment inlets may leak from the supply system out of the containment system into the room or hot aisles.

Return air is provided from the outlet(s) of the IT equipment to the room without the use of any return air ducting or distribution system in the room. Return cooling system total airflow matches or exceeds IT equipment total airflow plus any leakage in the supply system. Once the design amount of hot return air from the IT equipment enters the room and is available to the cooling unit inlets, the cooling unit fans automatically draw from the room air the design airflow per cooling unit(s). The room air is the return air for the cooling unit(s).

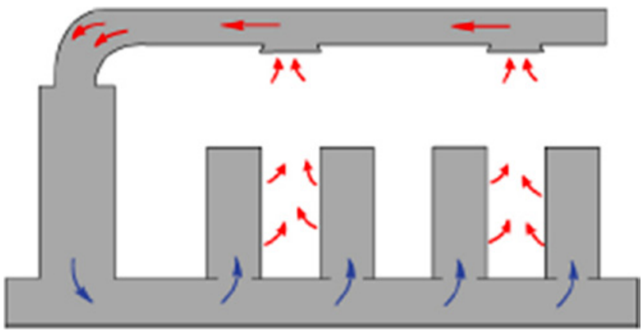


FIGURE 24.6 Locally ducted return—fully ducted supply. Courtesy of Schneider Electric.

24.2.2.2 Locally Ducted Return: Fully Ducted Supply In this architecture model (Fig. 24.6), cool supply air is managed in the same manner as Flooded Return—Fully Ducted Supply architecture. Cooling system total airflow matches total IT equipment airflow requirement plus any leakage in the supply duct system.

Warm return air is provided from the outlet of the IT equipment units to the room and from the room to the cooling unit(s) by means of a return air ducting or distribution system in the room. Once the total amount of exhaust air from the IT equipment units enters the hot aisle or room, some means of ducting the air closer to the cooling unit combined inlets are used. This is typically a drop ceiling return plenum system with perforated or partially open ceiling panel grates to provide return air from above the hot aisle(s). It may also be ducting overhead from the room space above the hot aisles, where return air is available to mix with the room air and eventually be drawn into the cooling unit inlets. The cooling unit fans automatically draw from the room air above hot aisle air the full capacity cooling airflow per unit. The room and hot aisle air is the return air for the cooling equipment.

This case more often applies in an existing data center with previously installed drop ceiling or overhead return duct system, where hot aisle return as a Locally Ducted Return system is retrofit for efficiency and performance gains. In a new data center, with Fully Ducted Supply architecture, there is no benefit to installing any return duct system, provided that any other peripheral IT equipment in the room/hot aisle is tolerant of the IT Exhaust air temperatures that result in the room air temperature equal to the IT equipment exhaust air temperature.

Where other room-based or peripheral IT equipment outside the CACS will not tolerate high temperatures equal to row-based IT equipment exhaust air temperatures, the Locally Ducted Return system may be useful in directing the IT equipment exhaust airflow into the return system with minimum mixing with room air that is in the vicinity of the peripheral IT equipment.

24.3 RETURN AIR TEMPERATURE TRENDS IN HAC AND CAC

RAT to the cooling units can have a significant impact on unit coil heat removal capacity, especially in Chilled Water (CW) cooling unit systems where the flow of CW is limited only by velocity in the cooling unit coils and pressure drop capacity of the CW pumping system. Subsequently, “the water does the work” in CW systems, and more flow through the same coil at the same entering water temperature (EWT) can provide a proportional increase in coil capacity up to the published maximum flow rate per cooling unit coil.

An example is the APC-Schneider Model ACRC100 with 45°F EWT (Table 24.1).

To summarize, a 5°F increase in RAT results in a 37.8% increase in unit cooling capacity, with a 2.5% decrease in Water Usage rate (GPM/kW) (0.724/0.742).

Correspondingly, a 10°F increase in RAT (from 80 to 90°F), with 21.7 kW cooling unit capacity, results in a 64.3% increase in unit cooling capacity, with a 4.8% decrease in Water Usage rate (GPM/kW) (0.714/0.742).

In direct expansion (DX) cooling units, the impact of higher RAT is less significant, due to the limitation of refrigerant flow capacity at maximum compressor output/speed.

For Emerson–Liebert CRV row-based DX air-cooled units, see Table 24.2.
In summary, a 10°F increase in RAT results in a 16.2% increase in unit cooling capacity.

Whether it is the 64.3% increase in a CW cooling unit or the 16.2% increase in a DX cooling unit, these gains in net sensible cooling capacity per unit can have an impact on the total number of cooling units required to manage the total heat load of any data center space.

Schneider Electric White Paper #25 [2]—*Calculating total cooling requirements for Data Centers*—can be used to calculate total cooling needed. For determining the quantity of cooling units in a data center, N quantity of required cooling units would be calculated as follows:

Rounding up to nearest whole number:

N = Total Heat Load kW/Unit Capacity kW.

For example, in a 200 kW total heat load data center with ACRC100 CW cooling units and 13.2 kW per unit,

$$N = 200\,\text{kW}/13.2\,\text{kW} = 15.15; N = 16 \text{ cooling units}$$

In a 200 kW total heat load data center with ACRC100 CW cooling units and 21.7 kW per unit,

$$N = 200\,\text{kW}/21.7\,\text{kW} = 9.22; N = 10 \text{ cooling units.}$$

TABLE 24.1 ACRC100 performance data[a]

Temperature DB, Wet-bulb	CW Delta T	Total net capacity	Sensible net capacity	Sensible heat ratio	CW flow rate*	Total CW pressure drop
°F (°C)	°F (°C)	BTU/h (kW)	BTU/h (kW)	SHR	GPM (l/s)	ft H_2O (kPa)
80°F DB, 62.8°F WB (26.7°C DB, 17.1°C WB)						
	10°F (5.5°C)	45,000 (13.2)	45,000 (13.2)	1.00	9.8 (0.62)	9.8 (29.2)
	12°F (6.6°C)	39,000 (11.4)	39,000 (11.4)	1.00	7 (0.44)	5.4 (16.11)
	14°F (7.7°C)	36,000 (10.5)	36,000 (10.5)	1.00	5.7 (036)	3.8 (11.22)
	16°F (8.8°C)	35,000 (10.2)	35,000 (10.2)	1.00	4.8 (0.3)	2.8 (8.3)
	18°F (10°C)	33,000 (9.7)	33,000 (9.7)	1.00	4.1 (0.26)	2.14 (6.4)
	20°F (11.1°C)	33,000 (9.7)	33,000 (9.7)	1.00	3.6 (0.23)	1.7 (5.18)
85°F DB, 64.5°F WB (29.4°C DB, 18.1°C WB)						
	10°F (5.5°C)	62,000 (18.2)	62,000 (18.2)	1.00	13.2 (0.83)	16.9 (50.56)
	12°F (6.6°C)	52,000 (15.2)	52,000 (15.2)	1.00	9.3 (0.59)	8.9 (26.59)
	14°F (7.7°C)	46,000 (13.5)	46,000 (13.5)	1.00	7.0 (0.45)	5.4 (16.16)
	16°F (8.8°C)	44,000 (12.9)	44,000 (12.9)	1.00	6.0 (0.38)	4.05 (12.09)
	18°F (10°C)	43,000 (12.6)	43,000 (12.6)	1.00	5.2 (0.33)	3.2 (9.46)
	20°F (11.1°C)	43,000 (12.6)	43,000 (12.6)	1.00	4.6 (0.29)	2.6 (7.77)
90°F DB, 66.1°F WB (32.2°C DB, 18.9°C WB)						
	10°F (5.5°C)	74,000 (21.7)	74,000 (21.7)	1.00	15.5 (0.98)	23.06 (68.78)
	12°F (6.6°C)	66,000 (19.3)	66,000 (19.3)	1.00	115 (0.73)	13.2 (39.51)
	14°F (7.7°C)	60,000 (17.6)	60,000 (17.6)	1.00	9.0 (0.57)	8.5 (25.2)
	16°F (8.8°C)	53,000 (15.5)	53,000 (15.5)	1.00	7.1 (0.45)	5.5 (16.48)
	18°F (10°C)	53,000 (15.5)	53,000 (15.5)	1.00	6.3 (0.4)	4.4 (13.26)
	20°F (11.1°C)	53,000 (15.5)	53,000 (15.5)	1.00	5.6 (0.36)	3.7 (10.98)

[a] Courtesy of Schneider Electric.

Note: Unit net sensible capacity with 10°F CW ΔT, 80°F DB/62.8°F WB RAT, is 13.2 kW* (45 MBH, 3.75 tons) with 9.8 GPM waterflow. Water usage rate = 9.8 GPM/13.2 kW = 0.742 GPM/kW.

Unit Net Sensible Capacity with 10°F CW Delta T, 85°F DB/64.5°F WB RAT, is 18.2 kW (62 MBH, 5.16 tons) with 13.2 GPM waterflow. Water Usage rate = 13.2 GPM/18.2 kW = 0.724 GPM/kW.

Therefore, the data center design would require (6) fewer ACRC100 CW cooling units with a 90°F RAT compared to 80°F RAT. This eliminates (6) CW cooling unit equipment and installation cost, maintenance and operating costs, and space in the data center floor plan.

CW flow rates would also be affected:

$$200\,\text{kW} \times 0.742\,\text{GPM/kW for } 80°\text{F RAT} \\ = 148.4\,\text{GPM CW flow required}$$

$$200\,\text{kW} \times 0.714\,\text{GPM/kW for } 90°\text{F RAT} \\ = 142.8\,\text{GPM CW flow required}$$

Since CW system pump HP/kW power is directly proportional to flow rate and system pressure drop, the savings in 4.5% decrease in total flow would be offset by the 235% (23.06 ft/9.8 ft total dynamic head [TDH]) increase in pressure drop across the CW cooling unit. How much this offset represents would need to account for the increase in CW cooling unit pressure drop as a percentage of total CW System pressure drop. By example, if the total CW System pressure drop is 110 ft TDH, an increase of 13.26 in CW cooling unit pressure drop would only calculate an offset pump HP/kW power increase of (13.26/110) = 12%.

With DX cooling units and 30.3 kW per unit, in a 200 kW total heat load data center,

$$N = 200\,\text{kW}/30.3\,\text{kW} = 6.6;\ N = 7 \text{ cooling units}$$

In a 200 kW total heat load data center with DX cooling units and 35.2 kW per unit,

$$N = 200\,\text{kW} / 21.7\,\text{kW} = 5.68;\ N = 6\,\text{cooling units}$$

So the Data Center design would require (1) fewer DX cooling units with a 90°F RAT compared to 80°F RAT.

One can conclude from this that installing an airflow architecture that uses either HAC or CAC to increase RAT can have significant impact on the cooling unit coil capacity, efficiency, and total number of cooling units required to manage the total data center heat load.

* 1 ton=12 MBH = 12,000 BTU h; 1 BTU h= 3.412 W

TABLE 24.2 Liebert CRV air-cooled data[a]

	Cond. temp. 120°F (48.9°C)	
	CR035RA	CR020RA
105°FB DB, 71°F WB (40.6°C DB, 21.6°C WB) 17% RH		
Total BTU/h (kW)	137,885 (40.4)	83,960 (24.6)
Sensible BTU/h (kW)	137,885 (40.4)	83,960 (24.6)
100°F DB, 69.5°F WB (37.8°C DB, 20.8°C WB) 20% RH		
Total BTU/h (kW)	131,401 (38.5)	79,864 (23.4)
Sensible BTU/h	131,401 (38.5)	79,864 (23.4)
95°F DB, 67.9°F WB (35°C DB, 19.9°C WB) 23% RH		
Total BTU/h (kW)	125,257 (36.7)	76,110 (22.3)
Sensible BTU/h (kW)	125,257 (36.7)	76,110 (22.3)
90°F DB, 66.2°F WB (32.2°C DB, 19.0°C WB) 27% RH		
Total BTU/h (kW)	120,138 (35.2)	72,356 (21.2)
Sensible BTU/h (kW)	120,138 (35.2)	72,356 (21.2)
85°F DB, 64.5°F WB (29.4°C DB, 18.1°C WB) 31% RH		
Total BTU/h (kW)	117,066 (34.3)	68,601 (20.1)
Sensible BTU/h (kW)	113,994 (33.4)	68,601 (20.1)
80°F DB, 62.8°F WB (26.7°C DB, 17.1°C WB) 37% RH		
Total BTU/h	113,994 (33.4)	67,919 (19.9)
Sensible BTU/h (kW)	103,414 (30.3)	67,919 (19.9)
80°F DB, 66.5°F WB (26.7°C DB, 19.2°C WB) 50% RH		
Total BTU/h (kW)	121,503 (35.6)	72,697 (21.3)
Sensible BTU/h	88,738 (26)	59,045 (17.3)

[a]Courtesy of Emerson Electric Corp.
Note: Unit net sensible capacity with 80°F DB/62.8°F WB RAT is 30.3 kW (103.47 MBH, 8.62 tons).
Unit Net Sensible Capacity with 85°F DB/64.5°F WB RAT is 33.4 kW (114.06 MBH, 9.5 tons).
Unit Net Sensible Capacity with 90°F DB/66.2°F WB RAT is 35.2 kW (120.2 MBH, 10.0 tons).

24.4 RUN- OR RIDE-THROUGH IMPACT OF HIGHER RAT

In DX and CW cooling units, power loss results in an immediate loss of fans and cooling. In CW cooling units, loss of waterflow, due to pump system failure, can result in less immediate loss of cooling capacity, as fans would continue to run if power is available.

With CAC, consider the availability of cool supply air during a loss of power/cooling. Containing the cold aisle *minimizes* the overall pool or reservoir of cool or cold aisle air available to the IT equipment inlets, in the event of power loss or cooling failure. A small portion of the room air volume is at the correct IT equipment inlet air temperature, but most of the room volume of air is at the incorrect temperature.

With HAC, containing the hot aisle *maximizes* the overall pool or reservoir of cool air available to the IT equipment inlets in the event of power loss or cooling failure. A small portion of the room air volume is at the incorrect IT equipment inlet air temperature, but most of the room volume of air is at the correct temperature.

Some examples (based on estimates, not testing; actual computational fluid dynamics (CFD) model is recommended for any specific data center) are discussed in the following.

24.4.1 Low-Density CAC versus HAC Ride-Through

Example 1: Low density, 2500 ft² room

Consider an example data center with 200 kW total heat load, 50 ft × 50 ft white space, 14 ft high ceilings, and (2) cold aisle zones, each with two rows of 20 racks per row. Average IT rack heat load density = 200 kW/80 racks = 2.5 kW per rack.

At 160 CFM per kW of IT load, airflow through the IT equipment = 200 kW × 160 CFM per kW = 32,000 CFM IT equipment total airflow or 16,000 CFM per zone:

$$\text{Room total volume} = 50\,\text{ft} \times 50\,\text{ft} \times 14\,\text{ft} = 35{,}000\,\text{cu ft}$$

$$\begin{aligned}\text{Room IT air exchange rate} &= 35{,}000\,\text{cu ft}/32{,}000\,\text{CFM} \\ &= 1.09\,\text{min}.\end{aligned}$$

For IT equipment with 160 CFM/kW airflow, air temperature rise across the IT equipment = 19.7°F ΔT (3415 BTU/h per kW/160 CFM/1.08 Air Constant).

Another way to state this is as follows: every 1.09 min, the entire room volume of air passes through the IT equipment and experiences the IT equipment temperature increase ΔT, so that when cooling is absent, in 1.09 min the room air temperature will rise (+)19.7°F.

24.4.1.1 CAC: Low Density The CAC volume = 40 ft (rows length) × 4 ft (containment aisle width) × 7 ft (containment rack height typical for 42U racks) = 1120 ft^3 × 2 CACS zones = total 2240 ft^3 in CAC zones.

The room/hot aisle temperature would be a maximum of Rack Inlet Air Temperature (RIAT) plus IT equipment ΔT: with a RIAT = 75°F, room/hot aisle temperature = 94.7°F.

CAC volume as compared to IT equipment airflow = 2,240 ft^3/32,000 CFM = 0.07 min, such that in ~4.2 s, the entire volume of CACS air at 75°F temperature will be drawn into the IT equipment and increase in temperature from 75 to 94.7°F, joining the rest of the 94.7°F air in the room/hot aisles. In the next 4.2 s with no cooling, RIAT will be 94.7°F in the cold aisles at the IT equipment air inlets.

Even without power to the cooling system unit fans, warm room/hot aisle air will continue to be drawn into the CACS volume at the constant flow rate as needed to match total IT equipment airflow. Airflow follows the path of least resistance, driven by differential pressure as it moves from higher pressure zones to lower pressure zones. No currently available containment system is completely airtight or leak-free. There are always open containment spaces, such as between server inlet blanking panels, rack cable access openings, space between the IT equipment racks and each other, and between the racks and the floor. In the case of internal cooling units, collocated in the CAC zone with the racks, containment leakage and airflow paths through the cooling units with static fans will permit the required amount of IT equipment airflow to pass from the room/hot aisle into the containment system.

In the case of external cooling units and fans outside the CAC Zone with the fans not operating, the airflow to match IT equipment inlet air will come either from the room/hot aisle via the leakage in the containment system or from the route of supply and return airflow between the CAC zone and the static external cooling unit(s). Some equilibrium of total airflow balanced between room leakage and external routes will develop, at the same pressure differential between both systems with respect to the IT equipment inlets at the CAC Zone.

In either case, the IT equipment fans will continue to draw the required 32,000 CFM of airflow from their inlet sources.

With no further cooling, in another 1.02 min, for the entire room volume, including cold aisles, air temperature will rise another 19.7°F, to reach 114.4°F, and, in the next 1.09 min, further potential increase up to 134.1°F room temperature, at which time the fire sensors may start to engage. This is a room temperature rise of 18°F increase for every minute of lost cooling.

Starting at a room/hot aisle temperature of 94.7°F, this causes a room temperature rise to an unacceptable 120°F in ~1.4 min or to 140°F in 2.5 min.

For external cooling systems, Rack inlet temperature rise per measure of time would be adjusted to account for the additional "room volume" as allocated to the volume of the supply and return air duct systems. Hot air volume would initially be counted on the return duct side and cool air on the supply duct side.

24.4.1.2 HAC: Low Density As in the same room example, the HAC volume = 40 ft (rows length) × 4 ft (containment aisle width) × 7 ft (containment rack height typical for 42U racks) = 1120 ft^3, × 2 HACS zones = total 2240 ft^3 in HAC zones.

The hot aisle temperature would be a maximum of RIAT plus IT equipment ΔT: with an RIAT = 75°F, hot aisle temperature = 94.7°F.

The room/cold aisle temperature would be the same as RIAT: with an RIAT = 75°F, room/cold aisle temperature = 75°F.

HAC containment volume as compared to IT equipment airflow = 2,240 ft^3/32,000 CFM = 0.07 min, such that in ~4.2 s, the entire volume of HACS air at 94.7°F temperature will be drawn into the room through nonfunctioning cooling units or containment leakage and join the rest of the 75°F air in the room/cold aisles. After these 4.2 s, the blended room/cold aisle temperature can be calculated as follows:

$$\{(94.7°\text{F} \times 2,240\,\text{cu ft}) + [(35,000\,\text{cu ft} - 2,240\,\text{cu ft}) \times 75°\text{F}$$
$$35,000\,\text{cu ft.}$$

This is equal to 76.3°F blended room air temperature after 4.2 s of cooling loss.

Even without power to the cooling system unit fans, cool/room–cold aisle air will continue to be moved into the HACS volume at the constant flow rate to match total IT equipment airflow. In the case of internal cooling units, located in the HAC zone with the racks, containment leakage and airflow paths through the cooling units with static fans will permit the required amount of IT equipment airflow to pass from the HACS volume out into the room/cold aisle.

In the case of external cooling units and fans, outside the HAC Zone, with the fans not operating, the airflow to match IT equipment inlet air will come either from the room/cold aisle via the leakage through the containment system or from the route of supply and return airflow between the HAC Zone and the static external cooling unit(s). Some equilibrium of total airflow balanced between room leakage and external routes will develop, at the same pressure differential between both systems with respect to the IT equipment outlets at the Hot Aisle Zone.

In either case, the IT equipment fans will continue to draw the required 32,000 CFM of airflow from their inlet sources and exhaust it to the HACS volume.

With no further cooling, in the next 1.05 min, the entire room volume air temperature will rise another 19.7°F, to reach 96°F, and, in another 1.09 min, further potential increase up to 115.7°F room temperature. This is a room temperature rise of 18°F increase for every minute of lost cooling.

Starting at a room/cold aisle temperature of 75°F, this causes a room temperature rise to an unacceptable 120°F in approx. 2.5 min or 140°F in 3.6 min.

For external cooling systems, Rack Inlet temperature rise per measure of time would be adjusted to account for the additional "room volume" as allocated to the volume of the supply and return air duct systems. Hot air volume would initially be counted on the return duct side, and cool air on the supply duct side.

24.4.1.3 Low-Density CAC versus HAC Ride-Through Summary

For CAC, RIAT = 94.7°F after 8.4 s of cooling loss.

For HAC, RIAT = 94.7°F after 1.09 min of cooling loss.

Considering the ride-through time from loss of cooling power to 120°F room temperature in the data center, compare 1.4 min for CAC with 2.5 min for HAC. From loss of cooling power to 140°F room temperature in the data center, 2.51 min is for CAC and 3.6 min for HAC.

24.4.2 Medium-Density CAC versus HAC Ride-Through

Example 2: Medium density, 2500 ft^2 room

Consider an example data center with 800 kW total heat load, 50 ft × 50 ft white space, 14 ft high ceilings, and (2) cold aisle zones, each with 2 rows of 20 racks per row. Average heat load density per rack is 800/80 = 10 kW per rack.

At 135 CFM per kW of IT Load, airflow through the IT equipment = 800 kW × 135 CFM per kW = 108,000 CFM IT equipment total airflow or 54,000 CFM per zone:

$$\text{Room total volume} = 50\,\text{ft} \times 50\,\text{ft} \times 14\,\text{ft} = 35{,}000\,\text{cu ft}.$$

$$\begin{aligned}\text{Room IT air exchange rate} &= 35{,}000\,\text{cu ft}/108{,}000\,\text{CFM}\\ &= 0.324\,\text{min}(19.4\,\text{s}).\end{aligned}$$

For IT equipment with 135 CFM/kW airflow = 23.4°F ΔT (3415 BTU/h/kW/135 CFM/1.08 Air Constant).

In other words, every 19.4 s, the entire room volume of air passes through the IT equipment and experiences the IT equipment temperature increase ΔT, so that when cooling is absent, in 19.4 s the room air temperature will rise 23.4°F.

Following the same type of calculations as for the low-density example, starting at a room/cold aisle temperature of 75°F, this causes a room temperature rise to an unacceptable 120°F in ~37 s or to 140°F in 54 s.

24.4.2.1 Medium-Density CAC versus HAC Ride-Through Summary

For CAC, RIAT = 98.4°F after 1.2 s of cooling loss.

For HAC, RIAT = 98.4°F after 19.4 s of cooling loss.

Considering the ride-through time from loss of cooling power to 120°F room temperature in the data center, compare 18 s for CAC with 37 s for HAC. From loss of cooling power to 140°F room temperature in the data center, 34.7 is for CAC and 54 s for HAC.

24.5 SINGLE-GEOMETRY PASSIVE CHIMNEY DUCTS AS PART OF HAC

24.5.1 Pressure Drop Dependency

Single-geometry passive chimney exhaust ducts from IT cabinets with no other means of airflow regulation are dependent primarily on different pressure drops across the exhaust duct for different airflow rates. Any (2) or more IT cabinets with the same chimney duct geometry (rack access opening, cross-sectional area, and height) will have approximately the same airflow regardless of rack kilowatt IT load or IT equipment airflow, since the main force moving the air out of the cabinet is the pressure difference between the bottom of the cabinet chimney and the negative air pressure in the ceiling plenum or return duct above the chimney. Passive rack designs that claim to contain the exhaust air from the server fans and move it ***only*** into the chimney duct are in most cases not airtight "balloons," since they offer many exit paths for airflow out of the racks other than the ducted chimney or other directed exit path.

Consider these potential leakage openings in a typical rack, for example: the rack bottom and space between the lowest server and the floor, the cable access ports, rack accessory mounting holes, the small spaces between the front blanking panels, rack door interfaces, and sheet metal junctures.

Air as a fluid medium is capable of CFM movement through small areas with small differential pressure across the area. In common terms, attempting to push air without complete containment (ductwork or racks with no holes) is comparable to carrying water in a bucket full of holes. To illustrate, 300 CFM of airflow can move through 10 ft of a 10 in. diameter round duct with only 0.005 in. WC differential pressure.

Air movement can be more effectively influenced by creation of negative pressure from the desired direction of flow, such as a blower fan inlet at the end of a duct. This is particularly applicable when standard server fans with low capacity for external static pressure (ESP) are attempting to move air on the positive pressure side of the fan into an airflow resistance area. Server fans by design have been selected for the ability to move the required airflow across the heat sinks in the servers with enough pressure differential on the upstream-negative side of the fans, to overcome only the internal resistance of the airflow path through the server. It would be determined on a case-by-case basis if any specific server fan had sufficient "extra" pressure capacity on the downstream-positive side of the fan to provide sufficient server airflow internally in the server airflow paths.

24.5.2 Server Fan Issues

Depending on the server fans to exhaust the air from the servers out, the return duct system presents other issues:

1. Server fans have a typical "steep" Airflow-versus-Pressure performance curve, so that any slight increase in static discharge or inlet pressure drop causes a significant decrease in airflow.
2. Server fans are selected and designed for "draw through" airflow, pulling air from the front of the server across the heat sinks and out the back, with very small design pressure drops across the system. Using this type of fans for pushing air against an ESP is like trying to push a rope. Any buildup of pressure will cause reduced airflow, and the air will find the path of least resistance, which may be out into the room and not up the exhaust chimney.

For these reasons, the flow of air through any passive ducted system is more dependent on the airflow generated by the external fan system, not the IT Server fans. In a typical data center, this is the airflow provided by the perimeter computer room air-conditioning (CRAC) A/C units or any other fan-coil cooling unit dedicated to the data center space.

24.5.3 Mixed IT Load Examples

If the pressure difference for a single-geometry cabinet chimney duct resulted in 600 CFM per chimney, the same 600 CFM would be the airflow in a 3 kW IT load cabinet and in a 12 kW IT load cabinet.

However, required server airflow in a 3 kW cabinet is 480 CFM, and Server Airflow in a 12 kW cabinet is 1920 CFM (IT airflow = kW × 160 CFM for 19.7°F temp rise across the servers). Airflow (based on pressure drop between the ceiling plenum pressure and the cabinet pressure) for all racks in the system would have to be adjusted to 1920 CFM per each cabinet, in order to capture the airflow from any one of the cabinets with the highest IT load. If there are 20 × 3 kW cabinets and 10 × 12 kW cabinets, the total IT airflow would be 9,600 CFM + 19,200 CFM = 28,800 CFM.

However, because of the single-geometry chimney duct, all 30 cabinets would need to move 1,920 CFM × (30) = 57,600 CFM. This is a 100% increase in airflow, with the appropriate kilowatt power cost to run the A/C fan motors at this airflow.

In addition, if the 1920 CFM per 12 kW cabinet is not captured and exhausted through the chimney duct, the excess airflow from the servers would leak or spill out into the room, causing the potential for recirculation and hot spots. If the system A/C units are designed for the actual IT airflow requirement, 28,800 IT airflow (30) cabinet system, the average airflow per chimney duct would be 960 CFM (28,800/30). Under this condition, for each of the 10 12 kW cabinets, (1920 − 960) = 960 CFM of 90°F IT exhaust air per cabinet (total 9600 CFM) would not be drafted up through the return duct and would be available to recirculate and cause hot spots.

Consider an example scenario: a large data center room with Compute Rows J and M, consisting of 40 cabinets at 18 kW per rack. Total IT Airflow = 115,200 CFM. If all cabinets are the same IT airflow, this is 2880 CFM per cabinet chimney duct average.

In this example, consider the same room also with Infrastructure Rows C and F, consisting of 32 cabinets at 18 kW per rack. Total airflow = 92,160 CFM. If all cabinets are the same IT airflow, this is 2880 CFM per cabinet chimney duct average.

Since the cabinets are all 18 kW, this is okay and the total 207,360 CFM is the required A/C unit airflow in the ducted return system.

If one of the cabinets is changed to 25 kW IT Load, the new cabinet airflow requirement is 4000 CFM if the 160 CFM per kW applies or, if blade servers, maybe 130 CFM per kW = 3250 CFM. Since only 2880 CFM is being captured and collected by the return air system, 370 CFM per cabinet (Blade Servers) of warm IT airflow (maybe 94°F) would be available to recirculate and cause hot spots. If 10 of the 72 cabinets are upgraded to 25 kW IT Blade Server Load, then 3700 CFM of warm air would be available to recirculate and cause hot spots.

To correct for this, each of the 72 cabinets would require the new airflow of 3,250 CFM, for a total of 234,000 CFM required from the group of A/C units.

Again, instead of 62 cabinet ducts at 2,880 CFM (178,560 CFM) plus 10 cabinet ducts at 3,250 CFM (32,500), the IT airflow total is 211,060 CFM, but the A/C units group airflow is 234,000; this is an 11% penalty on Blower HP.

24.5.4 Solution: HACS with Baffle Ducts

Using the HACS plenum, for air containment in a common hot aisle between the rows of cabinets, allows aggregation of all the cabinet airflows into a shared/common return duct system, which can be adjusted for the HACS zone IT airflow average, regardless of individual cabinet IT airflow. So from the earlier example, 20 × 3 kW cabinets and (10) × 12 kW cabinets, 28,800 CFM is the airflow needed from the A/C blower units, not 57,600 CFM.

Since the Hot Aisle Zone serves as a buffer space for the IT Airflow to occupy before the total CFM is exhausted through

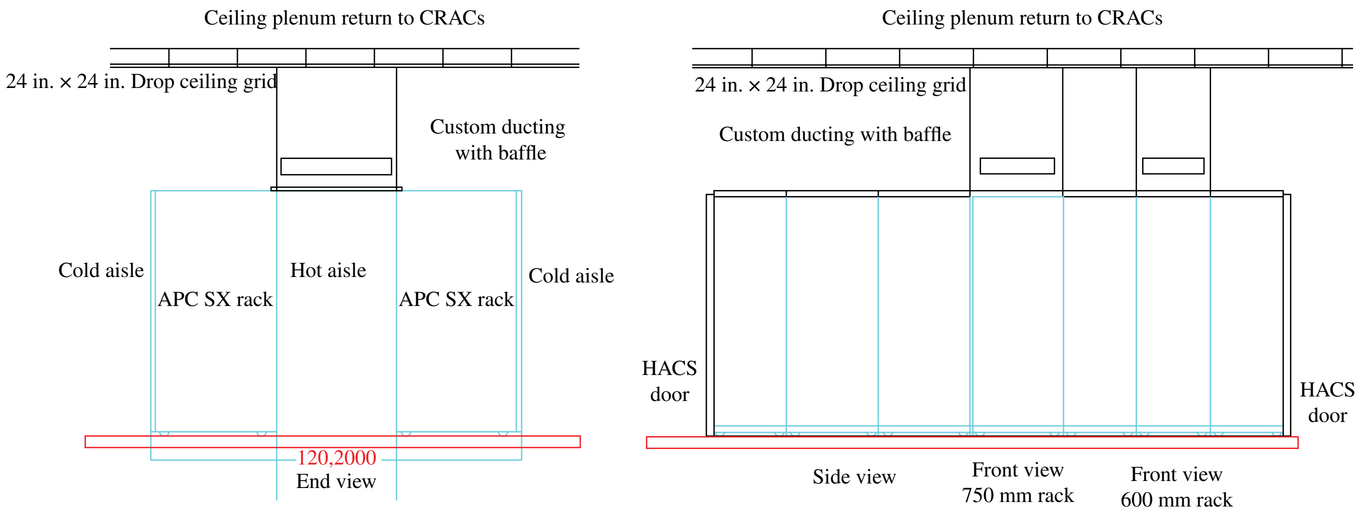


FIGURE 24.7 Baffle duct kits for HACS. Courtesy of Schneider Electric ITB.

the common Duct Kits, individual cabinet airflows are determined by the servers, not by the duct system (Fig. 24.7). With equal amounts of airflow coming into the hot aisle zone from the server outlets and being drawn out of the hot aisle zone by the A/C unit fans, there is no static pressure buildup in the hot aisle zone to affect the server fan airflows.

Adjustable baffles can be provided in the Duct Kits, to balance the airflows between multiple HACS Zones in the data center served by the same group of A/C units and ceiling plenum. The use of a "Hot Tip" anemometer inserted into the return duct kit is one way to manually determine the baffle adjustment needed for the required airflow per zone. This also allows adjustment to compensate for different negative pressures around the ceiling plenum space above the hot aisles.

In the example, the Infrastructure (32) rack hot aisle zone duct kit system will be adjusted for 92,160 CFM total per zone, and the Compute (40) racks hot aisle zone duct kit system will be adjusted for 115,200 CFM total per zone.

The HAC with ducted return is considered a **Passive** system since it contains no intermediate fans or air movers between the rack outlet and the A/C units. For discussion of **Active** ducted return systems, see Schneider Electric ITB Application Note #96, Rack Air Removal Unit.

24.6 PSYCHOLOGICAL IMPACTS OF HIGHER RAT

24.6.1 Maximum Room Temperature Comfort Level

Some industry documents claim that "…CAC yields high (return) air temperature to the cooling unit to increase capacity and efficiency." However, there may be a psychological penalty and compromise associated with higher RAT using CAC, as compared to HAC.

Consider a CACS where the small volume of air in the containment system is at RIAT = 75°F. With standard IT equipment servers, temperature rise across the IT equipment is ~ (+) 19°F ΔT. This means that the larger volume of room air/Hot Aisle temperature is ~75°F + 19°F = 94°F. This room temperature exceeds the recommended maximum ASHRAE TC9.9 RIAT of 80°F by 14°F.

This component and result of an otherwise efficient CACS may not be acceptable as a comfortable Room Air temperature for some Data Center managers. If this is the case, room temperatures would be required to remain by design at some value less than the 80°F ASHRAE standard maximum. This would limit the Cooling unit RAT to 80°F and affect the maximum available cooling unit capacity and efficiency at that RAT. To achieve a Room and RAT less than 80°F, adjustment of the cooling system set points would be required to account for the maximum RAT. Considering a 19°F ΔT across the IT equipment, to achieve an 80°F maximum RAT would require (80 − 19) = 61°F Cooling unit supply/cold aisle temperature. This 61°F Cold Aisle/RIAT is less than the ASHRAE standard minimum recommended 64°F RIAT.

So in order to maintain the RIAT and cooling unit SAT within the minimum ASHRAE limits, a RAT/Room/Hot Aisle temperature of 64°F + 19°F = 83°F is required.

Appropriate consultation with the data center management and engineering design team is recommended to resolve these competing design paths: either higher Room/Hot Aisle/RAT that provide more cooling unit capacity or management comfort level with room air temperatures above the maximum ASHRAE-recommended limit or management comfort level with RIAT lower than the minimum ASHRAE-recommended limit.

By comparison, consider an HACS where the small volume of air in the containment system hot aisle is at

RAT = RIAT plus IT equipment ΔT. With standard IT equipment servers, temperature rise across the IT equipment is ~ (+) 19°F ΔT. This means that the smaller contained volume of Hot Aisle temperature is ~75°F + 19°F = 94°F. With maximum recommended ASHRAE RIAT of 80°F, the hot aisle temperature may be at 99°F, resulting in even more cooling unit capacity and increased efficiency.

This component and result of an efficient HACS is likely to be acceptable as a RAT/Hot Aisle temperature for most Data Center managers. If this is the case, Room/Cold Aisle temperatures would remain by design at some value less than the ASHRAE maximum standard 80°F. Cooling unit and cold aisle supply temperatures would be consistent with recommended ASHRAE RIAT (64–80°F) and not lower than minimum recommended of 64°F.

24.6.2 Personnel Working Conditions in HAC versus CAC

The use of modern enclosed hot aisles to address increasing power densities and more efficient cooling systems in the data center has brought into question the suitability of working conditions in these hot aisle environments.

Traditional approaches to cooling a data center use a loosely coupled method of conveying air to and from the heat load. This means that cooling system supply air is widely dispersed within the room, typically via raised floor. There are no rigid boundaries to prevent cool air from mixing freely with air throughout the entire volume of the room. This is done as an attempt to hold the average bulk temperature of the room down to a level acceptable to IT equipment and personnel. The return airstream to the cooling system in such a room is representative of bulk room conditions, usually around 75°F.

If a hot aisle/cold aisle layout is used in the room, temperatures observed should be higher than 75°F in the hot aisle and lower than 75°F in the cold aisle. In practice, these deviations from the overall average room temperature are almost never as large as desired. This is because traditional designs often allow a large portion of cool supply air to bypass its intended delivery points and directly enter the hot aisle, usually through cable access penetrations in the floor. This lowers the hot aisle temperature. Similarly, historical facilities generally contain no provisions to prevent hot IT equipment exhaust air from being drawn over the tops of racks or around the ends of rows into the cold aisle. The cool air leakage serves to lessen the heat stress experienced by an individual working in the hot aisle, and the hot air leakage might warm an uncomfortable cold aisle.

A conservative approach and set of assumptions can be made and comparisons drawn between worker heat stress in a relatively open legacy data center and a modern compacted and zero-recirculation data center.

Heat and mass transfer analysis helps to illustrate the potential effects of recirculation on worker heat stress. By enforcing conservation of mass and energy, it is shown that for a worker standing in the hot aisle, heat stress could either *decrease* or *increase* as recirculation is reduced or eliminated. The actual result depends on where the worker is standing in relation to the onset of air mixing. If the worker positions himself such that he is mostly exposed to unmixed server exhaust air, he will be exposed to higher temperatures when recirculation is occurring than when it is not. This would likely be the case for someone working in very close proximity to the IT equipment exhaust vents on servers. However, a worker whose job involved moving around in the general volume bounded by the open hot aisle would likely experience lower temperatures in a data center where recirculation is allowed to occur. This is because the IT equipment exhaust air to which this worker is exposed has had more of an opportunity to be tempered by cool supply air leaking from the cold aisle into the hot aisle.

Most temperature sensors in data centers are dry-bulb temperature sensors. However, the dry-bulb temperature alone is a poor measure of worker heat stress because it does not take into account the physiological effect of humidity levels or the presence of radiant heat. A much better measure of heat stress is Wet-Bulb Globe Temperature, or WBGT. WBGT is different from a normal thermometer reading in that it takes into consideration air temperature, humidity, and radiant heat. Each of these factors can contribute to heat stress experienced by a worker.

Guidance for maximum WBGT exposure for a variety of conditions is provided in the *Occupational Safety and Health Administration (OSHA) Technical Manual* Section III: Chapter 4.

Equation (24.1) is for WBGT when no solar radiation is present, as is assumed to be the case in data centers:

$$\text{WBGT} = .7\,\text{NWB} + .3\,\text{GT} \qquad (24.1)$$

where

NWB is the natural wet-bulb temperature.
GT is the globe temperature.

The natural wet-bulb (NWB) temperature is a function of both the dry-bulb temperature and relative humidity in a room. It is measured by placing a water-soaked wick-type material over the bulb of a normal mercury thermometer. The latent heat removed from the thermometer bulb by evaporation of the water reduces the temperature relative to dry bulb and is a direct representation of the ease with which a worker can dissipate heat by sweating. A psychrometric chart can also be used to indicate wet-bulb temperature if relative humidity is known. It can be assumed that both data centers are controlled to 50% RH. The Globe Temperature (GT) is defined as the reading of a temperature sensor in the center of a thin-walled

TABLE 24.3 OSHA guidelines: Permissible heat exposure threshold limit value

Work/rest regimen (each hour)	Light work load max permissible WBGT	Moderate work load max permissible WBGT	Heavy work load max permissible WBGT
Continuous work	86°F (30°C)	80°F (26.7°C)	77°F (25°C)
75% work/25% rest	87°F (30.6°C)	82°F (28.0°C)	78°F (25.9°C)
50% work/50% rest	89°F (31.4°C)	85°F (29.4°C)	82°F (27.9°C)
25% work/75% rest	90°F (32.2°C)	88°F (31.1°C)	86°F (30°C)

TABLE 24.4 WBGT calculations room and HACS[a]

	Room values	HACS zone values
DB temp (°F)	75	95
WB temp (°F)	62.7	69
RH (%)	50	27
Calculated WBGT	66.4	76.8

[a] Courtesy of Schneider Electric.

blackened copper sphere. Both radiant heat and ambient dry-bulb temperature contribute to this reading. The level of radiant heat absorbed by an individual standing in the hot aisle is negligible because visible solid surfaces are not substantially hotter than the worker's body temperature. Because of this, the dry-bulb temperature from the analysis can be used in place of GT without compromising accuracy. NWB and GT are then summed using Equation (24.1) to arrive at WBGT.

Table 24.3 represents current OSHA Guidelines for work load levels with respect to associated WBGT Values.

The OSHA description of "light work load" is thought to most nearly match the work of a typical IT worker task such as installing rack-mounted equipment or routing network cables, and the worker is assumed to be dressed in normal single-layer clothes. For comparison, OSHA defines "heavy work load" as heavy full-body work and cites "laying railroad ties" as an example of work fitting this category.

In an HACS, the IT equipment exhaust air temperature represents the rise in dry-bulb temperature added to RIAT, with IT load contributing no change in moisture levels in the airflow. With a typical RIAT of 75 and 19.7°F ΔT for IT equipment, HACS dry-bulb temperature may be as high as 95°F.

The following are the results of WBGT Calculations (Table 24.4) for the comparison of Room and Hot Aisle conditions in a typical data center.

This demonstrates that with Hot Aisle Zone Dry-Bulb temperatures of 95°F, working conditions per OSHA are still less than the WBGT permitting heavy work load and certainly 10°F less than that WBGT value for normal data center working loads.

24.6.3 Adjustments for Hot Aisle Temperatures

Notwithstanding this regulatory analysis, personnel may still object or raise concerns about the perceived conditions in the hot aisle zone. A reasonable data center manager would be responsive to these concerns, by taking some action to address the issue, as follows.

24.6.3.1 CAC Temperature Adjustments With CAC, when room/hot aisle temperature and humidity conditions are determined to be too high for working personnel comfort, set points for cooling unit supply temperatures may be adjusted lower, to result in lower room/hot aisle temperatures with the same IT equipment ΔT. This may temporarily reduce cooling unit capacity and so must be planned for ahead of actual commissioning. Lower room/hot aisle temperatures may also impact RAT and cooling system efficiency, and the power increase for the same required cooling capacity must be absorbed by the data center electrical system and operations budget.

As a partially effective alternate, set points for cooling unit airflow may be adjusted to achieve more than the required airflow equal to zone IT equipment airflow. In a typical CACS cooling plan, $N+1$ cooling units per CAC zone are deployed, resulting in lower than 100% fan speed for all units operating at the same time. For example, if $N=5$ units, $N+1=6$ units; 100% airflow at maximum fan speed for five units = 83% of fan speed airflow for six units; therefore, there is an additional 20% of N airflow available if all six units are operating at 100% fan speed.

This additional 20% of N Cooling airflow would be supplied to the CAC zone, not drawn out and heated by the IT equipment, and instead leak into the room/hot aisle volume at the cooling unit SAT. The effect of this additional cooling airflow on the room/hot aisle temperature can be calculated, and the impact on cooling unit performance determined. In summary, the effect of 20% more cooled supply air on a large room/hot aisle volume of air will be less than the effect of 20% more cooled air on a smaller hot aisle volume of air.

24.6.3.2 HAC Temperature Adjustments With HAC, when hot aisle temperature and humidity conditions are determined to be too high for working personnel comfort, set points that control cooling unit airflow may be adjusted for higher airflow. In a typical HACS cooling plan, $N+1$ cooling units per HAC zone are deployed, resulting in less than 100% fan speed for all units operating at the same time. As with CAC, if $N=5$, there is an additional 20% of N airflow available if all six units are operating at 100% fan speed.

This additional 20% of N cooling airflow would be drawn out of the HAC zone and replaced by air leakage into the hot aisle volume at the room air temperature and RIAT. For best psychological impact on working personnel, HAC zone doors can be opened to facilitate this leakage of cooler air into the hot aisle zone volume. The effect of this additional cool/room temperature airflow on the hot aisle temperature can be calculated, and the impact on cooling unit performance determined.

For example, if IT equipment total airflow in a 95°F HAC zone = 15,000 CFM and an additional 20% of N airflow is added at room/cold aisle temperature, with addition of 3,000 CFM at 75°F, the net effect would be a hot aisle temperature of 91.7°F instead of 95°F with N airflow. Based on the reaction and control logic for the cooling unit SAT, this additional airflow should have minimal effect on any change in RIAT.

If there is no control logic for control of the cooling unit SAT, lowering the RAT set point may result in lowering the RIAT. These lower SAT and resulting RIAT can produce thermal shock on the IT equipment. Therefore, this must be done gradually, within the not more than 1°F per 6.7 min (not to exceed 9°F ΔT in 1 h) rate as recommended for RIAT rate of change in ASHRAE TC9.9.

Higher airflow and fan speeds may also impact cooling system efficiency, and the power increase for the same required cooling capacity must be absorbed by the data center electrical system and operations budget.

In summary, the effect of 20% more cooled supply air on a small hot aisle volume of air will be greater than the effect of 20% more cooled air on a larger room/hot aisle volume of air. Based on this, lower cooling unit fan speeds to achieve the same reduction in hot aisle temperature would be expected with HAC zones.

With HAC, as an alternate to airflow adjustment, set points for cooling unit SAT may be adjusted lower, to result in lower room/cold aisle temperatures with the appropriate result in lower room/hot aisle temperatures with the same IT equipment ΔT. This may temporarily reduce cooling unit capacity and so must be planned for ahead of actual commissioning. Lower room/hot aisle temperatures may also impact cooling system efficiency, and the power increase for the same required cooling capacity must be absorbed by the data center electrical system and operations budget.

Note: There will be an impact RIAT from the lower SAT set point method.

With CAC or HAC, lowering the SAT can produce thermal shock on the IT equipment and therefore must be done gradually, within the not more than 1°F per 6.7 min (not to exceed 9°F ΔT in 1 h) rate as recommended for RIAT rate of change in ASHRAE TC9.9. So a decrease of 1°F in SAT set point should be allowed to operate for approx. 10 min before the next 1°F set point change, with a net change of 5°F change accomplished over a period of ~1 h.

24.7 COOLING SYSTEM AIRFLOW AND FAN POWER

24.7.1 Room-Based Cooling Systems

Traditional data center cooling systems were based on a room cooling unit(s) and a mixing model for obtaining heat removal and some ability to achieve IT equipment RAITs within the range recommended by ASHRAE standard TC9.9 (Fig. 24.8).

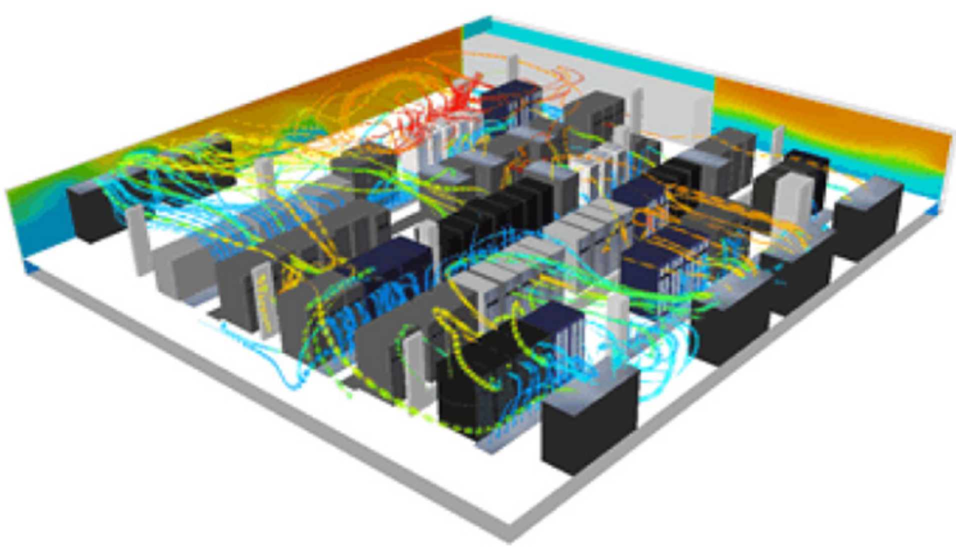


FIGURE 24.8 Room-based cooling system. Courtesy of Schneider Electric.

This design approach appears to have been sufficient for low- and medium-density data center rack heat loads (up to 5 kW per rack), equivalent to ~170 W/ft^2 of data center space, and larger available white space to distribute the mixed air from the IT equipment exhaust locations to the hot aisles, to the room, and eventually to the return/inlet of the cooling unit(s).

Cooling unit air supply temperatures were lower than the design average RIAT, and cooling unit total airflow (CFM/kW) was typically less than IT equipment total airflow, such that the mixing of cooling supply air with some room air and some IT equipment exhaust air from the hot aisles recirculated into the cold aisles was sufficient to blend into an acceptable RIAT at some or most of the rack IT equipment locations. Also consistent with this approach and model was the likelihood of localized RIAT above or below the recommended range, with regard to individual IT equipment mounted in any particular U Space in any given rack. This was particularly true in a raised-floor supply air delivery system, where supply air/cold aisle temperatures at the lower U spaces were 60–70°F and the supply air/cold aisle temperatures at the upper rack U spaces were 80–90°F.

With flooded room return air system to the cooling units, cooling system fan power was in direct proportion to the airflow rate CFM and total pressure drop in the system. Pressure drop consists of ESP drop, in particular the pressure below the raised floor, and the internal cooling unit pressure drop across the coil and filter(s). There is also a velocity pressure drop as the airflow changes directions to disperse under the raised floor.

In overhead supply air systems locally ducted from upflow perimeter cooling units to the cold aisles, the external pressure drop is dependent on the flow and geometry of the supply duct system.

In some newer room cooling systems, supplying sufficient IT inlet air at the cold aisles is one of the design considerations. In this case, supply air out of the perforated tiles in the cold aisle is designed to be sufficient to match the required IT inlet airflow for that row of rack-based IT equipment. Floor tile manufacturers publish CFM flow rates per tile based on the underfloor static pressure and can provide detailed professional analysis and recommendations.

In an example of 25% open space floor tile, with no damper, an underfloor static pressure of 0.10 in WG is sufficient to provide ~800 CFM of supply airflow per tile. In a typical cold aisle design, a 20 ft row length of 10 tiles and 10 racks under this condition could provide 8000 CFM supply airflow to the inlet of the rack-mounted IT equipment. For IT equipment with 160 CFM required, and a corresponding 19.7°F ΔT air temperature increase across the IT equipment these 10 floor tiles could support airflow for 8000 CFM/160 CFM/kW = 50 kW of IT equipment load. This equates to an average of 5 kW per rack.

In this model, floor leakage outside the cold aisle tiles must be taken into account. Typical floor leakage rate for 0.10 in WG Static pressure is 10–15% of total airflow. So in this example, between 8800 and 9200 CFM must be delivered from the supply air system to account for leakage and still provide the required airflow in the cold aisle tiles. Since fan or blower power is directly proportional to airflow CFM, this represents an additional 10–15% more fan power than other nonleaking supply air systems.

One potential shortcoming of room-oriented architecture is that in many cases the full rated capacity of the CRAC units cannot be utilized. This condition is a result of room design and occurs when a significant fraction of the air distribution pathways from the CRAC units bypasses the IT loads and returns directly to the CRAC. This bypass air represents CRAC airflow that is not assisting directly with cooling of the IT equipment heat loads, in essence a decrease in overall cooling capacity. The result is that cooling requirements of the IT layout can exceed the cooling capacity of the CRAC units even, while additional bulk cooling (kW) capacity of the CRAC is not fully utilized.

Another design or operating defect of room cooling systems may occur that is known as "stranded capacity," which is present when the CRAC or external cooling capacity and airflow are too far removed from the IT equipment heat loads to be effectively available for the heat transfer to occur from the IT equipment outlet airflow to the cooling unit airflow inlet. This can also occur in dynamic data centers, where IT equipment power consumption and heat load are changing per server, and changing locations in the data center group of servers. In such cases, supply airflow out of the floor tiles may not always match required IT equipment airflow requirement, as it was designed for averages, not actual airflow in real time, and real-time adjustments are not possible. The results are hot spots and server RIAT that may at any time exceed allowable values per ASHRAE Standards. Some types of floor tiles address this with automated variable openings.

A design that accounts for maximum required floor tile airflow in any cold aisle under any conditions would become oversized for the actual airflow requirement and wasteful on fan or blower power. Designing with variable speed motors on the blowers is inadequate for a raised-floor supply air delivery system common to multiple cold aisles, since a higher underfloor pressure would be needed in some cold aisles to achieve the higher rated airflow per tile, while lower underfloor pressure would be needed in other cold aisles due to lower IT equipment airflow requirements. These two incompatible and opposite operating conditions could not be achieved by variable speed controls on the blower motors assigned to airflow for the entire raised-floor system.

Fan power in room cooling units may be characterized according to fan type and CFM per kW of power usage. Legacy CRAC units with motor- and belt-driven double width, double inlet centrifugal blowers typically range from 1200 to 2500 CFM/kW. More recent models for CRAC units may include Direct EC Motor-driven plenum or plug-type centrifugal blowers. Capacity ranges for these are dependent on ESP, but for ranges of ESP from 0 to 0.2 in WG, fan capacity may range from 2200 to 4000 CFM/kW.

24.7.1.1 Summary: Room Cooling Airflow and Fan Power For room-based cooling systems, the design impact of underfloor or other supply air delivery systems increases fan power, compared to supply air delivery systems that do not use ducting or a raised-floor system for supply air delivery to the IT equipment inlets. Raised-floor supply air delivery systems also may increase required airflow, and consequently fan power, due to leakage into white space outside the cold aisles, where cool supply air is "wasted" as it is not available to the IT equipment inlet(s).

The design engineer's choice of supply air delivery system also impacts the type of fan and motor available to meet the ESP design conditions, based on pressure drop in the supply air delivery system. Less efficient fan designs may be required, as compared to supply air delivery systems that do not use ducting or a raised-floor system for supply air delivery to the IT equipment inlets.

In a virtual data center, excessive design cooling supply air beyond the minimum required to match actual IT equipment total inlet airflow may be required, to avoid less than required airflow in the vicinity of the more active IT equipment, or to avoid hot spots and warm IT exhaust air recirculation.

24.7.2 Row-Based Cooling Systems

With a row-oriented architecture, the cooling units are associated with a row and are assumed to be dedicated to a row system for design purposes. The cooling units may be mounted in or among the IT equipment racks: they may be mounted overhead, they may be mounted at the back of the racks, or they may be mounted under the floor. Compared with the room-oriented architecture, the airflow paths are shorter and more clearly defined. In addition, airflows are much more predictable, as all of the rated capacity of the cooling system can be utilized, and higher IT equipment load density can be achieved.

The row-oriented architecture has a number of benefits other than cooling performance. Since ESP is eliminated, the reduction in the airflow path length reduces the cooling unit fan power required, increasing efficiency. This is not a minor benefit when one considers that in many lightly loaded data centers, the fixed speed cooling system fan power losses alone exceed the total IT load power consumption.

A row-oriented design allows cooling capacity and redundancy to be targeted to the actual needs of specific rows or two-row zones. For example, row-based architecture allows one row of racks to run high-density IT heat load applications such as blade server, while another row in the same zone satisfies lower power density applications such as communication enclosures. In a similar way, racks with Low and Medium IT Load density may be mixed in the same row, or in the same two-row zone, while achieving adequate airflow and cooling capacity for each rack and IT equipment in the entire row system. Furthermore, $N+1$ or $2N$ redundancy can be targeted at specific rows or two-row zones.

A row-oriented architecture can be implemented without a raised floor. This potentially increases the floor load-bearing capacity, reduces installation costs, eliminates the need for access ramps, and allows data centers to exist in building spaces that otherwise do not have the headroom to permit the installation of a sufficient raised floor. This is particularly an issue for high-density installations where a raised floor height of 3 ft or more is recommended. Examples are the following:

Overhead cooling unit solution (Fig. 24.9)
The row-based cooling unit solution (Fig. 24.10)

FIGURE 24.9 Overhead cooling unit solution. Courtesy of Schneider Electric.

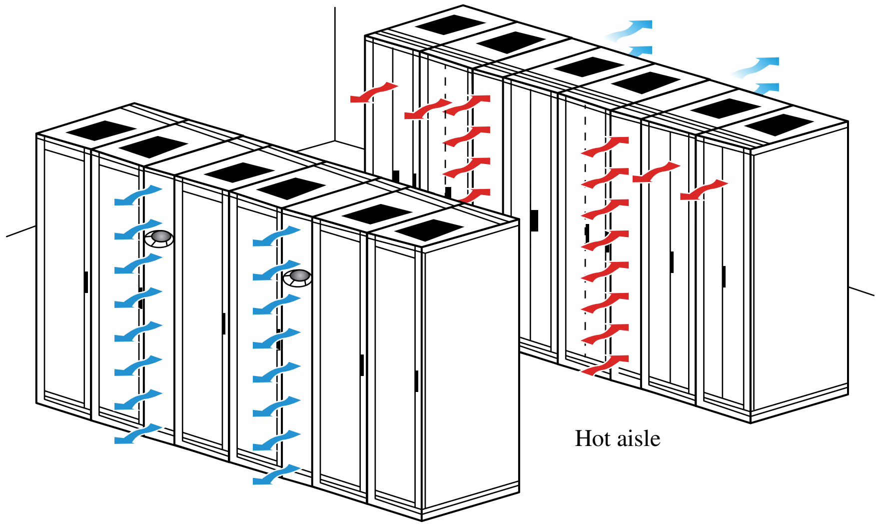


FIGURE 24.10 Row-based cooling unit solution. Courtesy of Schneider Electric.

With row-based air system from the cooling units, cooling system fan power is in direct proportion to the airflow rate CFM and total pressure drop in the system, consisting of the internal cooling unit pressure drop across the coil and filter(s), as well as any velocity pressure drops as the airflow changed directions to disperse in the supply air system route to the IT equipment inlet(s).

There is typically no ESP; in particular, there is no pressure below any raised floor or in any supply air ductwork, since the cooling units supply/outlet and return/inlet are in close proximity to the IT equipment inlet(s) and outlet(s).

The design engineer's choice of supply air delivery system impacts the type of fan and motor available to meet the ESP design conditions, based on pressure drop in the supply air delivery system. More efficient fan designs may be selected, as compared to supply air delivery systems that use ducting or a raised-floor system for supply air delivery to the IT equipment inlets.

For row-based systems, cooling unit total airflow may be affected by the use of open aisles or closed/contained aisles.

24.7.2.1 Open Cold Aisle Consider the open cold aisle system with floor-mounted, row-based, and horizontal airflow cooling units. Since there is no aisle containment system blocking air between the cold aisle and the room, airflow from the row-based cooling units is delivered to the common cold aisle and to a portion of the room above the cold aisle. This portion of supply air that goes into the room above the open cold aisle may be as much as 25% of the total supply airflow from the cooling units, typical for a 4 ft wide cold aisle. It is increased with IT equipment low-density loads in more racks per cooling unit that are located further away from the cooling units.

For example (Fig. 24.11), consider a two-row open cold aisle system with 20 racks and 2 kW average load per rack, using a floor-mounted, row-based CW cooling unit with design cooling capacity of 43 kW and 6900 CFM airflow each. N cooling units would be (1), so $N+1$ would be (2) cooling units. Symmetrical design of the zone would likely place the (2) cooling units opposite each other in the middle of the zone, with five racks to each side per row, a total of (20) racks. Since the last (2) racks at the end of each row are significantly closer to the room air (0–2 ft) than to the cooling units (8–10 ft), the inlet air to the IT equipment mounted in these (4) racks would be almost 100% room air.

IT equipment inlet airflow is proportioned between three vertical zones:

1. Supply/inlet air from the room, especially the into upper rack U spaces
2. A combination of room air and supply air in the middle rack U spaces
3. Supply air from the cooling units, in the lower rack U spaces

To prevent this vertical stratification and higher IT equipment RIAT at the higher U spaces, the cooling units need to ramp up their airflow and extend their sphere of influence, resulting in increased fan power. In the aforementioned example, both row-based cooling units would operate with 100% fan speed, providing $2 \times 6{,}900$ CFM $= 13{,}800$ CFM airflow for the zone, at 100% fan motor power.

24.7.2.2 Closed Hot Aisle Consider the closed contained HACS (Figs. 24.12 and 24.13) with floor-mounted, row-based, and horizontal airflow cooling units. Since there are HACS components blocking air from the room, airflow from

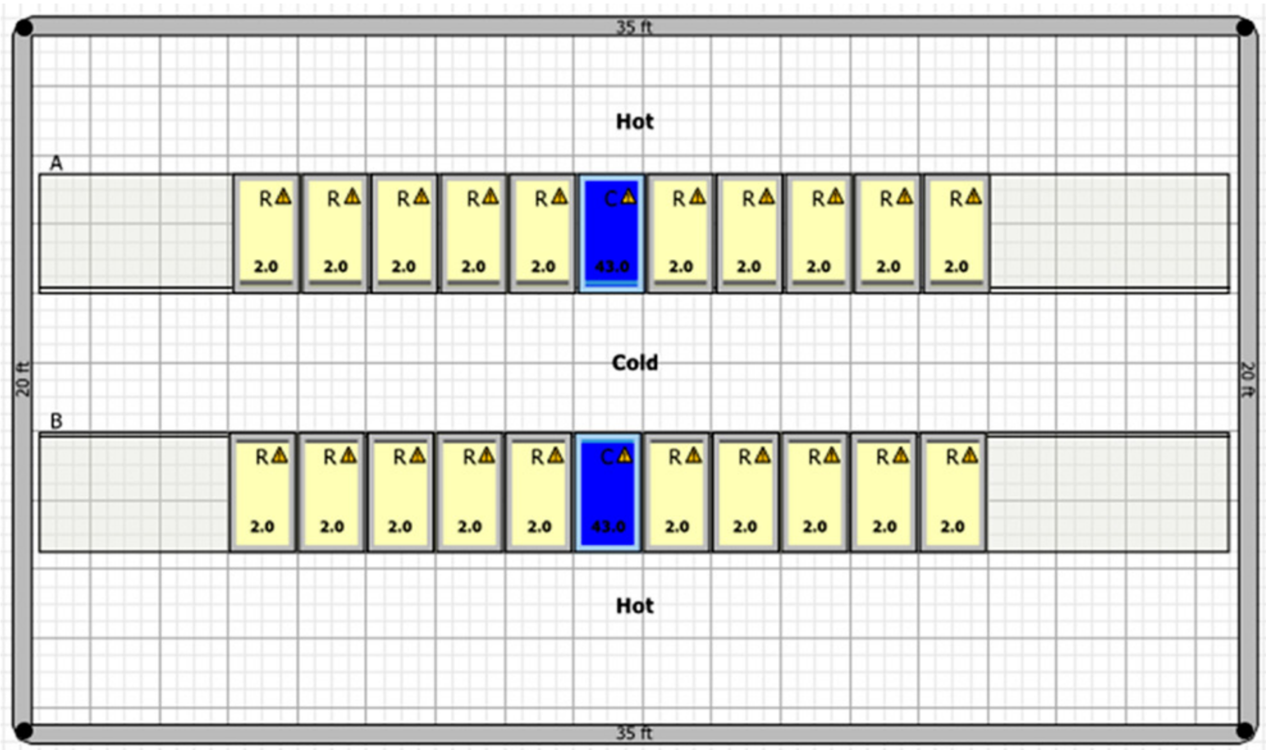


FIGURE 24.11 Cold aisle open system configuration. Courtesy of Schneider Electric.

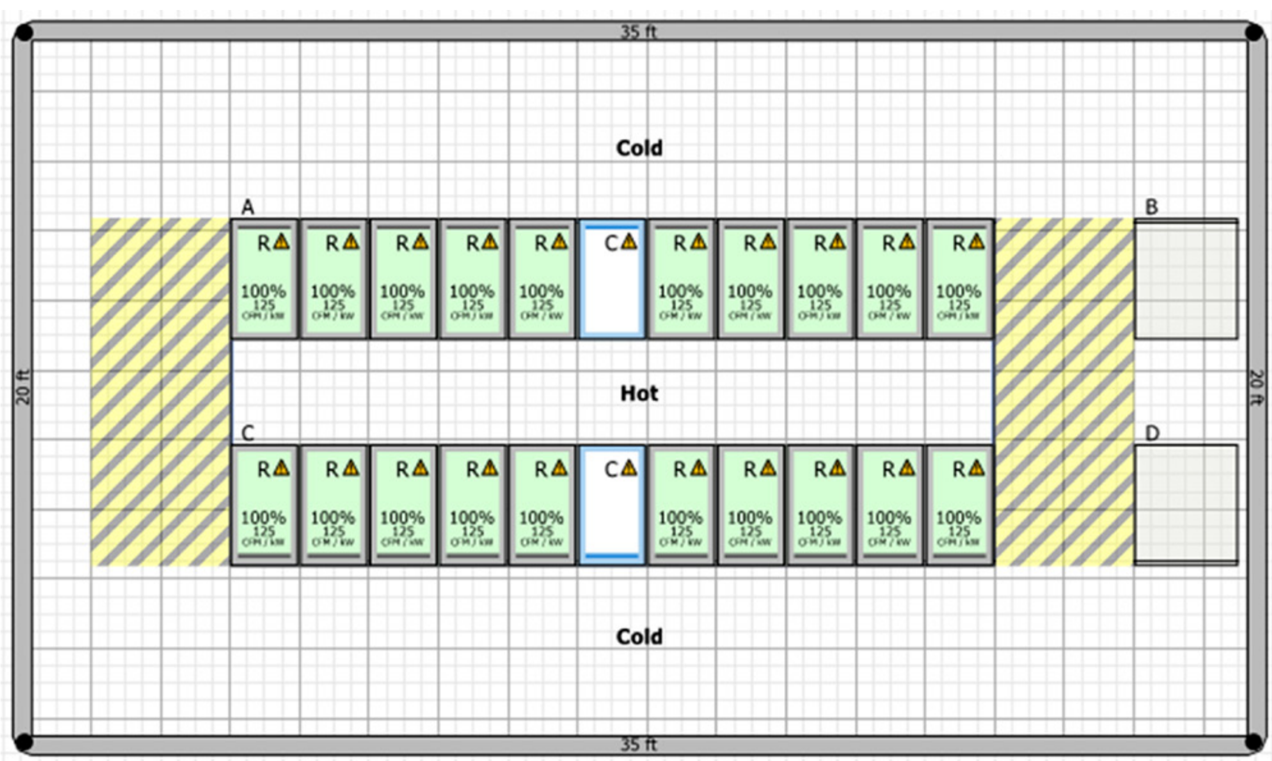


FIGURE 24.12 Hot aisle containment system configuration. Courtesy of Schneider Electric.

the row-based cooling units is removed from the common hot aisle only and not from any portion of the room above the hot aisle or at the ends of the rows. Cooling airflow is automatically controlled to almost match IT equipment airflow and is unaffected by IT equipment at low-, medium-, or high-density loads in racks that are located further away from the cooling units.

Vertical stratification and higher IT equipment RIAT at the higher U spaces are avoided, through the turbulence in the hot aisle zone as the IT exhaust air enters (40 kW × 160 CFM/kW = 6400 CFM IT exhaust airflow). The cooling units adjust their airflow and extend their sphere of influence into the closed space of the HACS, resulting in decreased fan power. In the aforementioned example, both row-based cooling units would operate with 51% fan speed, providing 2 × 3536 CFM = 7072 CFM airflow for the zone, only 10% more cooling unit total airflow than the total IT equipment airflow (7072/6400 = 1.10).

In this model, any leakage between the room and the hot aisle zone is from the room into the zone, since the cooling unit total airflow (7072 CFM) out of the zone is some 10% higher than the IT equipment total airflow (6400 CFM) into the zone.

Based on the Fan Power Affinity Curve (Fig. 24.14), where the power of the fan motor is proportional to the cube of the speed, operating each cooling unit at 51% speed uses ~28% of full fan motor power.

So, (2) cooling units at 51% speed use less (56%) fan power than (1) cooling unit operating at 100% speed. This represents real and continuous energy savings to the data center operations.

Airflow and cooling unit fan power operations would be comparable in a CACS. In this case, even with containment, any leakage of cool supply air would be from the cold aisle zone out into the room and not from the room into the cold aisle zone.

The improvement in redundancy for row-based cooling system is achieved by using containment (either HAC or CAC) to extend the sphere of influence for the cooling units. Consider the (20) rack (2) cooling unit system and HAC above. With (1) cooling unit out of service, the remaining cooling unit with 6900 CFM airflow at 100% fan speed has sufficient airflow capacity to exceed the 6400 CFM of IT equipment airflow with only one unit operating. Cooling power usage would increase to 100%, compared to (2) units operating at 51% speed each. However, sufficient cooling coil capacity and airflow are available from (1) cooling unit for the 40 kW IT equipment load and airflow in this zone.

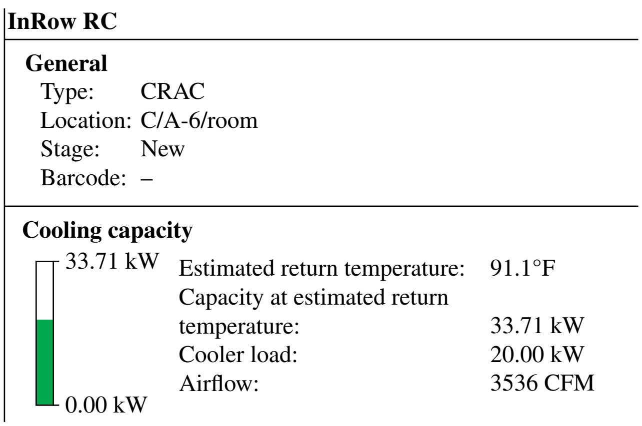


FIGURE 24.13 Model from CFD analysis in design portal (APC-Schneider). Courtesy of Schneider Electric.

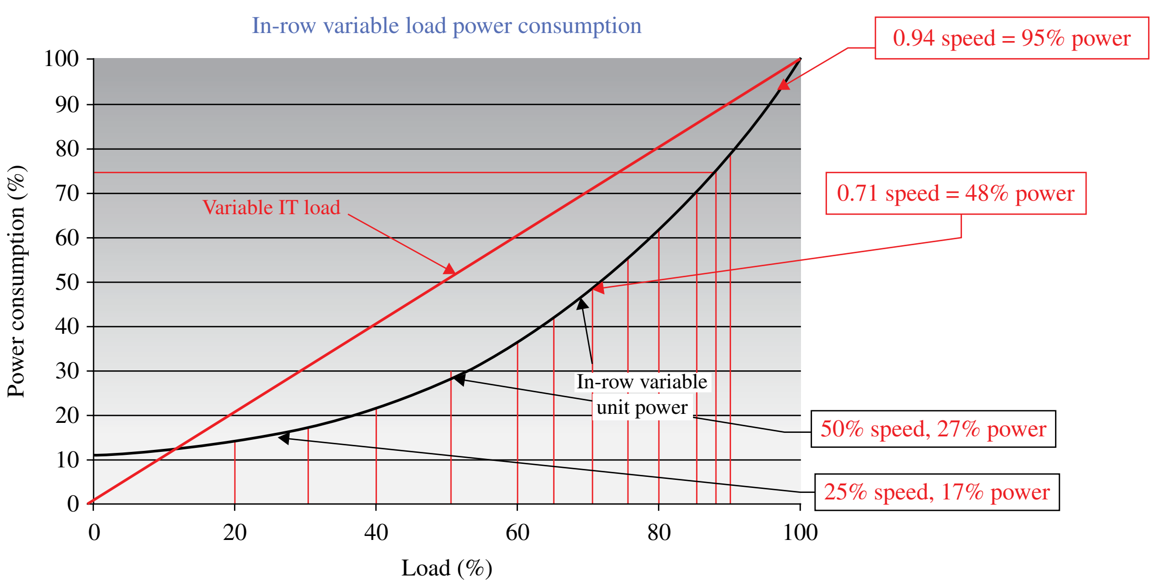


FIGURE 24.14 Fan power curve affinity law. Courtesy of Schneider Electric.

Thus, either HAC or CAC provides predictable and reliable IT equipment RIAT.

In the HAC model, with matching total cooling airflow, all warm IT exhaust airflow is contained and captured from the inside of the hot aisle zone. As it passes through the cooling units, heat is removed, and SAT airflow is delivered to the room/cold aisle side and made available at the IT equipment inlets. This provides a "room-neutral" zone where heat and airflow generated by the IT equipment in an HAC zone are managed, captured, and neutralized by the cooling units before being supplied to the room/cold aisle side of the zone.

In the CAC model, all warm IT exhaust airflow is sent into the hot aisle/room. With matching total cooling unit airflow, cooling supply airflow in sufficient CFM to match IT equipment inlet total is provided in the CAC. A slightly positive pressure is maintained inside of the cold aisle zone, so that cool supply air does not mix with any warm room air inside the cold aisle. As it passes through the cooling units, heat is removed, and SAT is delivered to the cold aisle side and made available at the IT equipment inlets. The containment system components extend the cooling supply airflow sphere of influence, allowing the required airflow to move down the cold aisle away from the cooling units while still preventing any entry of warm room air from the room/hot aisle into the CACS.

The simple and predefined layout geometries of row-oriented architecture give rise to predictable performance that can be completely characterized by the manufacturer and are relatively unaffected by room geometry or other room constraints. This simplifies both the specification and the implementation of designs, at all IT equipment Load densities from less than 2 kW per rack up to 45 kW per rack.

24.8 REDUNDANCY AND COOLING UNIT LOCATION IMPACT

In both HAC and CAC models, $N+1$ or more redundancy is not dependent on cooling unit location with reference to any particular set of racks or IT equipment locations in the same zone.

24.8.1 Room Cooling Redundancy

Consider a room airflow cooling plan (Fig. 24.15) using a raised-floor supply air delivery system. Based on Schneider Electric White Paper #55 [1], this would be characterized as a Flooded Return, Locally Ducted Supply system.

With all (5) CRAC units operating (upper left CFD Model), sufficient cooling and airflow are provided to maintain IT equipment RIAT at some reasonable ASHRAE-recommended limit between 65 and 74°F.

However, when one of the (5) CRAC units is out of service, the raised-floor air static pressure distribution changes, and there are rack locations where the RIAT available to the IT equipment is higher than the ASHRAE-recommended 80°F. Overall room average temperatures may continue to be maintained within recommended limits, but individual rack inlets are insufficiently provided, and temperatures exceed recommended limit.

Compare this lack of adequate cooling to a redundancy plan with HACS or CACS. Two types of systems can be developed with these models: internal cooling units and external cooling units.

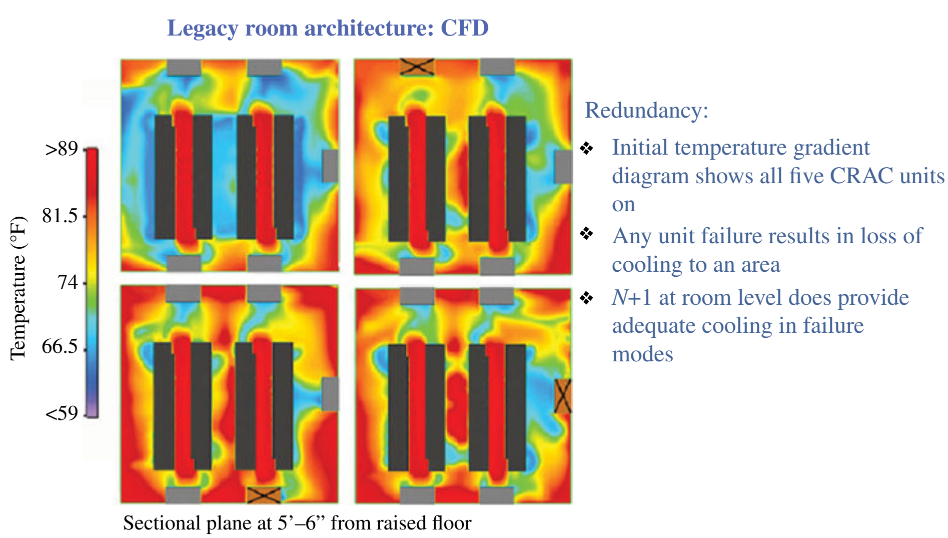


FIGURE 24.15 Legacy room cooling redundancy check CFD. Courtesy of Schneider Electric.

24.8.2 Examples of HAC or CAC with External Cooling Units

24.8.2.1 HAC with External Cooling In HACS, IT equipment exhaust warm air is contained within the HAC enclosed hot aisle and ducted to the cooling unit inlet/return connections. Provided all the cooling unit inlets are collocated on the same common return duct system, location of the cooling units is not significant. The total cooling unit airflow should match or slightly exceed the total IT equipment exhaust/outlet airflow. Provision must be made in the total cooling unit airflow to account for any leakage into the return duct system when N cooling units are operating.

A special case exists where the cooling units are traditional design DX refrigerant evaporator type (as compared to CW coils) where the CRAC unit RAT must be maintained below the maximum allowable for that compressor and refrigerant-based system (Table 24.5). Exceeding the maximum CRAC RAT may cause pressure issues and unit shutdown. With this design, additional room air must be drawn into the return duct system, so as to temper and reduce the hot aisle zone return air system temperature. For example, consider $N+1$ (2) DX CRAC units each with a capacity of 78 kW cooling and 12,600 CFM airflow. If IT equipment heat load is 78 kW, the total airflow in the hot aisle zone = 78 kW × 160 CFM/kW = 12,480 CFM. This 12,480 CFM would be the minimum return airflow requirements in the CRAC(s) to effectively draw all of the hot IT equipment exhaust air out of the HAC and into the CRAC units.

If RIAT is 75°F and IT equipment ΔT is 20°F, the resultant HAC Zone air temperature would be 94°F. With both CRAC units operating at full fan speed, the total cooling airflow would equal 2 units × 12,600 CFM each = 25,200 CFM. The resultant CRAC unit RAT would be 84.4°F. This is higher than the maximum 80°F design CRAC inlet RAT and therefore would require additional CRAC unit(s) operating at less than maximum coil performance capacity, to avoid the resulting inlet RAT higher than allowable.

As an alternate design, with (2) CRAC units in service, the required SAT for the IT equipment would need to be 66°F, in order to achieve an average 80°F inlet RAT in the return airflow to the CRAC unit inlets. This would be based on an IT equipment exhaust air temperature of 85°F.

Redundancy for such a case also requires more than (2) CRAC units, so that the inlet total airflow to operating CRAC units would never be less than the total 25,200 CFM, and using 66°F supply air temperature (SAT).

If only (2) CRAC units are installed, when one is not operational, CRAC inlet RAT would be close to the 85°F IT equipment exhaust air temperature, starting with 66°F RIAT.

This may exceed the DX refrigerant circuit maximum inlet air temperature and cause cooling unit pressure issues and shutdown, with manual reset required to begin operation again.

These high RAT DX CRAC issues are avoided by design in most row-based cooling units specifically engineered to allow up to 105°F RAT. Refer to the Typical Liebert CRV035A and CRV020A performance data given earlier. Similar high RAT design conditions are within the normal operating range for APC-Schneider InRow cooling units as well.

24.8.2.2 CAC with External Cooling For a CACS with external cooling units, IT equipment supply air is contained within the CAC enclosed cold aisle and ducted to the IT equipment inlet(s) in the common cold aisle. Provided all the cooling unit supply outlets are collocated on the same common supply duct system, location

TABLE 24.5 Typical DX air-cooled CRAC performance data[a]
Air-Cooled DX Units (TD/UAV)

Model		0511	0611	0921	1121	1422	1622	1822	2242	2542	3342
Net cooling capacity data											
80°F DB, 67°F WB (26.7°C DB, 19.4°C WB) 50% RH											
Total	BTU/h	58,000	61,000	87,000	116,000	154,000	168,000	171,000	212,000	233,000	267,000
	kW	17.0	17.9	25.5	34.0	45.1	49.2	50.1	62.0	68.3	78.3
Sensible	BTU/h	58,000	61,000	87,000	116,000	154,000	168,000	171,000	212,000	233,000	267,000
	kW	17.0	17.9	25.5	34.0	45.1	49.2	50.1	62.0	68.3	78.3
Evaporator blower/motor: Direct drive electronic commutation (EC) backward curved fans											
Nominal horsepower		3.8	4	4	4	4	4	4	4	4	4
CFM at .20 in WC ESP		3,500	3,500	4,800	7,100	9,200	9,300	9,300	12,600	12,600	12,600
Quantity		1	1	1	2	2	2	2	3	3	3

[a] Courtesy of Schneider Electric.

of the cooling units is not significant. The total N cooling unit airflow should match or slightly exceed the total IT equipment exhaust/outlet airflow. Provision must be made in the total cooling unit airflow to account for any leakage out of the supply duct system when N cooling units are operating. Typical leakage rates for a raised-floor supply air delivery system are 10–15%, depending on pressure required under the floor for the rated airflow through all tiles in the room. If cooling unit supply total airflow in the CAC zone matches IT equipment inlet total airflow, any leakage out of the CAC zone is not significant for maintaining the required IT equipment RIAT.

The same consideration should be given with regard to not exceeding the CRAC unit maximum RAT limit, as with HAC.

24.8.3 Examples of HAC or CAC with Internal Cooling Units

24.8.3.1 HAC with Internal Cooling In HACS, IT equipment exhaust warm air is contained within the HAC enclosed hot aisle and close-coupled via the HAC zone to the cooling unit inlet/return connections. Provided all the cooling unit inlets are collocated on the same common return HACS, location of the cooling units is not significant. The total cooling unit airflow should match or slightly exceed the total IT equipment exhaust/outlet airflow. Because the IT equipment outlet/exhaust and the cooling unit inlet are close-coupled, there is no leakage into any return duct system when N cooling units are operating.

24.8.3.2 CAC with Internal Cooling For a CACS with internal cooling units, IT equipment supply air is contained within the CAC enclosed cold aisle and ducted to the IT equipment inlet(s) in the common cold aisle. Provided all the cooling unit supply outlets are collocated on the same common CACS, location of the individual cooling units is not significant. The total N cooling unit airflow should match or slightly exceed the total IT equipment exhaust/outlet airflow. Because the IT equipment inlets and the cooling unit supply/outlet are close-coupled, there is no leakage into any supply duct system when N cooling units are operating. With cooling unit supply total airflow in the CAC zone matching or slightly exceeding IT equipment inlet total airflow, leakage out of the CAC zone is not significant for maintaining the required IT equipment inlet air temperatures.

The best practices for HAC or CAC internal cooling unit locations are developed with symmetry and cooling units opposite each other across their common hot or cold aisles (summarized in APC-Schneider Application Note #92 [3]—*Best Practices for Designing Data Centers with the InRow RC*).

24.9 IMPACT ON CONDITIONS FOR PERIPHERAL EQUIPMENT IN THE DATA CENTER OUTSIDE ANY OF THE HAC OR CAC ZONE(S)

24.9.1 CAC and Peripheral Equipment

With CAC, the room air temperature is the same as the hot aisle temperature and likely to be some 2°F lower than the IT equipment exhaust air temperature. This allows for skin heat losses to the surrounding building space if it is cooler than the data center room air temperature. The basis of temperature gradient for the room, and hence for inlet air temperatures at any peripheral equipment, would be determined by adding the IT equipment airflow ΔT to the RIAT. This is a required calculation so the cooling unit performance can be determined in part based on the RAT. If the RIAT is 64°F and the IT equipment ΔT is 19°F, the resultant room air temperature may be 81°F. This is at the low value for RIAT as established by ASHRAE standard TC9.9. On the midrange ASHRAE standard value of 75°F RIAT, the resultant room air temperature may be 92°F. Data Center operations personnel will need to determine what constitutes a suitable temperature for the room gradient and adjust RIAT and cooling unit SAT set point accordingly.

Any heat load from peripheral equipment outside the CAC zone will have a corresponding impact on increase in room air temperature and a resulting increase in Cooling unit RAT. For relatively low and low-density peripheral heat loads (up to 5 kW), the room airflow drawn into the cooling units will adequately transfer this warmed air from the vicinity of the peripheral equipment to the inlet of the cooling units. Once drawn into the cooling units, the heat will be automatically rejected by transfer to the cooling unit coil. For peripheral loads greater than 5 kW and any peripheral loads not within 10 ft or so of the CAC zone cooling units, a supplementary cooling system may be required to manage the peripheral heat load and airflow contribution to the room heat profile.

24.9.2 HAC and Peripheral Equipment

With HAC, the room air temperature is the same as the cold aisle temperature and likely to be some 2°F higher than the cooling unit SAT. This allows for skin heat loads from the surrounding building space, if it is warmer than the data center room air temperature.

The basis of temperature gradient for the HAC zone, and hence for return/inlet air temperatures at any cooling unit, would be determined by adding the IT equipment airflow ΔT to the RIAT. This is a required calculation, so the cooling unit performance can be determined, in part based on the RAT. If the RIAT is 64°F and the IT equipment ΔT is 19°F, the resultant cooling unit RAT may be 81°F. This is at the low value for RIAT as established by ASHRAE standard TC9.9. On the midrange ASHRAE standard value of 75°F RIAT, the resultant cooling unit return/inlet air temperature may be 92°F. Data center operations personnel will need to determine what constitutes a suitable temperature for the room gradient and adjust RIAT and determine cooling unit RAT and performance accordingly.

Any heat load from peripheral equipment outside the HAC zone will have a corresponding impact on increase in room air temperature and a resulting increase in RIAT. Based on typical row-based cooling unit control logic, this increase in RIAT will call for an increase in cooling unit airflow and/or cooling coil capacity, to manage the additional heat load while automatically maintaining the set points for cooling unit SAT and RIAT.

For relatively low and low-density peripheral heat loads (up to 5 kW), the peripheral load airflow will mix with the cooling unit supply temperature air and average to the RIAT set point. Room airflow drawn into the IT equipment will adequately transfer this air from the vicinity of the peripheral equipment to the inlet of the IT equipment. Once drawn into the IT equipment, the air will be automatically heated up, drawn into the cooling units, and rejected by transfer to the cooling unit coil. For peripheral loads greater than 5 kW and any peripheral loads not within 10 ft or so of the HAC zone cooling units, a supplementary cooling system in the vicinity of the peripheral equipment may be required to manage the peripheral heat load and airflow contribution to the room heat profile.

In both CACS and HACS, any peripheral heat loads and their location, as well as other contributors to total room heat load, should be factored into the overall cooling system performance requirements. Thus, with proper design, row-based cooling units in a CAC or HAC zone can be adequate for managing the entire room heat load, even considering some peripheral heat load equipment.

In some specific cases, row-based cooling unit alone may not be sufficient, and additional room-based cooling units may be required.

24.10 IMPACT ON ECONOMIZER OPERATION TIME PERIODS DURING COOLER OUTSIDE AMBIENT TEMPERATURES

HACS and CACS differ in their practical ability to allow use of outside economizer heat rejection systems, based on the typical temperature gradients of the two systems. This applies specifically to Dual Cooling (DX and glycol coils) CW systems and Air-to Air economizer systems, where the hours of economizer operation are determined by the temperature differences between ambient temperatures and IT equipment airflow temperatures.

Consider a water-side economizer in Appleton, Wisconsin, USA (Fig. 24.16) and mean, maximum, and minimum temperatures historically as per the following.

Consider APC-Schneider ACRC series with 45°F EWT, RAT = 80°F DB (Table 24.6).

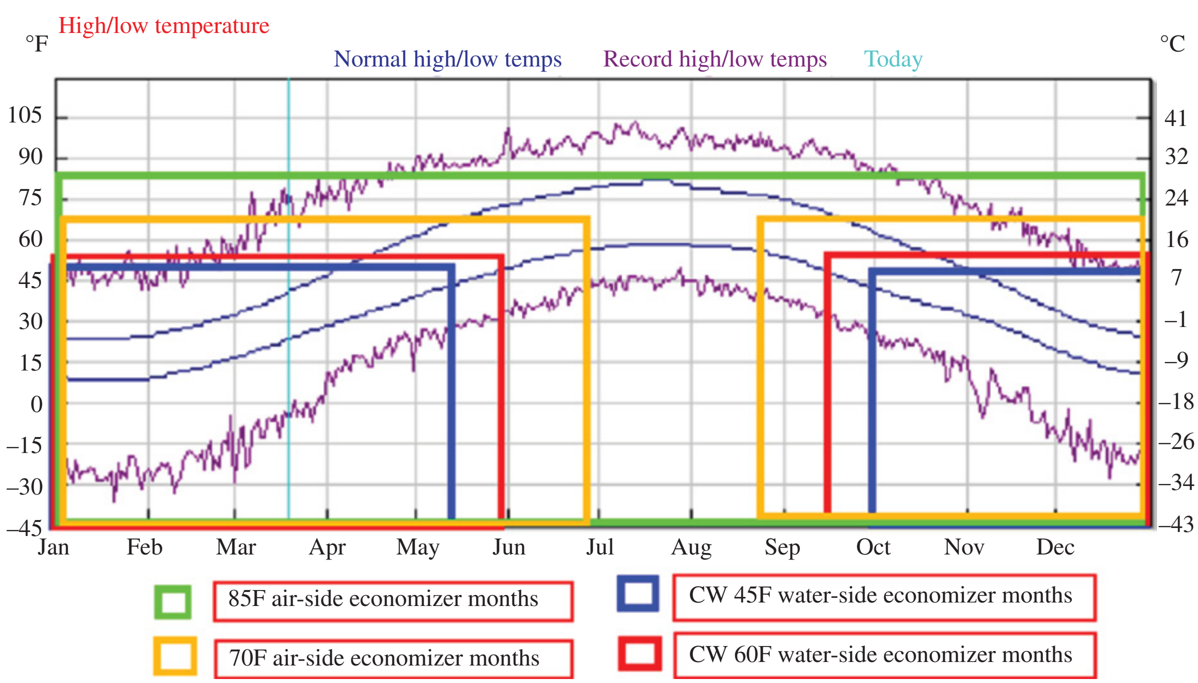


FIGURE 24.16 Appleton, Wisconsin, temperatures and economizer months. Courtesy of Weather Underground.

TABLE 24.6 ACRC100 with 80°F RAT[a] CW 45°F EWT

Temperature DB, WB	CW Delta T	SKU series	Total net capacity	Sensible net capacity	Sensible heat ratio	CW flow rate	Total CW pressure drop
°C (°F)	°C (°F)		kW (BTU/h)	kW (BTU/h)	SHR	l/s (GPM)	kPa (ft H_2O)
26.7°C DB, 17.1°C WB (80°F DB, 62.8°F WB)							
	5.5°C (10°F)	ACRC100	13.2 (45,000)	13.2 (45,000)	1.00	0.62 (9.8)	29.2 (9.8)
		ACRC500	40.8 (139,000)	39.6 (135,000)	0.97	1.8 (28.8)	81 (27.3)

[a] Courtesy of Schneider Electric.

TABLE 24.7 ACRC100 with 95°F RAT[a] CW 60°F EWT

Temperature DB, WB	CW Delta T	SKU series	Total net capacity	Sensible net capacity	Sensible heat ratio	CW flow rate	Total CW pressure drop
°C (°F)	°C (°F)		kW (BTU/h)	kW (BTU/h)	SHR	l/s (GPM)	kPa (ft H_2O)
35.0°C DB, 19.8°C WB (95°F DB, 67.7°F WB)							
	5.5°C (10°F)	ACRC100	17.3 (59,000)	17.3 (59,000)	1.00	0.79 (12.5)	46.17 (15.5)
		ACRC500	40.6 (139,000)	40.6 (139,000)	1.00	1.8 (28.7)	80 (26.8)

[a] Courtesy of Schneider Electric.

Consider ACRC series with 60°F EWT, RAT = 95°F DB (Table 24.7).

From these performance tables, one can conclude that HAC with higher RAT allows the design CW supply temperature to increase from 45 to 60°F, with a corresponding increase of ~1 month in economizer CW mode while still achieving a required cooling unit net sensible capacity above the 45°F EWT capacity for 80°F RAT.

For Model ACRC100, without HAC, 45°F EWT is required to achieve 13.2 kW cooling capacity with 80°F RAT, compared with 17.3 kW cooling capacity with 60°F EWT and HACS that provides a RAT = 95°F.

For Model ACRC500, without HAC, 45°F EWT is required to achieve 40.8 kW cooling capacity with 80°F RAT, compared with 40.6 kW cooling capacity with 60°F EWT and HACS that provides a RAT = 95°F.

A savings of chiller compressors not operating while chiller is in economizer mode can be calculated based on the specific chiller size. As an example, consider a 500 kW IT Load Data Center, using a nominal 150 ton chiller. Typical power savings when a 150 ton chiller is in full economizer mode amount to 182 kW in chiller condenser and compressor power. As a result, one extra month of savings equates to 131,040 kWh total savings for the month.

In a similar way, operating the fluid-cooled coil in a dual-coil cooling unit (DX and glycol) with higher RAT can also extend the economizer coil operating hours when the same unit net sensible capacity can be achieved with higher EWT to the glycol cooling coil. Again, this represents the power cost savings of not running the compressors or, in certain unit designs, not running the compressors at full capacity, for some additional time each year.

As with any cooling unit serving an HACS with fully ducted return air, consideration must be given to the maximum allowable inlet air temperature for the DX operating side.

For Air-to-Air economizer systems, economizer cooling capacity is typically related to a 5°F split between the airflow RAT and the ambient temperature. Thus, a 15°F increase in RAT, from 75 to 90°F (70–85°F ambient temps), would correspond to another 2 months of full Air-to-Air economizer operation in Appleton, Wisconsin, USA.

24.11 CONCLUSION AND FUTURE TRENDS

Four key elements contribute to cooling unit performance and efficiency in the data center:

1. ***Contain*** the Heat
2. ***Close-Coupled*** Cooling
3. Best in Class Right-sized ***Components***
4. Manage ***Capacity***

24.11.1 HAC Meets All Four Elements

HACS can be designed and deployed to meet all four of these, providing maximum cooling performance and efficiency, as well as allowing for capacity match and room-neutral IT Zones in any data center layout. This design can also accommodate any room peripheral heat loads outside the HAC zone(s).

24.11.2 CAC Meets Only Two Elements

CACS by design only satisfy elements (2) and (3), by failing to contain the heat and not allowing maximum capacity cooling matched to IT heat loads, defaulting to a room cooling model and lower RAT, which is critical to cooling unit performance.

24.11.3 HAC Has Design Advantage

For future data centers, the clear advantages of HACS as compared to CAC should determine the design criteria and deployments in favor of HAC.

24.11.4 Chimney Rack Ducts Are a Special Case

Single-geometry passive chimney exhaust ducts from IT cabinets with no other means of airflow regulation are dependent primarily on different pressure drops across the exhaust duct for different airflow rates. Caution in design is needed for varying IT loads per rack and chimney ducts.

REFERENCES

[1] Rasmussen N. Airflow architecture in data center designs. West Kingston: Schneider Electric ITB; 2012. APC-Schneider White Paper #55.

[2] Rasmussen N. Calculating total cooling requirements for data centres. West Kingston: Schneider Electric ITB; 2003. APC-Schneider White Paper #25.

[3] Niemann J. Best practices for designing data centers with the InfraStruXure InRow. West Kingston: Schneider Electric ITB; 2006. APC-Schneider Application Note #92.

FURTHER READING

ASHRAE TC9 Committee. ASHRAE TC9 Addendum and Appendix E. Table E-1. ASHRAE 2011 Thermal guidelines (I-P Units). Atlanta: ASHRAE Headquarters; 2011.

Capes J. APC-Schneider Presentation by Joe Capes. Power Point—data centre trends. West Kingston: Schneider Electric ITB; 2010.

Dunlap K, Rasmussen N. The advantages of row and rack-oriented cooling architectures for data centers. West Kingston: Schneider Electric ITB; 2006. APC-Schneider White Paper #130.

Emerson Technical Team CRV Tech Manual. Emerson-Liebert Tech Data Manual for CRV Model: Document SL-11978_REV1_02-10. Columbus: Emerson Network Power; 2013.

Emerson Technical Team—WP166-19. Emerson-Liebert White Paper: focused cooling using cold aisle containment. Columbus: Emerson Network Power; 2013.

Fink J. Hot aisle containment working conditions. West Kingston: Schneider Electric ITB; 2005. APC-Schneider White Paper #123.

Lemke K. Hot aisle containment. West Kingston: Schneider Electric ITB; 2008. APC-Schneider Application Note #146.

Niemann J, Brown K, Avelar V. Hot aisle vs. cold aisle containment. West Kingston: Schneider Electric ITB; 2010. APC-Schneider White Paper #135.

Niles S, Donovan P. Virtualization and cloud computing: optimized power, cooling, and management maximizes benefits. West Kingston: Schneider Electric ITB; 2012. APC-Schneider White Paper #118.

Uhrhan G, Buell S. Selection procedure for InRow chilled water products. West Kingston: Schneider Electric ITB; 2007. APC-Schneider Application Note #126.

VanGilder JW, Schmidt RR. Airflow uniformity through perforated tiles in a raised-floor data center. The Proceedings of IPACK2005, ASME Interpack 05; July 17–22, 2005; San Francisco.

Weather Underground. WebPage for U.S. Temperatures. San Francisco: Weather Underground, Inc. Available at St. Louis to USA: http://www.wunderground.com/ Accessed on May 27, 2014.

25

FREE COOLING TECHNOLOGIES IN DATA CENTERS

Nicholas H. Des Champs and Keith Dunnavant
Munters Corporation, Buena Vista, VA, USA

25.1 INTRODUCTION

Development and use of computers for business and science were a result of attempts to remove the drudgery of many office functions and to speed the time required to do mathematically intensive scientific computations. As computers developed from the 1950s tube-type mainframes, such as the IBM 705, through the minicomputers of the 1980s, they were typically housed in a facility that was also home to many of the operation's top-level employees. And because of the cost of these early computers and the security surrounding them, they were housed in a secure area within the main facility. It was not uncommon to have them in an area enclosed in lots of glass so that the computers and peripheral hardware could be seen by visitors and employees. It was an asset that presented the operation as one that was at the leading edge of technology.

These early systems generated considerably more heat per instruction than today's servers. Also, the electronic equipment was more sensitive to temperature, moisture, and dust. As a result, the computer room was essentially treated as a modern-day clean room. That is, high-efficiency filtration, humidity control, and temperatures comparable to operating rooms were standard. Since the computer room was an integral part of the main facility and had numerous personnel operating the computers and the many varied pieces of peripheral equipment, maintaining the environment was considered by the facilities personnel as a more precise form of "air-conditioning (AC)."

Development of the single-chip microprocessor during the mid-1970s is considered to be the beginning of an era in which computers would be low enough in cost and had the power to perform office and scientific calculation, allowing individuals to have access to their own "personal" computers. The early processors and their host computers produced very little heat and were usually scattered throughout a department. For instance, an 8086 processor (refer to Table 25.1) generated less than 2 W of heat and its host computer generated on the order of 25 W of heat (without monitor). Today's servers generate around 500 W of heat and, when used in a modern data center, are loaded 40 servers to a rack or 20,000 W of heat in a very small space. Considering a data center with 200 racks, there is 20,000 W/rack × 200 racks or 4 mW of heat to dissipate.

Of course, there would be no demand for combining thousands of servers in large data centers had it not been for the development of the Internet and the launching of the World Wide Web in 1991 (at the beginning of 1993, only 50 servers were known to exist on the WWW), development of sophisticated routers, and many other ancillary hardware and software products. During the 1990s, the use of the Internet and personal computers mushroomed as is illustrated by the rapid growth in routers: in 1991, Cisco had 251 employees and $70 million in sales, and by 1997, it had 11,000 employees and $7 billion in sales. Another example of demand for server capacity is as follows: the total number of websites at the end of 2011 is 555 million, and of that number, 300 million were added in 2011. The total number of Internet servers worldwide is estimated to be 75 million.

As technology has evolved, so have the cooling requirements. No longer is a new data center "air-conditioned," but instead, it is considered "process cooling" where a cooling fluid is delivered to a "cold aisle," traverses the process to a hot aisle, and then is either discarded to ambient or returned to the process-fluid machines for extraction of heat and then sent back to the cold aisle.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

TABLE 25.1 Chronology of computing processors

Processor	Clock speed	Introduction	Mfg. process	Transistors
4004	108 kHz	November 1971	10 μm	2,300
8086	10 MHz	June 1978	3 μm	29,000 1.87 W (sustained)
386	33 MHz	June 1988	1.5 μm	275,000
486	33 MHz	November 1992	0.8 μm	1.4 million
Pentium	66 MHz	March 1993	0.8 μm	3.1 million
Pentium II	233 MHz	May 1997	0.35 μm	7.5 million
Pentium III	900 MHz	March 2001	0.18 μm	28 million
Celeron	2.66 GHz	April 2008	65 nm	105 million
Xeon MP X7460	2.66 GHz	September 2008	45 nm	190 billion 170.25 W (sustained)

Processor Designations are copyright of Intel Corporation.

Today's allowable cooling temperatures reflect the conceptual change from AC to process cooling. There have been three changes in ASHRAE's cooling guidelines during the last 9 years. In 2004, ASHRAE recommended Class 1 temperature as 68–77°F (20–25°C); in 2008, it was 64.4–80.6°F (18–27°C). Today, the 2011s guidelines remain the same in terms of ***recommended*** range but greatly expand the ***allowable*** range of temperatures and humidity in order to give operators more flexibility in doing compressor-less cooling (using ambient air directly or indirectly) to remove the heat from the data center with the goal of increasing the data center cooling efficiency and reducing the energy efficiency metric Power Usage Effectiveness (PUE).

25.2 USING PROPERTIES OF AMBIENT AIR TO COOL A DATA CENTER

In some instances, it is the ambient conditions that are the principal criteria that determine the future location of a data center, but most often, the location is based on acceptance by the community, access to Internet backbone, and adequate supply and cost of utilities in addition to being near the market it serves. Ambient conditions have become a more important factor as a result of an increase in allowable cooling temperature for the IT equipment. The cooler, and sometimes dryer, the ambient, the greater period of time a data center can be cooled by using ambient air. For instance, in Reno, NV, cooling air can be supplied all year at 72°F (22°C) with no mechanical refrigeration by using evaporative cooling techniques.

Major considerations by the design engineers when selecting the cooling system for a specific site are:

a. Cold aisle temperature and maximum temperature rise across server rack
b. Critical nature of continuous operation for individual servers and peripheral equipment
c. Availability of sufficient water for use with evaporative cooling
d. Yearly ambient dry-bulb (db) and wet-bulb (wb) temperatures, extremes of db and wb temperatures, and air quality, that is, particulate and gases
e. Utility costs

Other factors are projections of initial capital cost; full-year cooling cost; reliability; maintenance cost; and the effectiveness of the system in maintaining the desired space temperature, humidity, and air quality during normal operation and during a power or water supply failure.

Going forward, four commonly used economizer cooling systems are discussed in greater detail: three *A*ir-to-*A*ir *H*eat e*X*changers (AtoAHXs) systems and one *D*irect *E*vaporative *C*ooling (DEC) system. The systems using AtoAHXs are considered *I*ndirect *A*ir-*S*ide *E*conomizers (IASEs) because ambient air is used to indirectly cool the recirculating airstream without delivering ambient air to the space. The DEC system is a *D*irect *A*ir-*S*ide *E*conomizer (DASE) because ambient air traverses the media, reducing the db temperature, and is delivered directly to the space. Any form of AtoAHX may be used that does not transfer latent energy between airstreams. Typically, plate-type heat exchanger, sensible heat wheels, and heat pipes are used [1].

25.3 ECONOMIZER THERMODYNAMIC PROCESS AND SCHEMATIC OF EQUIPMENT LAYOUT

25.3.1 DASE

25.3.1.1 Cooling with Ambient DB Temperature The simplest form of economizer uses ambient air directly supplied to the space to cool equipment. Figure 25.1 shows a schematic of a typical DASE arrangement and includes a DEC, item 1, and a cooling coil, item 2. Without item 1, this schematic would represent a DASE that uses the db temperature of the ambient air to cool the Datacom equipment. For this case, ambient air can be used to perform all of the cooling when its temperature is below the design cold aisle

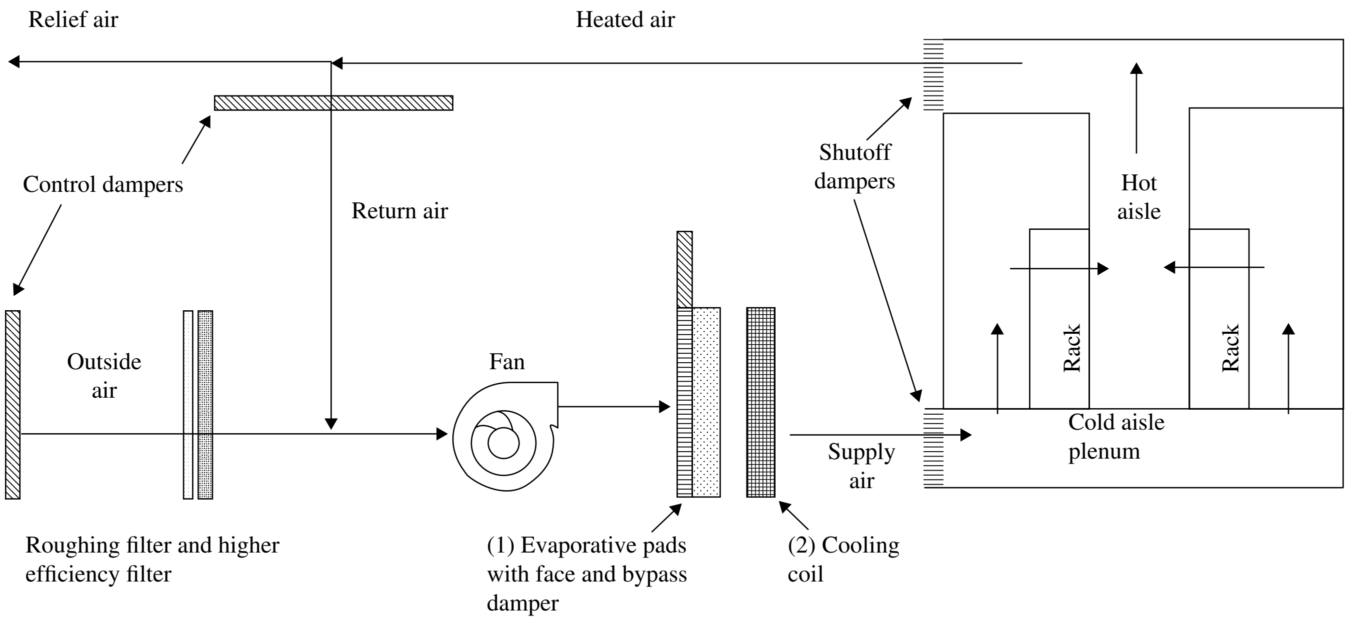


FIGURE 25.1 Schematic of a typical direct air-side economizer.

temperature and a portion of the cooling when it is below the design hot aisle temperature. When ambient is above hot aisle temperature, the system must resort to all mechanical cooling, and when ambient is below the design cold aisle temperature, some of the heated process air is returned to the inlet plenum to mix with the incoming outdoor air to yield the desired delivery temperature. In almost all cases, except in extreme cold climates, some level of mechanical cooling is required to meet the space-cooling requirements, and in most cases, the mechanical supplement will be designed to handle the full cooling load. The result is that for most regions of the world, the full-year energy reduction is appreciable but the capital equipment cost reflects the cost of having considerable mechanical refrigeration on board. Other factors to consider are costs associated with bringing high levels of outdoor air into the building, which result in higher rate of filter changes and less control of space humidity. Also possible gaseous contaminants, not captured by standard high-efficiency filters, could pose a problem.

25.3.1.2 Cooling with Ambient wb Temperature If a source of usable water is available at the site, then an economical approach to extend the yearly accumulated hours of cooling, as discussed in the previous paragraph, is to add DEC as shown in Figure 25.1. The Evaporative Pads typically can achieve 90–95% efficiency in cooling the ambient air to wb temperature from db, resulting in a db being delivered to space at only a few degrees above the ambient wb temperature. The result is that the amount of trim mechanical cooling required is considerably reduced from using ambient db. In addition, there is greater space humidity control by using the DEC arrangement (1) to add water to the air during colder ambient conditions. The relative humidity within the space, during cooler periods, is controlled with the face and bypass dampers on the DEC. There would be no humidity control during the warmer ambient conditions. In fact, lack of humidity control is the single biggest drawback in using DASE with DEC. As with the db cooling, factors to consider are costs associated with bringing high levels of outdoor air into the building, which results in higher rates of filter change and less control of space humidity. Also possible gaseous contaminants, not captured by standard high-efficiency filters, could pose a problem. Even with these operating issues, the DASE using DEC is arguably the most efficient and least costly of the many techniques for removing waste heat from Datacom facilities, except for DASE used on facilities in extreme climates where the maximum ambient db temperature never exceeds the specified maximum cold aisle temperature.

A DASE using a DEC process is illustrated in Figure 25.2. In this instance, the cold aisle temperature is 75°F, the hot aisle is 95°F, and the design wb is 67.7°F. With a 90% effective evaporative process, the supply air (SA) to the space can be cooled to 70°F from 91.2°F, lower than specified. Under this type of condition, there are several control schemes that are used to satisfy the space cooling requirements:

1. Allow the hot aisle temperature to remain at 95°F, thereby increasing the ΔT between cold aisle and hot aisle temperatures. This is accomplished by reducing process fan speed, which results in less recirculating airflow and consequently considerably less fan power. This scheme is shown as the process between the two square end marks.

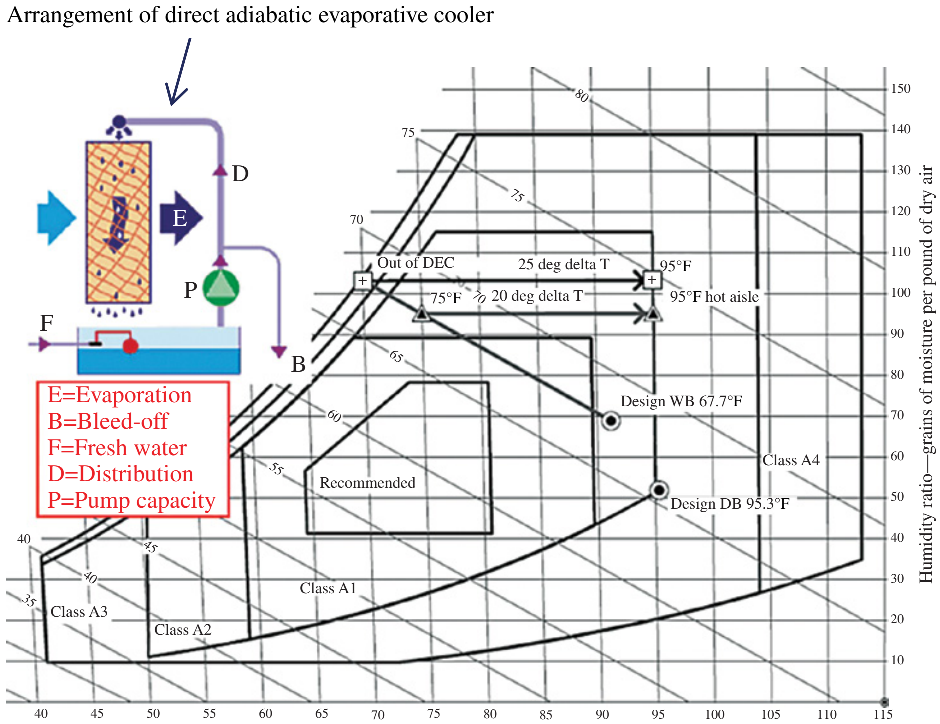


FIGURE 25.2 Direct cooling processes shown with recommended and allowable envelopes for datacom supply temperature and moisture levels.

(a)

(b)

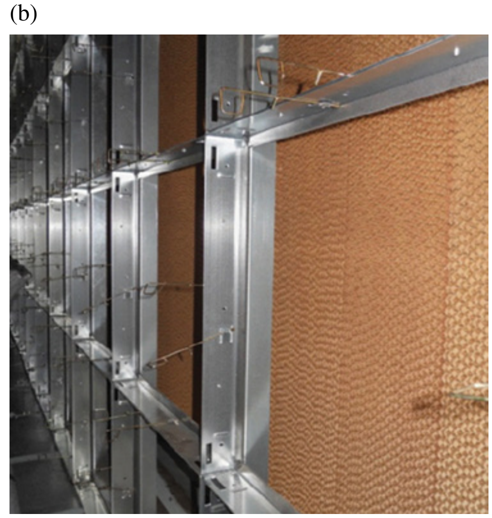


FIGURE 25.3 (a) Cooling system using bank of direct evaporative cooler units at the end wall of a data center; (b) The array of evaporative cooling media.

2. Simply let the airflow remain constant and end up with a cold aisle of 70°F and a hot aisle of 90°F with the specified airflow as shown in the horizontal process line from "Out of DEC" to 90°F db temperature. This condition will occur because the heat load of the space remains constant.
3. Use face and bypass dampers on the DEC to control the cold aisle SA temperature to 75°F as shown in the process between the two triangular end marks.

A bank of five DEC units arranged in parallel is shown in Figure 25.3. There are a total of 20 units on the job, each supplying 40,000 cfm of adiabatically cooled outdoor air during warm periods and a blend of outdoor air and recirculated air, as illustrated in Figure 25.1, during colder periods. The cooling air is supplied directly to the cold aisle, travels through the servers and other IT equipment, and is then directed to the relief dampers on the roof. Also shown in Figure 25.3 is a commonly used type of rigid, fluted DEC media.

FIGURE 25.4 From left, plate-type heat exchanger, heat pipe, and sensible-only rotary wheel.

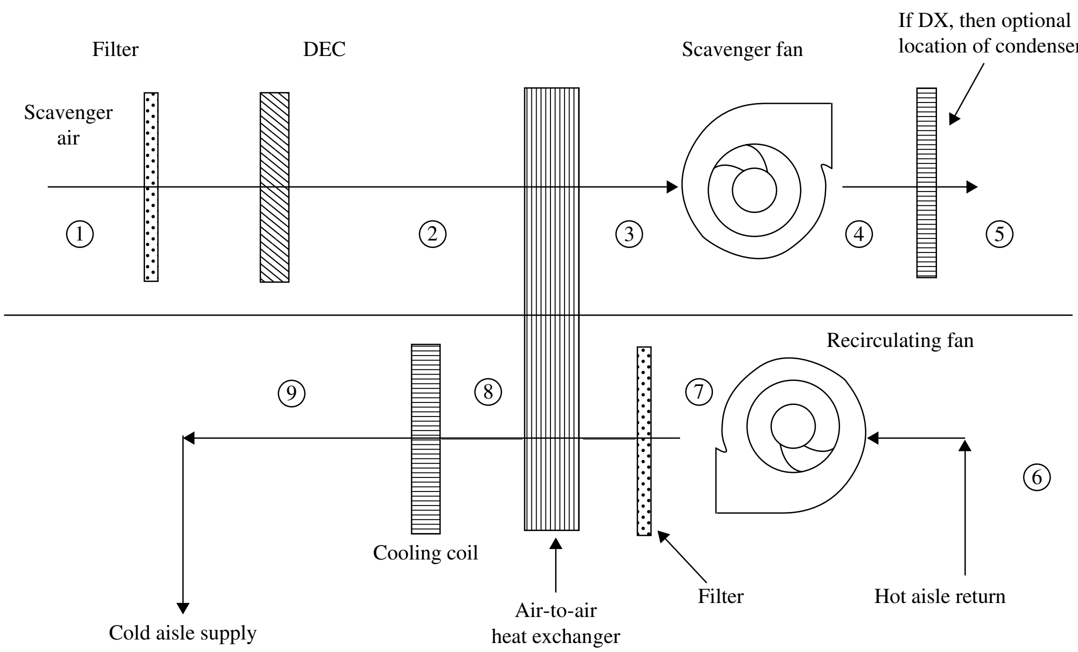


FIGURE 25.5 Schematic of typical indirect air-side economizer.

25.3.2 IASE

25.3.2.1 AtoAHXs In most Datacom cooling applications, it is desirable to cool the recirculated air as opposed to delivering ambient air directly into the space for cooling. This indirect technique allows for better space humidity control and reduces the potential for airborne contaminants entering the space. When cooling recirculated air, dedicated makeup air units control space humidity and building pressure. AtoAHXs serve as the intermediary that permits the use of ambient air to cool the space without introducing the ambient air to the space. The most commonly used types of AtoAHXs used for this purpose are the plate, heat pipe, and sensible wheel as shown in Figure 25.4. (See the 2012 *ASHRAE Handbook*, Chapter 26, for further information regarding performance and descriptions of AtoAHXs.) Figure 25.5 illustrates the manner in which the AtoAHX is used to transfer the heat from the hot aisle return air (RA) to the cooling air, commonly referred to as scavenger air (ScA), since it is discarded to ambient after it performs its intended purpose, that of absorbing heat. The effectiveness of an AtoAHX, when taking into consideration the cost, size, and pressure drop, is usually selected to be between 65 and 75% when operating at equal airflows for the ScA and recirculating air.

As shown in Figure 25.5, the ScA enters the system through a roughing filter ① that removes materials that are contained in the outdoor air that might hamper the operation of the DEC, AtoAHX, or the optional refrigeration

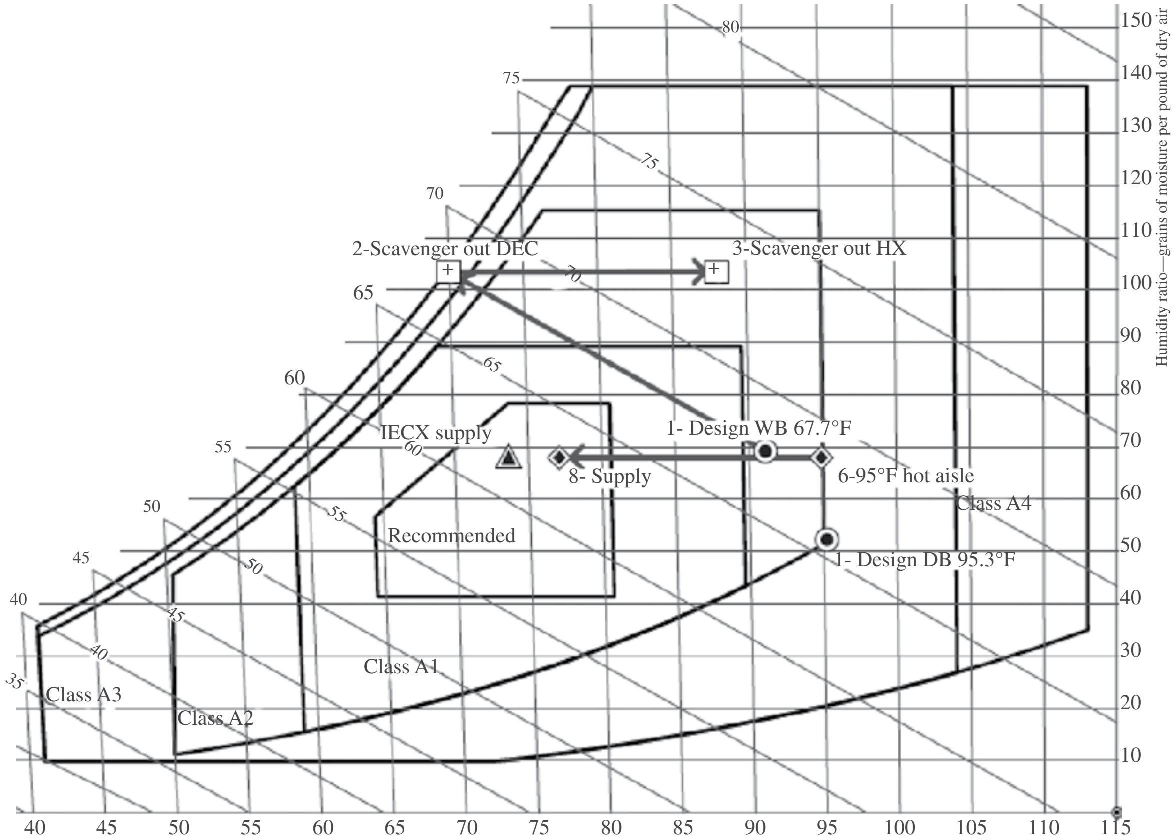
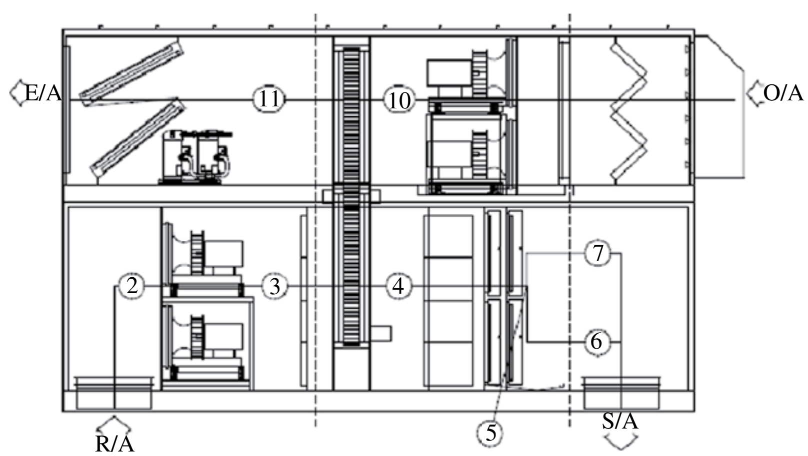


FIGURE 25.6 Psychrometric chart showing recommended and allowable envelopes of temperature and moisture content of delivered cooling air for datacom facilities.



Operating point	Summer (normal)			Summer max			Winter		
	DB (°F)	WB (°F)	ACFM	DB (°F)	WB (°F)	ACFM	DB (°F)	WB (°F)	ACFM
R/A	94	69.2	79,424	94	69.2	79,424	94	69.2	79,424
2	94	69.2	79,424	94	69.2	79,424	94	69.2	79,424
3	98.3	70.4	79,995	98.3	70.4	79,995	98.3	70.4	79,995
4	75.4	63.4	76,169	98.3	70.4	79,195	74	62.9	75,982
5	74	62.9	75,982	74	62.9	75,982	74	62.9	75,982
6	74	62.9	75,982	74	62.9	75,982	74	62.9	75,982
7	74	62.9	75,982	74	62.9	75,982	74	62.9	75,982
S/A	74	62.9	75,982	74	62.9	75,982	74	62.9	75,982
O/A	65.4	54.6	80,000	105.4	72.7	80,000	–14.5	–15.5	13,027
10	67	55.2	80,228	107	73.1	80,207	–14.5	–15.5	13,027
11	86.9	62.4	83,608	107	73.1	81,019	94.5	54	16,884
E/A	88.7	63	83,608	147.4	82.8	86,300	94.5	54	16,884

FIGURE 25.7 Schematic of 80,000 cfm IASE using a 14 ft diameter heat wheel.

condenser coil. If a sufficient amount of acceptable water is available at the site, then cooling the scavenger air with a DEC before it enters the AtoAHX at ② should definitely be considered, since in all cases evaporatively cooling the air will extend the energy-saving capability of the IASE over a greater period of time and also reduce the amount of trim mechanical refrigeration required to meet the cooling requirements on the extreme ambient design conditions. The ambient conditions used for design of cooling equipment are generally extreme db temperature if just an AtoAHX is used or extreme wb temperature if a form of evaporative cooling is used to precool the scavenger air before it enters the heat exchanger. What is considered to be the extreme ambient condition is job dependent and is usually selected using either Typical Meteorological Year 3 (TMY3) data, the extreme 50-year ASHRAE data, or even the 0.4% ASHRAE Design Conditions.

When DEC is used as shown in Figure 25.5 and trim DX (Direct Expansion Refrigeration) cooling is required, then it is advantageous to place the condenser coil in the leaving scavenger airstream since its db temperature, in almost all cases, is lower than the ambient db temperature. If no DEC is used and trim DX cooling is required and placed in the leaving ScA, then the scavenger air could have a temperature level above the recirculating air temperature, which would result in the air cooling the condenser being above ambient temperature. Under these circumstances, there should be a means to prevent the heat exchanger from transferring heat in the wrong direction; otherwise, heat will be transferred from the ScA to the recirculating air, and the trim mechanical refrigeration will not be able to cool the recirculating air to the specified cold aisle temperature. Vertical heat pipe heat exchangers automatically prevent heat transfer at these extreme conditions, because if the ambient is hotter than the return air, then no condensing of the heat pipe working fluid will occur (processes ②–③ as shown in Figure 25.5) and therefore no liquid will be returned to the portion of the heat pipe in the recirculating airstream (processes ⑦–⑧). Heat wheels can be made to cease rotating, thereby eliminating heat transfer if the ambient temperature exceeds the return air temperature. With the plate heat exchanger, a face and bypass section to direct ScA around the heat exchanger may be necessary in order to prevent heat transfer.

As an example, when using just an AtoAHX without DEC and assuming an effectiveness of 72.5% (again using 75°F cold aisle and 95°F hot aisle), the economizer can do all of the cooling when the ambient db temperature is below 67.4°F. At lower ambient temperatures, the scavenger fans are slowed in order to remove the correct amount of heat and save on scavenger fan energy. Above 67.4°F ambient, the mechanical cooling system is staged on until at an ambient of 95°F or higher; the entire cooling load is borne by the mechanical cooling system.

FIGURE 25.8 160,000 cfm IASEs with heat wheels removing a total of 10 mW of heat from a data center.

When precooling ScA with a DEC, it is necessary to discuss the cooling performance with the aid of a psychometric chart. The numbered points on Figure 25.6 correspond to the numbered locations shown in Figure 25.5. On a design wb day ① of 67.7°F, the DEC lowers the ScA to 70.1°F ②. The ScA then enters the heat exchanger and heats to 88.2°F ③. During this process, air returning from the hot aisle ⑥ is cooled from 95°F (no fan heat added) to 77.2°F ⑧ or 89% of the required cooling load. Therefore, on a design day, using DEC and an AtoAHX, the amount of trim mechanical cooling required, ⑨ in Figure 25.5, is only 11% of the full cooling load, and the trim would only be called into operation for a short period of time during the year.

Figures 25.7 and 25.8 represent specifications and design for a total of 12 IASE Heat Wheel Units, (2) 80,000 and (10) 160,000 cfm units that remove 10 mW of heat from a data center. The maximum outdoor air temperature is 105.4°F and the return air from the hot aisle is 94°F, so on this application, there is a requirement that the DX cooling on board be sufficient to supply the full AC load, that is, the design ambient temperature is greater than the hot aisle temperature. Even though the full complement of DX is required, the AtoAHXs supply over 90% of the full-year ton-hour cooling requirement, leaving less than 10% of the ton-hours to be supplied by the DX.

25.3.2.2 Integral AtoAHX Cooling Tower The previous section used a separate DEC and AtoAHX to perform an indirect evaporative cooling process. The two processes can be integrated into a single piece of equipment, known as an Indirect Evaporative Cooling heat exchanger (IECX). The IECX approach to using wb

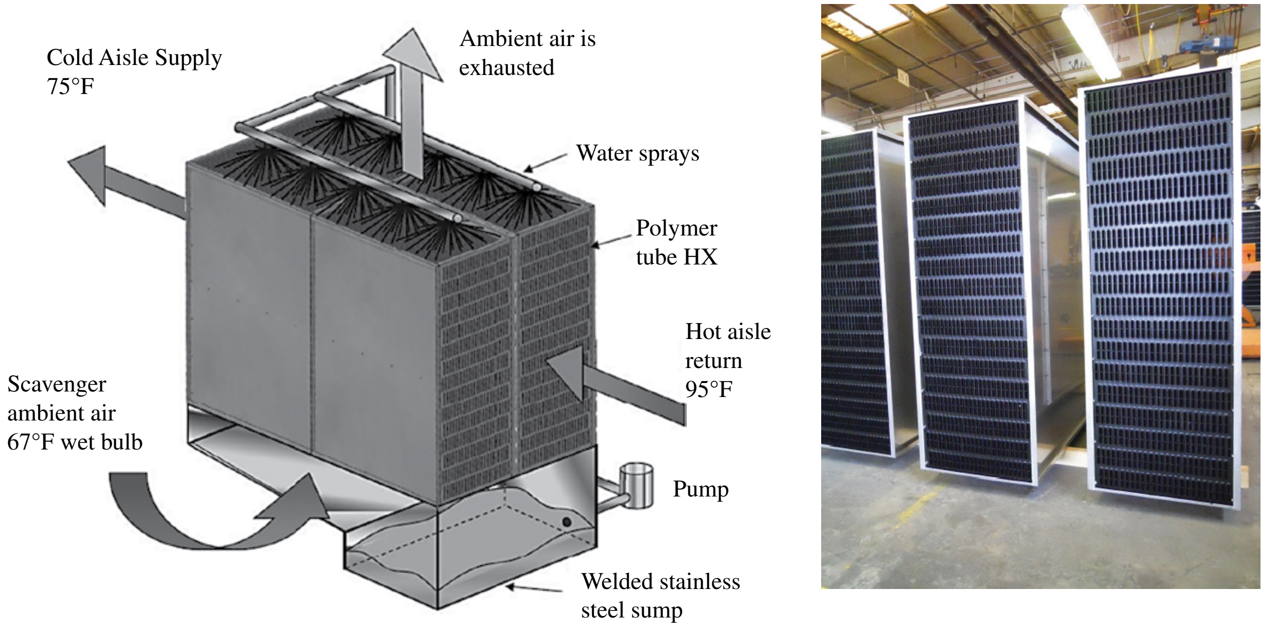


FIGURE 25.9 Integral heat exchanger/cooling tower or IECX.

temperature as the driving potential to cool Datacom facilities can be more efficient than using a combination of DEC and heat exchanger, the reason being the thermodynamic processes involved that are beyond the scope of this chapter.

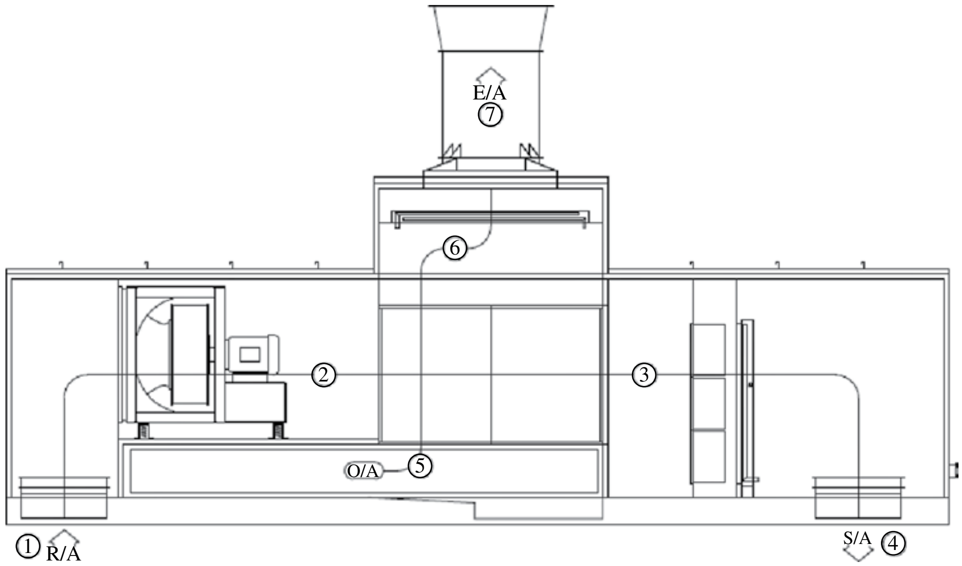
Configuration of a typical IECX is illustrated in Figure 25.9. The recirculating Datacom air returns from the hot aisle at 95°F and enters the horizontal tubes from the right side and travels through the inside of the tubes where it cools to 75°F. The recirculating air cools as a result of the cooling tower effect of ScA evaporating water that is flowing downward over the outside of the tubes. Because of the evaporative cooling effect, the water flowing over the tubes and the tubes themselves cool to within a few degrees of the scavenger inlet wb temperature. Typically, an IECX is designed to have wet-bulb depression efficiency (WBDE) in the range of 70–80%. Referring to Figure 25.6, with all conditions remaining the same as the dry air-to-air heat exchanger with a DEC precooler on the ScA, a 78% efficient IECX process is shown to deliver a cold aisle temperature of 73.7°F, shown as a triangle, which is below the required 75°F. Under these conditions, the ScA fan speed is controlled to move less air in order to reduce the heat removal and maintain the specified cold aisle temperature at 75°F instead of 73.7°F.

The unit schematic and operating conditions shown in Figure 25.10 are for a Hewlett Packard Datacom facility in Sydney, Australia. Referring to the airflow pattern in the schematic, the return air at 90°F comes back to the unit from the hot aisle ①, heats to 92°F through the fan ②, and enters the tubes of the IECX where it cools to 83.2°F ③ on a design ambient day of 113°F/80.1°F (db/wb). The trim DX then cools the supply to the specified cold aisle temperature of 70°F. At these extreme operating conditions, the economizer section removes 40% of the heat load, and the DX removes the remaining 60% of the heat. This condition occurs once every 10 or 15 years, but in this case, the DX had to be on board to handle this extreme. For the entire year, the economizer removes 99.71% of the total annual heat load.

The period of time during the year that an economizer is performing the cooling function is extremely important because a Datacom facility utilizing economizer cooling has a lower PUE than a facility with conventional cooling using chillers and Computer Room Air Handler (CRAH—Chilled Water Coil) units or Computer Room Air Conditioner (CRAC—DX) units. PUE is a metric used to determine the energy efficiency of a Datacom facility. PUE is determined by dividing the amount of power entering a data center by the power used to run the computer infrastructure within it. PUE is therefore expressed as a ratio, with overall efficiency improving as the quotient decreases toward 1. There is no firm consensus of the average PUE; the value for data centers is 1.8, according to a survey of more than 500 data centers conducted by **the Uptime Institute** in 2011, and in 2012, the CTO of Digital Realty indicated that the average PUE for a data center was 2.5. Economizers range from as low as 1.07 PUE for a DASE using DEC to a high of about 1.3. The IECX ranges from 1.1 to 1.2 depending upon the efficiency of the integral cooling tower/heat exchanger and the site location. So, if the economizer at the HP Sydney location reduced the time that the mechanical refrigeration was operating by 99.7% during a year, then the cooling costs were reduced by a factor of around 5 relative to a data center with a PUE of 2.0.

Referring to Figure 25.10, the ScA enters at the bottom of the IECX ⑤ and flows upward and over the tubes where it

	Location	Supply	Scavenger
Fan data	Airflow (CFM/L/s)	16,166/7630	18,003/8496
	Ext. static pressure (in WG/Pa)	1.25/311	N/A
	Total static pressure (in WG/Pa)	3.86/961	1.16/289
	Motor size (HP/kW)	15/11.2	7.5/5.6



Operating point	Design					
	DB (°F)	DB (°C)	WB (°F)	WB (°C)	CFM	L/s
1 (R/A)	90.0	32.2	67.4	19.7	16,166	7630
2	92.0	33.3	68.0	20.0	16,225	7657
3	83.2	28.4	65.2	18.4	15,966	7535
4 (S/A)	70.0	21.1	60.7	15.9	15,578	7352
5 (O/A)	113.0	45.0	80.1	26.7	18,200	8589
6	88.6	31.4	82.2	27.9	17,425	8224
7 (E/A)	106.8	41.6	86.1	30.1	18,003	8496

FIGURE 25.10 Schematic and operating conditions for units installed on a 5 mW datacom facility.

evaporates water flowing downward and simultaneously absorbs the heat from the recirculating air. It leaves the top of the IECX ⑥ in a near-saturated condition and at a db temperature of 88.6°F, 24.4°F below the ambient temperature, before performing its second job, that of removing heat from the refrigeration condenser coil and thus improving compressor performance with the resulting lower condensing temperature.

Figure 25.11 is an aerial view of the partially completed Hewlett Packard Datacom facility in Sydney, Australia. When completed, it will house 10 mW of computing power; 42 cooling units are shown, and 42 more will be required upon full build-out. What appear to be stacks on top of the units is actually the scavenger fan outlet diffuser that improves the fan performance, reduces horizontal radiated noise, and thrusts the moist, heated ScA sufficiently above the roof so that it does not circulate back into the ScA inlet and cause a reduction in cooling performance.

The shaded area in Figure 25.12 represents the Bin Hours (right ordinate) that a typical IECX unit might operate at each wb Bin. Most of the hours are between about 11 and 75°F. The upper curve, medium dashed line, is the total operating power of the economizer. The short dashed curve is the DX power, and the dot–dash curve is the scavenger fan motor, both of which operate at full capacity at the extreme wb temperatures. The average weighted total power for the year is 117 kW. Typically, the lights and other electrical within the Datacom facility is about 3% of the IT load, so the total average load into the facility is 1500 kW × 1.03 + 117 kW or 1662 kW. This yields an average value of PUE of 1662/1500 or 1.108, an impressive value when compared to conventional cooling PUEs of 1.8–2.5. For this example, the onboard trim DX represented 24% of the 452.9 tons of heat rejection.

In order to give a better understanding of how the IECX performs at different climatic conditions and altitudes, Figure 25.13 shows the percentage of cooling ton-hours performed during the year: first, the IECX operating wet (warm conditions using wb temperature); second, the IECX

FIGURE 25.11 Aerial view of (42) indirect air-side economizers (IASEs).

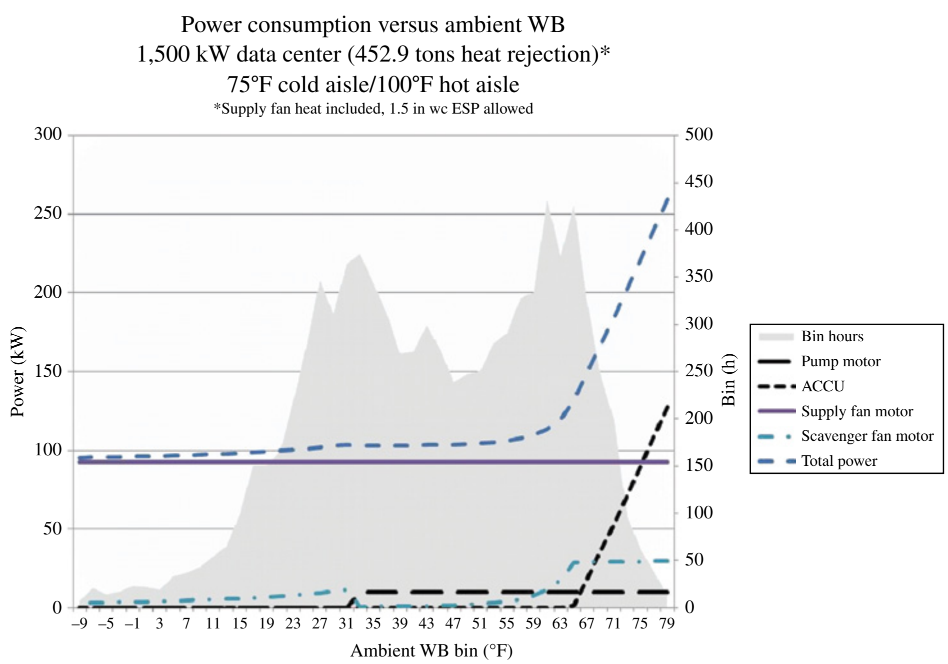


FIGURE 25.12 Power consumption for a typical IECX IASE cooling unit.

Location	Elevation (ft)	0.4% WB design (MCDB/WB °F)	% reduction of peak mechanical cooling requirement*	% annual ton-h IASE (wet)	% annual ton-h IASE (dry)	% annual ton-h mechanical cooling	% annual hours mechanical cooling is off
Ashburn, VA (IAD)	325	88.8/77.7	65.7	53.0	44.1	2.9	78.7
Atlanta, GA	1027	88.2/77.2	67.2	73.7	22.0	4.3	70.8
Boston, MA	0	86.3/76.2	70.1	51.5	47.8	0.7	91.6
Chicago, IL	673	88.2/77.9	65.1	46.4	52.3	1.3	88.8
Dallas, TX	597	91.4/78.6	63.0	69.8	22.8	7.4	62.1
Denver, CO	5285	81.8/64.6	100.0	51.3	48.7	0.0	100.0
Houston, TX	105	89.0/80.1	58.4	74.8	15.2	10.0	48.0
Los Angeles, CA	325	78.0/70.2	87.4	99.2	0.7	0.1	97.9
Miami, FL	0	86.8/80.2	58.1	84.1	0.3	15.6	24.5
Minneapolis, MN	837	87.5/76.9	68.1	46.4	52.3	1.3	90.3
Newark, NJ	0	88.8/77.7	65.7	54.6	43.7	1.7	84.9
Phoenix, AZ	1106	96.4/76.1	70.3	83.1	14.5	2.4	80.7
Salt Lake City, UT	4226	86.8/67.0	95.9	50.1	49.9	0.0	99.8
San Francisco, CA	0	78.2/65.4	100.0	70.7	29.3	0.0	100.0
Seattle, WA	433	82.2/66.5	97.5	58.6	41.4	0.0	99.8

System Design Parameters:
1 MW load, $N = 4$
Target supply air = 75°F, target return air = 96°F
N+1 redundancy, with redundant unit operating for annual analysis
MERV 13 filtration consolidated in only (2) units
Water sprays turned off below 50°F ambient db
1.0 in. ESP (supply + return)
IECS WBDE ≈75% and Dry Effectiveness ≈56%

Notes:
System (wet) rejects 100% of ITe load when ambient wet-bulb temperature is below 67°F
System (dry) rejects 100% of ITe load when ambient dry-bulb temperature is below 55°F
System does not introduce any outside air into data hall; all cooling effects are produced indirectly
*Percentage reduction in mechanical cooling equipment normally required at peak load based on N units operating

Percentage of annual cooling contribution with IECX IASE

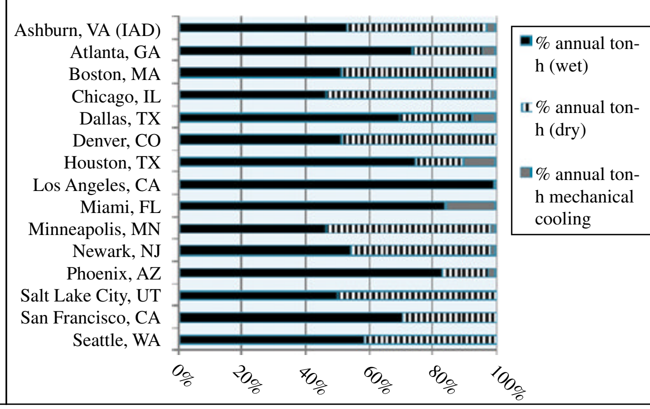


FIGURE 25.13 Analysis summary for modular data center cooling solution using IECX.

operating dry (cool conditions using db temperature); and third, at extreme conditions operating wet with aid of DX. Fifteen cities are listed with elevations ranging from sea level to over 5000 ft. The embedded chart gives a graphical representation of the energy saved during each operating mode. The last column is the percentage of time during the year that there are no compressors staged on and the IECX is handling the entire cooling load.

25.4 COMPARATIVE POTENTIAL ENERGY SAVINGS AND REQUIRED TRIM MECHANICAL REFRIGERATION

Numerous factors have an influence on the selection and design of a Datacom cooling system. Location, water availability, allowable cold aisle temperature, and extreme design conditions are four of the major factors. Table 25.2 compares the cooling concepts previously discussed as they relate to percentage of cooling load during the year that the economizer is capable of removing and the tons of trim mechanical cooling that has to be on board to supplement the economizer on hot days. The former represents full-year energy savings and the latter initial capital cost.

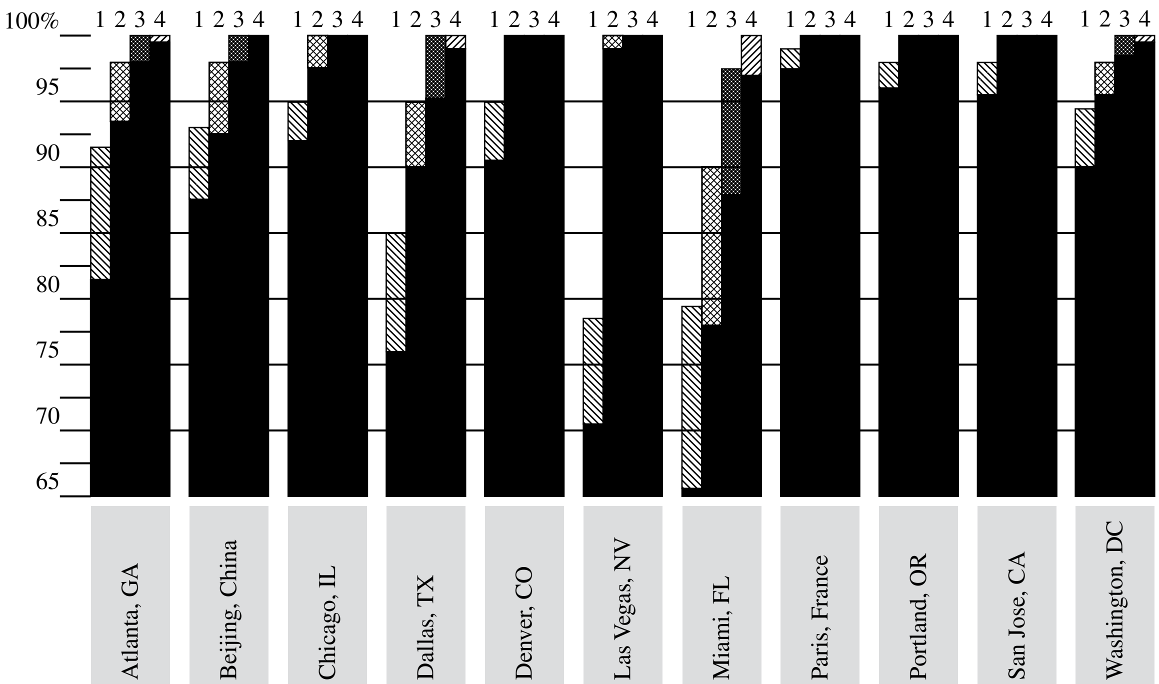
To use Table 25.2, take the following steps:

1. Select the city of interest and use that column to select the following parameters.
2. Select either TMY Maximum or 50-year Extreme section for the ambient cooling design.
3. Select the desired cold aisle/hot aisle temperature section within the section selected in step 2.
4. Compare the trim cooling required for each of the four cooling systems under the selected conditions.

Dallas, Texas, using an AtoAHX, represented by the No. 1 at the top of the column, will be used as the first example. Operating at a cold aisle temperature of 75°F and a hot aisle of 95°F, represented by the solid black bars, 76% of the

TABLE 25.2 Annualized economizer cooling capability based on TMY3 data

Solid black: 75°F/95°F (23.9°C/35°C) cold aisle/hot aisle
Hash marks: 80°F/100°F (26.7°C/37.8°C) cold aisle/hot aisle
1, Air-to-air HX; 2, DEC+Air-to-air HX; 3, IECX; 4, DEC



Trim DX using TMY maximum temperatures, tons

75/95°F (23.9/35°C) cold aisle/hot aisle temperature

	Atlanta, GA	Beijing, China	Chicago, IL	Dallas, TX	Denver, CO	Las Vegas, NV	Miami, FL	Paris, France	Portland, OR	San Jose, CA	Washington, DC
1	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.22	1.80	1.80	1.80
2	1.06	1.11	0.95	1.10	0.25	0.58	0.90	0.52	0.88	0.34	0.94
3	0.55	0.97	0.78	0.96	0.00	0.34	0.73	0.27	0.70	0.06	0.77
4	3.58°F	8.62°F	10.53°F	9.0°F	0°F	1.91°F	5.64°F	0°F	6.05°F	0°F	6.72°F

80/100°F (26.7/37.8°C) cold aisle/hot aisle temperature

	Atlanta, GA	Beijing, China	Chicago, IL	Dallas, TX	Denver, CO	Las Vegas, NV	Miami, FL	Paris, France	Portland, OR	San Jose, CA	Washington, DC
1	1.68	1.75	1.48	1.80	1.80	1.80	1.55	0.89	1.71	1.55	1.74
2	0.77	0.81	0.65	0.80	0.00	0.28	0.61	0.23	0.58	0.05	0.64
3	0.20	0.62	0.43	0.61	0.00	0.00	0.37	0.00	0.35	0.00	0.42
4	0°F	3.6°F	1.55°F	4.0°F	0°F	0°F	0.64°F	0°F	1.05°F	0°F	1.72°F

Trim DX using extreme 50-year maximum temperatures, tons

75/95°F (23.9/35°C) cold aisle/hot aisle temperature

	Atlanta, GA	Beijing, China	Chicago, IL	Dallas, TX	Denver, CO	Las Vegas, NV	Miami, FL	Paris, France	Portland, OR	San Jose, CA	Washington, DC
1	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80
2	1.06	1.38	1.11	1.09	0.29	1.00	1.20	0.85	1.29	0.85	1.15
3	0.92	1.29	0.98	0.95	0.00	0.84	1.08	0.66	1.20	0.66	1.03
4	9.6°F	15.9°F	6.55°F	10.86°F	0°F	9.93°F	11.17°F	6.24°F	13.57°F	6.7°F	11.2°F

80/100°F (26.7/37.8°C) cold aisle/hot aisle temperature

	Atlanta, GA	Beijing, China	Chicago, IL	Dallas, TX	Denver, CO	Las Vegas, NV	Miami, FL	Paris, France	Portland, OR	San Jose, CA	Washington, DC
1	1.80	1.80	1.80	1.80	1.80	1.80	1.76	1.80	1.80	1.80	1.80
2	0.77	1.08	0.82	0.80	0.00	0.70	0.90	0.56	1.00	0.56	0.86
3	0.56	0.94	0.63	0.60	0.00	0.49	0.73	0.31	0.85	0.31	0.68
4	4.66°F	9.9°F	5.53°F	5.86°F	0°F	4.93°F	6.17°F	1.24°F	8.57°F	1.7°F	6.2°F

Tons of additional mechanical AC per 1000 SCFM of cooling air required to achieve desired delivery temperature when using air economizers—with no economizer, the full AC load is 1.8 tons/1000 SCFM.

TABLE 25.3 Design temperatures that aid in determining the amount of trim cooling

	50-year extreme				Maximum from TMY 3 data			
	DB		WB		DB		WB	
	°F	°C	°F	°C	°F	°C	°F	°C
Atlanta	105.0	40.6	82.4	28.0	98.1	36.7	77.2	25.1
Beijing	108.8	42.7	87.8	31.0	99.3	37.4	83.2	28.4
Chicago	105.6	40.9	83.3	28.5	95.0	35.0	80.5	26.9
Dallas	112.5	44.7	82.9	28.3	104.0	40.0	83.0	28.3
Denver	104.8	40.4	69.3	20.7	104.0	40.0	68.6	20.3
Las Vegas	117.6	47.6	81.3	27.4	111.9	44.4	74.2	23.4
Miami	99.4	37.4	84.7	29.3	96.1	35.6	79.7	26.5
Paris	103.2	39.6	78.8	26.0	86.0	30.0	73.2	22.9
Portland	108.1	42.3	86.4	30.2	98.6	37.0	79.3	26.3
San Jose	107.8	42.1	78.8	26.0	96.1	35.6	70.2	21.2
Washington, DC	106.0	41.1	84.0	28.9	99.0	37.2	80.3	26.8

cooling ton-hours during the year will be supplied by the economizer. The other 24% will be supplied by a cooling coil. The size of the mechanical cooling system, termed trim cooling, is shown in the lower part of the table as 1.8 tons per 1000 scfm (Standard Cubic Feet per Minute) of cooling air, which is also the specified maximum cooling load that is required to dissipate the IT heat load. Therefore, for the AtoAHX in Dallas, the amount of trim cooling required is the same tonnage as would be required when no economizer is used. That is because the TMY3 design db temperature is 104°F, well above the return air temperature of 95°F. Even when the cold aisle/hot aisle is raised to 80°F/100°F, the full load of trim cooling is required. If a DEC (represented by No. 2 at top of column) is placed in the ScA (TMY3 maximum wb temperature is 83°F), then 90% of the yearly cooling is supplied by the economizer, and the trim cooling drops to 1.1 tons per 1000 scfm from 1.8 tons.

For the second example, we will examine Washington, DC, where the engineer has determined that the design ambient conditions will be based on TMY3 data. Using 75°F/95°F cold aisle/hot aisle conditions, the IECX and DEC, heat exchangers No. 3 and No. 4, can perform 98 and 99% of the yearly cooling, respectively, leaving only 2 and 1% of the energy to be supplied by the mechanical trim cooling. The Air-to-Air HX (No. 1) accomplishes 90% of the yearly cooling, and if a DEC (No. 2) is added to the scavenger airstream, the combination does 96% of the cooling. The trim cooling for heat exchangers 1, 2, and 3, respectively, is 1.8, 0.94, and 0.77 tons, where 1.8 is full-load tonnage. Increasing the cold aisle/hot aisle to 80°F/110°F allows No. 3 and No. 4 to supply all of the cooling with the economizers and reduces the amount of onboard trim cooling.

From Table 25.1, even in climates such as Miami, Florida, economizers should be investigated as an alternative to full mechanical cooling for Datacom facilities. In addition, the economizers presented in this section will become even more desirable for energy savings as engineers and owners become more familiar with the recently introduced allowable operating environments A1 through A4 as shown on the psychrometric charts of Figures 25.2 and 25.4. In fact, if the conditions of A1 and A2 were allowed for a small portion of the 8766 total hours per year, then for No. 2 and No. 3, all of the cooling could be accomplished with the economizers, and there would be no requirement for trim cooling when using TMY3 extremes. For No. 4, the cooling could also be fully done with the economizer, but the humidity would exceed the envelope during hot, humid periods.

There are instances when the cooling system is being selected and designed for a very critical application where the system has to hold space temperature under the worst possible ambient cooling condition. In this case, the ASHRAE 50-year Extreme Annual Design Conditions are used as referred to in Chapter 14 of Ref. [2] and designated as “complete data tables” and underlined in blue in the first paragraph. These data can only be accessed by means of the disk that accompanies the ASHRAE Handbook. The extreme conditions are shown in Table 25.3, which also includes for comparison the maximum conditions from TMY3 data.

Using the 50-year extreme temperatures of Table 25.3, the amount of trim cooling, which translates to additional initial capital cost, is shown in the lower portion of Table 25.2. All values of cooling tons are per 1000 scfm (1699 m^3/h) with a final cold aisle to hot aisle temperature rise of 20°F (11.1°C). For the DEC designated as No. 4, instead of showing tons, temperature rise above desired cold aisle temperature is given.

From a cost standpoint, just what does it mean when the economizer reduces or eliminates the need for mechanical cooling? This can best be illustrated by comparing the pPUE of an economizer system to that of a modern, conventional mechanical cooling system. pPUE in this case is a ratio of (IT cooling load + power consumed in cooling IT load)/(IT load). The pPUE value of economizers ranges from 1.07 to about 1.3. For refrigeration systems, the value ranges from 1.8 to 2.5. Taking the average of the economizer performance as being 1.13 and using the lower value of a refrigeration (better performance) system of 1.8, the economizer uses only 1/6 of the operating energy to cool the data center when all cooling is performed by the economizer.

As an example of cost savings, if a Datacom facility operated at an IT load of 5 mW for a full year and they paid $0.10 per kW-hour, then the power cost to operate the IT equipment would be $4,383,000 per year. To cool with mechanical refrigeration equipment with a PUE of 1.80, the cooling cost would be $3,506,400 per year for a total electrical cost of $7,889,000. If the economizer handled the entire cooling load, the cooling cost would be reduced to $570,000 per year. If the economizer could only do 95% of the full cooling load for the year, then the cooling cost would still be reduced from $3,506,400 to $717,000—a reduction worth investigating.

25.5 CONVENTIONAL MEANS FOR COOLING DATACOM FACILITIES

In this chapter, we have discussed techniques for cooling that first consider economization as the principal form of cooling. There are more than 20 ways to cool a data center using mechanical refrigeration with or without some form of economizer as part of the cooling strategy. References [3] and [4] cover these various mechanical cooling techniques. Chapter 19 of Ref. [5] discusses standard techniques for Datacom cooling.

REFERENCES

[1] ASHRAE. *ASHRAE Handbook—Systems and Equipment*. Atlanta: American Society of Heating Refrigeration and Air Conditioning Engineers, Inc.; 2012.

[2] ASHRAE. *ASHRAE Handbook—Fundamentals*. Atlanta: American Society of Heating Refrigeration and Air Conditioning Engineers, Inc.; 2013.

[3] Evans T. The different technologies for cooling data centers, Revision 2. Available at http://www.apcmedia.com/salestools/VAVR-5UDTU5/VAVR-5UDTU5_R2_EN.pdf. Accessed on May 27, 2014.

[4] Kennedy D. Understanding data center cooling energy usage & reduction methods. Urbana: Rittal; February 2009. Rittal White Paper 507.

[5] ASHRAE. *ASHRAE Handbook—Applications*. Atlanta: American Society of Heating Refrigeration and Air Conditioning Engineers, Inc.; 2011.

FURTHER READING

ASHRAE. *Thermal Guidelines for Data Processing Environments.* 3rd ed. Atlanta: ASHRAE; 2012.

Atwood D, Miner J. Reducing data center cost with an air economizer. Hillsboro: Intel; 2008.

Dunnavant K. Data center heat rejection. ASHRAE Journal 2011;53:11.

Quirk D, Sorell V. Economizers in datacom—risk mission vs. reward environment? ASHRAE Trans 2012;116(2):9, para. 2.

Scofield M, Weaver T. Using wet-bulb economizers, data center cooling. ASHRAE J 2008;50(8):52–58.

Scofield M, Weaver T, Dunnavant K, Fisher M. Reduce data center cooling cost by 75%. Engineered Syst 2009;51:44–54.

Yury YL. Waterside and airside economizers, design considerations for data center facilities. ASHRAE Trans 2010;116(1):98–108.

26

RACK-LEVEL COOLING AND COLD PLATE COOLING*

HENRY COLES[1], STEVE GREENBERG[1], AND PHIL HUGHES[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[2]Clustered Systems Company, Inc., Santa Clara, CA, USA

26.1 INTRODUCTION

This chapter provides a brief introduction to rack-level cooling as applied to Information and Communication Technology (ICT) equipment support. Rack-level cooling devices are available in a wide variety of designs and capabilities. How these devices fit with existing cooling, the pros and cons of a few common types, its selection, and installation considerations are discussed.

26.1.1 Fundamentals

A data center is typically a dedicated building used to house computer systems and associated equipment such as electronic data storage arrays and telecommunications hardware. For the purpose of brevity, "server" is synonymous with ICT equipment providing services to the end user.

The variety of end use, size, and configuration of data centers is diverse. On one end of the spectrum, there are a fleet of data centers located across the globe that support large social or purchasing networks consuming tens of megawatts each. On the other end of the spectrum, an ICT cooling load may involve a few pieces of electronic equipment consuming 1 kW or less. In this chapter, when we use the term "data center," we are referring to the entire spectrum of ICT equipment installations.

26.1.2 Energy Consumers

Data center energy consumption increased worldwide by 56% from 2005 to 2010. In 2010, data centers consumed approximately 2% (1.7–2.2%) of the electrical power produced in the United States [1]. The rapid increase in data center power consumption is attributed to the increase in the number and use of services available via the Internet.

Data center energy consumption in the United States is counted in tens of billions of kW hours per year, caused in part by a massive amount of ICT equipment operating $24 \times 7 \times 365$. But the energy consumed by just the ICT equipment is not the complete story.

The ICT equipment consists of processor nodes, storage, and networking. This category provides the functionality and services that generate business value. One hundred percent of the electrical energy supplied to this category is turned into heat energy inside the data center.

For a more complete understanding of the energy consumers inside a data center, let's review two additional categories of equipment: power distribution and cooling.

The power distribution equipment provides the electrical power for all equipment in the data center; it provides power to the ICT equipment often in the form of redundant power paths using uninterruptible power supplies (UPSs). UPSs are not 100% efficient; they consume power to keep batteries maintained or keep inertial energy storage devices moving.

It is common to find two or more voltage transformations starting with the power supplied by the utility and ending with the cord power supplied to individual ICT equipment. The voltage transformations are also not 100% efficient with efficiency varying considerably with load.

*This chapter was written by two independent teams in separate sections to give readers two different perspectives for rack cooling technologies. Sections 26.1–26.4 are written by Henry Coles and Steve Greenberg. Sections 26.5 and 26.6 are written by Phil Hughes. Each section represents the section author's view.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

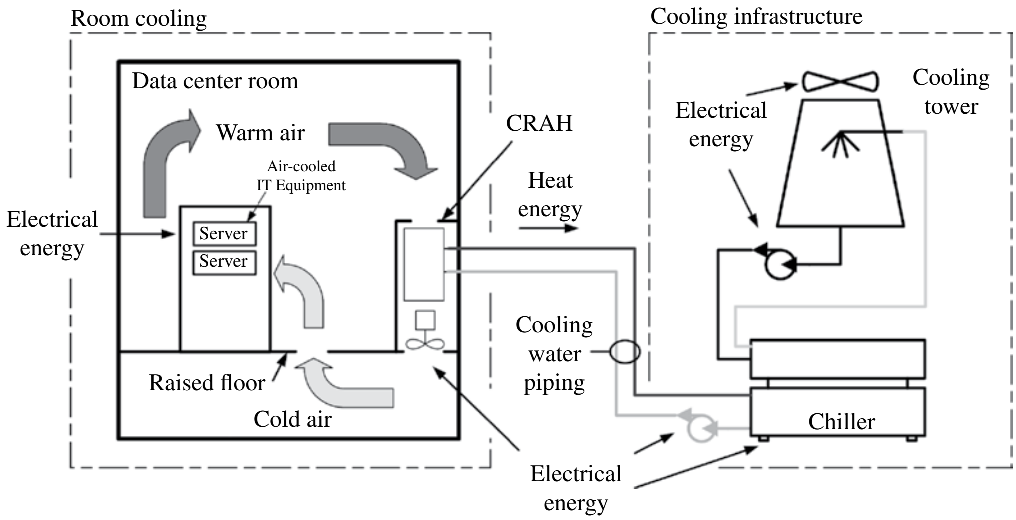


FIGURE 26.1 Data center cooling overview.

The inefficiencies of UPS systems and power transformations combined can be 10–15% of the energy consumed by the ICT equipment; these add to the overall energy requirements. These inefficiencies generate heat, much of which also ends up inside the data center.

The last category of equipment can be grouped together and termed the cooling system. All the heat released inside the data center must be removed and released to the outside environment, often in the form of water evaporation (using a cooling tower).

The total cooling system equipment and processes are split into two categories (Fig. 26.1):

1. Located inside the data center room—Room Cooling
2. Located outside the data center room—Cooling Infrastructure

Rack-level cooling is primarily focused on equipment and processes associated with room cooling. Existing room cooling equipment is typically comprised of equipment such as computer room air handlers (CRAHs) or computer room air conditioners (CRACs). These devices most commonly pull warm air from the ceiling area, cool it using a heat exchanger, and force it out to an underfloor plenum using fans. This method is often referred to as raised floor cooling, shown in Figure 26.1.

The heat from the ICT equipment, power distribution, and cooling systems inside the data center (including the energy required by the CRAH/CRAC units) must be transferred to the cooling infrastructure via the CRAH/CRAC units. This transfer typically takes place using a cooling water loop. The cooling infrastructure, commonly using a water-cooled chiller and cooling tower, receives the heated water from the room cooling systems and transfers the heat to the environment.

26.1.3 Data Center Energy Intensity

As explained, ICT equipment across the United States consumes large amounts of electricity. In addition, the density of the power use inside a data center is higher than almost any other type of building. Here, “density” is defined as the power supplied to the ICT equipment divided by the floor area inside a data center.

The energy use density in data centers has increased dramatically in the last 10 years. For example, traditional densities of 40–80 W/ft^2 (430–860 W/m^2) have given way to data centers required to support operating densities of 600–1000 W/ft^2 (6.5–11 kW/m^2) [2] with even greater annual increases since 2005 [1].

This recent density increase is more easily understood if we look at the change of density on a per-rack basis. Many legacy data centers were designed to support rack energy densities of 1.5 kW per rack. Current requirements for modern ICT equipment can be 20 kW or more per rack.

Adding or replacing legacy ICT equipment at higher densities presents challenges including increased power and cooling requirements. The addition of rack-level cooling with new high-density servers can provide excellent solutions to these challenges.

One can find a wide range of density and configurations across data centers or within a single rack. Rack-level cooling can often be successfully applied while improving overall energy use efficiency.

26.1.4 Data Center Cooling

26.1.4.1 Introduction We have described the sources of heat found inside a data center and the typical means of moving the heat outside.

There is a variety of cooling infrastructure options. In some of these options (e.g., chiller plant), the cooling infrastructure consumes most of the cooling system energy. That said, cooling processes inside the data center room also consume a considerable amount of energy. Note that the energy consumption of the external cooling infrastructure is affected by the efficiency and effectiveness of the room cooling systems.

Rack-level cooling is applied to the heat energy transport inside the data center. Therefore, a brief overview of the typical existing cooling equipment is provided so we can understand how rack-level cooling fits into the overall cooling system picture.

It is important to note that rack-level cooling depends on having a cooling water loop (chilled water or tower water). Facilities without cooling water systems are unlikely to be good candidates for rack-level cooling.

26.1.4.2 Transferring Heat Most of the heat generated inside a data center originates from the ICT equipment. As shown in Figure 26.2, electronic components are kept from overheating by a constant stream of air provided from internal fans.

Commercial ICT equipment is typically mounted in what are termed “standard racks.” A standard ICT equipment rack has the approximate overall dimensions of 24 in. wide by 80 in. tall and 40 in. deep. These racks containing ICT equipment are placed in rows with inlets on one side and exits on the other. This arrangement creates what is termed “hot aisles” and “cold aisles.”

The ICT equipment manufacturer specifies an acceptable range of inlet air temperature that allows the equipment to provide the maximum computing performance. In addition to a specified air temperature range, the equipment is designed for a very low external pressure difference between the inlet and exit.

In the case of low-density racks (e.g., 1 kW per rack), additional air-conditioning capacity is not an issue.

The situation for a moderate-sized data center, consuming, for example, 2 MW of electrical power for the ICT equipment, is more interesting. If the ICT equipment air inlet supply temperature is 70°F and the heated exiting air is 100°F, this provides a commonly found 30°F delta. For this case, the cooling system inside the room should be

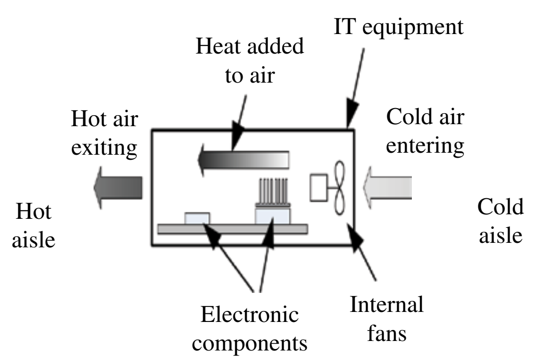


FIGURE 26.2 ICT equipment cooling basics.

capable of moving and cooling 210,000 ft^3/min from 100°F back to 70°F and providing this cool air at very close to zero delta pressure back to the front of all the ICT equipment. This task is the function of the data center room cooling system.

26.1.4.3 Conventional Room Cooling The task of moving the air from the exit of the servers, cooling it, and supplying it back to the inlet of the servers is commonly provided inside existing data center rooms by CRAHs arranged as shown in Figure 26.1. (Note that the room may be cooled by CRACs that are water cooled or use remote air-cooled condensers, but for simplicity, these are not shown in Figure 26.1.)

This air cooling method worked in the past but can pose a number of issues when high-density ICT equipment is added to or replaces existing equipment.

The problems arise when airflow increases are required from existing CRAHs or CRACs. These requirements can be as high as a 10-fold increase at a particular location inside the data center, for example, an increase from 150 cfm per rack (1.5 kW of ICT equipment) to 2000 cfm per rack (20 kW of ICT equipment). These systems were not originally designed to support these airflow rates, nor are their internal heat exchangers adequately sized to remove the increased quantity of heat and transfer it to the cooling infrastructure.

In addition, the space for airflow under the raised floor is often gradually reduced by an accumulation of cables and other equipment in use or abandoned. This reduction in underfloor volume creates additional airflow restriction and exacerbates the problem of inadequate airflow.

26.1.4.4 Conventional Cooling Equipment To understand how rack-level cooling equipment fits into room cooling, a brief listing of the pros and cons of conventional raised floor room cooling by CRAHs or CRACs is provided:

Pros: Cooling can be easily adjusted, within limits, by moving or changing the arrangement of perforated floor tiles.

Cons: Providing a significant increase in cooling at a desired location may not be practical due to airflow restrictions below the raised floor.

Raised floor cooling systems do not supply a uniform temperature of air presented at the ICT equipment inlets across the vertical rack array due to room-level air circulation. Therefore, the temperature of the air under the floor must be colder than it might otherwise be, causing the external cooling infrastructure to work harder and use more energy.

If the existing room cooling systems cannot be adjusted or modified, the additional load must be met via another method, such as with a rack-level cooling solution. In the next section, three common rack-level cooling solutions will be discussed.

26.2 RACK-LEVEL COOLING TYPES

26.2.1 Introduction

In the last several years, a number of technologies have been introduced addressing the challenges of cooling high-density ICT equipment. Before we look at a few common rack-level cooler types, three key functional requirements are discussed:

- Consistent temperature of cooling air at the ICT equipment inlet:
  The solution should provide for a consistent temperature environment, including air temperature in the specified range and a lack of rapid changes in temperature. See the ASHRAE Thermal Guidelines [3] for these limits.
- Near-neutral or slightly higher delta air pressure across the ICT equipment:
  ICT equipment needs adequate airflow via neutral or a positive delta air pressure to reduce the chance of issues caused by internal and external recirculation, including components operating above the maximum temperature limits.
- Minimal load addition to the existing room air conditioning:
  Ideally, a rack-level cooling solution should capture all the heat from the ICT equipment racks it is targeted to support. This will reduce the heat load on the existing room cooling equipment.

There are a few distinct types of rack-level cooling device designs that have been installed in many data centers and proven over a number of years. The description of these designs along with pros and cons is discussed in the following:

- Enclosed
- In-Row™
- Rear door

It should be noted that given the wide variety of situations where these devices might be considered or installed and with newer rack-level cooling devices frequently entering the market, there may be exceptions to the advantages or disadvantages listed.

26.2.2 Enclosed Type

The enclosed design approach is somewhat unique compared to the other two in that the required cooling is provided while having little or no heat exchange with the surrounding area. Additional cooling requirements on the CRAH or CRAC units can be avoided when adding ICT equipment using this rack-cooler type. The enclosed type consists of a rack of ICT equipment and a cooling unit directly attached and well-sealed. The cooling unit has an air-to-water heat exchanger and fans. All the heat transfer takes place inside the enclosure as shown in Figure 26.3. The heat captured by the enclosed rack-level device is then transferred directly to the cooling infrastructure outside the data center room. Typically, one or two racks of ICT equipment are supported, but larger enclosed coolers are available supporting six or more racks. There are a number of manufacturers of this type of rack-level cooler including Hewlett-Packard, Rittal, and APC by Schneider Electric.

Enclosed rack-level coolers require a supply of cooling water typically routed through the underfloor space. Overhead water supply is also an option. For some data centers, installing a cooling distribution unit (CDU) may be recommended depending on the water quality, leak mitigation strategy, temperature control, and condensation management considerations. A CDU provides a means of separating water cooling loops using a liquid-to-liquid heat exchanger and a pump. CDUs can be sized to provide for any number of enclosed rack-level-cooled ICT racks.

26.2.2.1 Advantages The main advantage of the enclosed solution is the ability to place high-density ICT equipment in almost any location inside an existing data center that has marginal room cooling capacity.

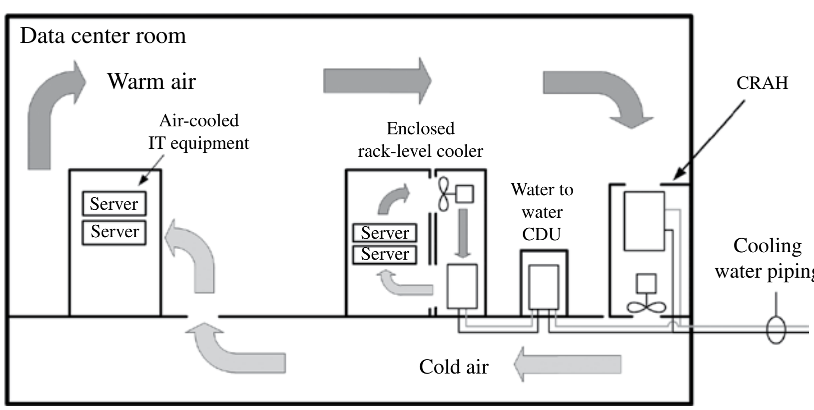


FIGURE 26.3 Side view: Enclosed rack-level cooler installation.

A proper enclosed design also provides a closely coupled, well-controlled, uniform temperature and pressure supply of cooling air to the ICT equipment in the rack. Because of this feature, there is an improved chance that adequate cooling can be provided with warmer water produced using a cooling tower. In these cases, the use of the chiller may be reduced, resulting in significant energy savings.

26.2.2.2 Disadvantages Enclosed rack coolers typically use row space that would normally be used for racks containing ICT equipment, thereby reducing the overall space available for ICT. If not carefully designed, low-pressure areas may be generated near the ICT inlets.

Because there is typically no redundant cooling water supply, a cooling water failure will cause the ICT equipment to overheat within a minute or less. To address this risk, some models are equipped with an automated enclosure opening system, activated during a cooling fluid system failure.

26.2.3 In-Row™ Type

The term In-Row™, a trademark of Schneider Electric, is commonly used to refer to a type of rack cooling solution. This rack-level cooling design approach is similar to the enclosed concept, but the cooling is typically provided to a larger number of racks; one such configuration is shown in Figure 26.4. These devices are typically larger in size, compared to those offering the enclosed approach, providing considerably more cooling and airflow rate capacities. There are a number of manufacturers of this type of rack-level cooler, including APC by Schneider Electric and Emerson Network Power (Liebert brand).

26.2.3.1 Advantages A wide variety of rack manufacturer models can be accommodated because the In-Row™ cooler does require an exacting mechanical connection to a particular model of rack. This approach works best with an air-management containment system that reduces mixing between the hot aisle and cold aisles. Either a hot aisle or cold aisle containment method can be used. Figure 26.4 shows an overhead view of a hot aisle containment installation. Because In-Row™ coolers are often a full rack width (24 in.), the cooling capacity can be substantial, thereby reducing the number of In-Row™ coolers needed. Half-rack-width models with less cooling capacity are also available.

26.2.3.2 Disadvantages The advantage of the ability to cool a large number of racks of different manufacturers containing a wide variety of ICT equipment also leads to a potential disadvantage. There is an increased likelihood that the temperature and air supply to the ICT equipment is not as tightly controlled compared to the enclosed approach.

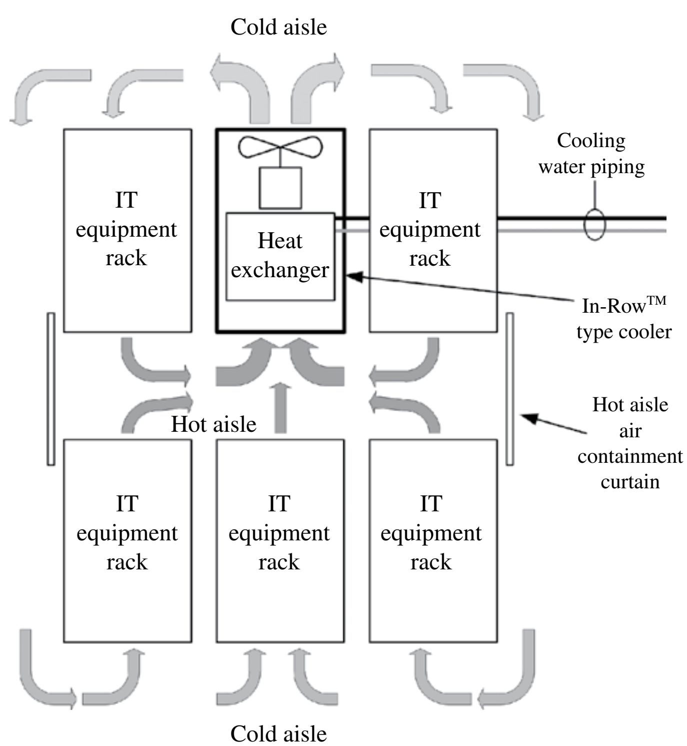


FIGURE 26.4 Overhead view: In-Row™ rack-cooler installation.

26.2.4 Rear-Door Type

Rear-door ICT equipment cooling was popularized in the mid-2000s when Vette, using technology licensed from IBM, brought the passive rear door to the market in quantity. Since that time, passive rear-door cooling has been used extensively on the IBM iDataPlex platform. Vette (now Coolcentric) passive rear doors have been operating for years at many locations.

Rear-door cooling works by placing a large air-to-water heat exchanger directly at the back of each rack of ICT equipment, replacing the original rack rear door.

The hot air exiting the rear of the ICT equipment is immediately forced to enter this heat exchanger without being mixed with other air and is cooled to the desired exit temperature as it reenters the room, as shown in Figure 26.5.

There are two types of rear-door coolers, passive and active. Passive coolers contain no fans to assist with pushing the hot air through the air-to-water heat exchanger. Instead they rely on the fans, shown in Figure 26.2, contained inside the ICT equipment to supply the airflow. If the added pressure of a passive rear door is a concern, "active" rear-door coolers are available containing fans that supply the needed pressure and flow through an air-to-water heat exchanger.

26.2.4.1 Advantages Rear-door coolers offer a simple and effective method to reduce or eliminate ICT equipment heat from reaching the existing data center room air-conditioning units. In some situations, depending on the cooling water supply, rear-door coolers can remove more heat than that supplied by the ICT equipment in the attached rack. Passive rear doors are typically very simple devices with relatively few failure modes. In the case of passive rear doors, they are typically installed without controls. For both passive and active rear doors, the risk of ICT equipment damage by condensation droplets formed on the heat exchanger and then released into the air stream is low. Potential damage by water droplets entering the ICT equipment is reduced or eliminated because these droplets would only be found in the airflow downstream of the ICT equipment. Rear-door coolers use less floor area than most other solutions.

26.2.4.2 Disadvantages Airflow restriction near the exit of the ICT equipment is the primary concern with rear-door coolers both active (with fans) and passive (no fans). The passive models restrict the ICT equipment airflow but possibly not more than the original rear door. While this concern is based on sound fluid dynamic principles, a literature review found nothing other than manufacturer reported data [4] of very small or negligible effects, which are consistent with users' anecdotal experience. For customers that have concerns regarding airflow restriction, active models containing fans are available.

26.2.5 Other Cooling Methods

In addition to the conventional air-based rack-level cooling solutions discussed earlier, there are other rack-level cooling solutions for high-density ICT equipment.

After 2009, a large number of innovative rack-level cooling solutions came to the market, including the following two examples.

In the 2013–2014 time frame, a cooling method commonly termed direct cooling was introduced for commercial ICT equipment. The concept of direct cooling is not new. It has been widely available for decades on large computer systems such as supercomputers used for scientific research. Direct cooling brings liquid, typically water, to the electronic component, replacing relatively inefficient cooling using air. Until recently, this technology was too costly to implement on commercial

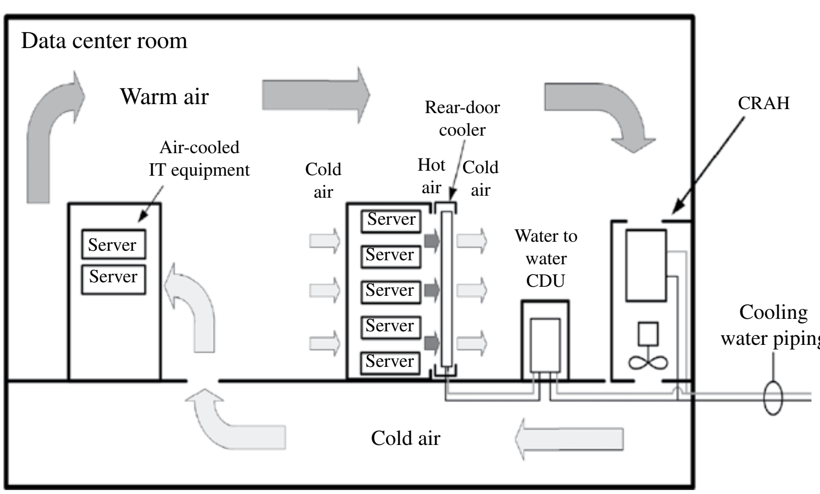


FIGURE 26.5 Side view: Rear-door cooling installation.

ICT equipment, but lower-cost solutions have been recently available from Asetek and CoolIT. These solutions cool high-heat-producing temperature-sensitive components inside the ICT equipment using small water-cooled cold plates or structures mounted near or contacting each direct-cooled component. Some solutions include miniature pumps integrated with the cool plates providing pump redundancy. Transferring heat directly to the facility cooling loop gives direct cooling an overall efficiency advantage. The heat captured by direct cooling allows the less efficient room air-conditioning systems to be turned down or off.

Clustered Systems offers a unique rack-level cooling solution. The heat from electronic components is transferred by conduction to a cold plate that covers the server. This cold plate is kept cold by refrigerant undergoing a phase change. The heat from the refrigerant is moved to the facility cooling loop by way of a refrigerant to water CDU (see Section 26.5).

26.3 RACK-LEVEL COOLER SELECTION AND INSTALLATION

There are many manufacturers and models of rack-level cooling solutions. One size does not fit all, and the best solution may not be initially obvious. This section provides some guidance to assist with the selection process. The data center owner can work with an engineering firm or with cooling device manufacturers directly to select a solution. In either case, some preparation described in the following will greatly improve the efficiency of that process.

Suggested selection process steps are presented as follows:

- Use of existing infrastructure
- ICT equipment and layout
- Facility requirements

26.3.1 Use Existing Infrastructure

When confronted with a change to cooling requirements, it is advised to apply some effort to clarify the problem and attempt to identify cost-effective solutions using existing equipment. For example, consider the following two illustrative cases:

Case 1: 25 kW of new ICT equipment needs to be added to an existing data center and the room air-conditioning systems are currently operating at maximum capacity.

First, explore solutions that use existing cooling systems. Ask the question, can the cooling for the new ICT equipment be provided without substantial cooling system changes?

Areas to consider may include decommissioning outdated or idle ICT equipment, adjusting the flow from the underfloor plenum or making low-cost or no-cost improvements to the existing air management and containment structures in the existing room. These considerations can lead to a low-cost solution that provides the needed capacity [5].

The addition of rack-level cooling solution can then be considered if a low-cost approach cannot be identified.

Case 2: To improve the energy efficiency of an existing data center.

Efficient data center cooling system design is widely known and understood (e.g., ASHRAE TC 9.9 Datacom Series). The barriers to a substantial retrofit or obtaining a new data center are largely economic. As possible modifications are reviewed, the return from the potential energy savings should be compared to other possible investments.

Before considering a substantial investment, first thoroughly investigate no-cost or low-cost solutions using the existing room cooling and cooling infrastructure as mentioned earlier.

26.3.2 ICT Equipment and Layout

A requirement of adding new or replacing existing ICT equipment is reviewing the existing cooling capability. When engaging with an engineering firm or cooling equipment manufacturer regarding cooling capacity issues, one of the first things you will be asked is to provide the layout and thermal details of the ICT equipment involved along with information on the complete existing cooling system.

26.3.2.1 Make/Model/Configuration Consider gathering the following information before approaching the engineering firm or manufacturer:

Make, model, and configuration of all the ICT equipment currently in place and anticipated in the near future.

26.3.2.2 Power Consumption For each piece of ICT equipment, obtain the manufacturer's estimate of the power requirements. This is commonly provided by the ICT equipment manufacturer by way of providing an online tool. If possible, consider measuring or making an estimate of the actual average and peak power draw for planned applications. This power will typically be much lower than information determined by an online tool. The online tools, in general, estimate maximum power consumption. Because of this, it is easy to overestimate additional cooling requirements.

26.3.2.3 Airflow Pieces of ICT equipment have fans that move air over internal components, keeping them within specified temperature limits (Fig. 26.2). The equipment is designed with the assumption that there will be no airflow restriction at the front or rear boundaries. The fan speeds are controlled by software, and the resulting airflow rates will change as a function of a number of factors including the air inlet temperature and the temperature of key electronic components such as CPUs. When this airflow is restricted, recirculation can occur inside or externally, causing higher than anticipated component temperatures. If rack-level cooling systems restrict this flow, temperature warnings are more likely when maximum performance is called for. The server fans may be commanded to speed up in an attempt to correct the effects of the restricted airflow and therefore consume more power. The reaction caused by airflow restriction will vary depending on the ICT equipment and should be kept in mind if wanting to operate at the air inlet temperature for optimum overall efficiency [6].

26.3.3 Facility Requirements

Rack-level cooling solutions transfer heat from the ICT equipment into water or in some cases a refrigerant. Transferring this heat to the outside cooling infrastructure is likely to require facility modifications and additional considerations. For example, consider the following.

26.3.3.1 Water Quality The heat exchangers provided with rack-level coolers may contain smaller passages compared to the existing room-level air-conditioning equipment. Reduced cooling performance may occur because these smaller passages can be more affected by particulate debris buildup or scaling caused by poor water quality.

26.3.3.2 Water Pressure Most rack-level solutions do not contain a pump providing for or assisting the water flow on the facility side. Therefore, it is a good idea to check the delta pressure that your facility will provide considering the piping required to get the water to the rack cooler or a CDU.

26.3.3.3 Condensation Some rack-level solutions contain a condensation management system that sheds liquid condensate when encountering certain combinations of temperature and humidity. In these cases, the facility needs to provide a condensation drain path.

26.3.3.4 CDU A CDU should be considered to address possible issues with poor water quality, temperature control, leak limitation, or condensation management.

26.4 CONCLUSION AND FUTURE TRENDS

Rack-level cooling technology can be used with success in many situations where the existing infrastructure or conventional cooling approaches present difficulties. The advantages come from one or more of these three attributes:

1. Rack-level cooling solutions offer energy-efficiency advantages due to their close proximity to the ICT equipment being cooled. Therefore, the heat is transferred at higher temperature differences and put into a water flow sooner.

   This proximity provides two potential advantages:

   a. The cooling water temperature supplied by the external cooling infrastructure can be higher, which opens opportunities for lower energy use.
   b. A larger percentage of heat is moved inside the data center using water and pumps compared to the less efficient method of moving large volumes of heated air using fans.

Note: When rack cooling is installed, the potential energy savings may be limited if the existing cooling systems are not optimized either manually or by automatic controls.

2. Rack-level cooling can solve hot spot problems when installed with high-density ICT equipment. This is especially true when the existing room cooling systems cannot be modified or adjusted to provide the needed cooling in a particular location.
3. Rack-level cooling systems are often provided with controls allowing efficiency improvements as the ICT equipment workload varies. Conventional data center room cooling systems historically have a limited ability to adjust efficiently to changes in load. This is particularly evident when CRAH [7] or CRAC [8] fan speeds are not reduced when the cooling load changes. However, recent data center control software companies such as Vigilent and Synapsense offer solutions to this problem by providing systems for CRAH or CRAC fan speed control.

As mentioned, new ICT equipment is providing an increase in heat load per square foot. To address this situation, rack-level cooling is constantly evolving with new models frequently coming to the market.

Recent trends in ICT equipment cooling indicate that new products will involve heat transfer close to or contacting high-heat-generating components that are temperature sensitive.

Many current and yet-to-be-introduced solutions will be successful in the market given the broad range of applications, starting with the requirements at a supercomputer center and ending with a single rack containing ICT equipment.

26.5 RACK-LEVEL COOLING USING COLD PLATES

26.5.1 Cooling Fundamentals

The phenomenon that we humans perceive as heat or cold is produced by the motion of molecules. Only at absolute zero (−273°C) do molecules have no motion. As they become more energetic, their temperature is perceived to rise and their state can change from solid to liquid to gas and even to plasma when the molecules themselves shake apart. As energy states increase, the rate of collisions between molecules increases and occasionally a photon is knocked off, causing the phenomenon of radiation. At lower-energy levels, radiation is in the infrared part of the spectrum, increasing into the visible and beyond at higher-energy levels.

The first law of thermodynamics holds that energy cannot be created nor destroyed but may change form. It is one of these changes that create our server-heating problems. Electrical energy arrives in a chip as a flow of electrons that bang into molecules and start them moving faster, producing heat. Those molecules must be slowed down enough (cooled) to avoid damaging the chip.

The second law of thermodynamics holds that when two systems are allowed to interact, they will achieve an energy equilibrium, that is, energy will flow from the more energetic to the less energetic system. The question, therefore, is: What is the best transfer mechanism to remove excess energy? We can choose from radiation, convection (forced or natural), conduction, and phase change.

26.5.2 Radiation

At the time of writing, most electronics are solid state, so we can assume that our high-energy system is a solid. The lower-energy system surrounding it could be a vacuum, gas, liquid, or another solid.

With a vacuum, the only way for energy to escape the first system is through radiation. According to Stefan–Boltzmann, the energy radiated by a black body is defined by $q = \sigma T^4 A$

where q = watts, $\sigma = 5.67 \times 10^{-8}$ (W/m^2 K^4) = the Stefan–Boltzmann Constant, T = absolute temperature, and A = body area in square meters.

Grinding though that lot, assuming a 33 mm × 33 mm chip package at a temperature of 70°C, we conclude that we can dissipate only 0.75 W through radiation with a perfect black body and surroundings at absolute zero, definitely insufficient.

26.5.3 Conduction

A gas is one step up from a vacuum. There are about 2.7×10^{22} molecules in a liter of air. Those molecules, if packed together at absolute zero, would occupy only 4.7×10^{-8} l. Not surprisingly, thermal conductivity, k, is only 0.028 W/m-K at room temperature. For every watt removed from the 33 mm × 33 mm chip, there would be a temperature difference of 800°C/in. (25 mm) of air between the hot chip and the cold body.

Water, one of the more popular coolants, has 3.3×10^{25} molecules/l, over a thousand times denser than air. Naturally, this implies a higher conductivity, 0.58 W/m-K, 20 times higher than air, dropping the temperature difference to 40°C/in.

Aluminum has 6.02×10^{25} molecules/l. Its conductivity, is 205 W/m-K, 350 times that of water. The one inch temperature gradient is just 0.11°C, over 7000 times better than air.

Clearly, Aluminum or other high-conductivity metals such as copper win hands down for conductivity. The molecules are trapped in a crystalline matrix where they vibrate and pass energy to all their neighbors. Liquid, on the other hand, is almost as dense but the molecules move freely (which is good for convection), but they don't readily pass on their energy to other molecules. In a gas, the molecules are so few they rarely collide, reducing conductivity even more.

26.5.4 Natural Convection

This type of convection occurs in both liquid and gas. When the fluid is heated, its molecules closest to the heat source become more energetic and tend to move above their less energetic neighbors. What we observe is that a portion of the fluid expands and rises to the top.

26.5.4.1 Air To compute the heat removed by natural convection in air, no less than 14 parameters must be taken into account. Even then, some are approximations or simplifications by worthies from centuries past such as Rayleigh, Reynolds, Prandtl, Nusselt, and Grashof.

Fortunately, there is a simplification; thus, h (heat transfer coefficient) = $C*[(T1 - T2)/L)^n$ = 3.77 W/m^{2}K.

C and n are dimensionless coefficients, which can be assumed to be 0.59 and 0.25, respectively. $T1$ and $T2$ are the temperatures of the hot body and cold plate, respectively. L is the distance between the hot body and a cold plate, 25 mm in this example.

Thus, for our 33 mm × 33 mm CPU, the gradient would be 5.8°C/W.

Conclusion: Natural convection may work well for lower-power chips (<5 W).

26.5.4.2 Liquid In systems using natural liquid convection, a fluid with a very high buoyancy-to-viscosity ratio is required. This can be expressed as the Grashof number, which should be as high as possible:

$$\text{Gr} = \text{buoyancy force / viscous force} = g \cdot \beta \cdot \Delta T \cdot L^3/v^2$$

where g = gravitational acceleration, β = volumetric expansion coefficient, L = characteristic length, 0.1 m, ΔT = temperature difference between vertical plane and fluid 30°C, and v = kinematic viscosity.

Typical Grashof numbers with aforementioned length and temperature parameters are as follows:

Fluorinert 3283 = 1.88×10^{10}, Florinert 70 = 5.19×10^5, and Mineral oil 1.01×10^5

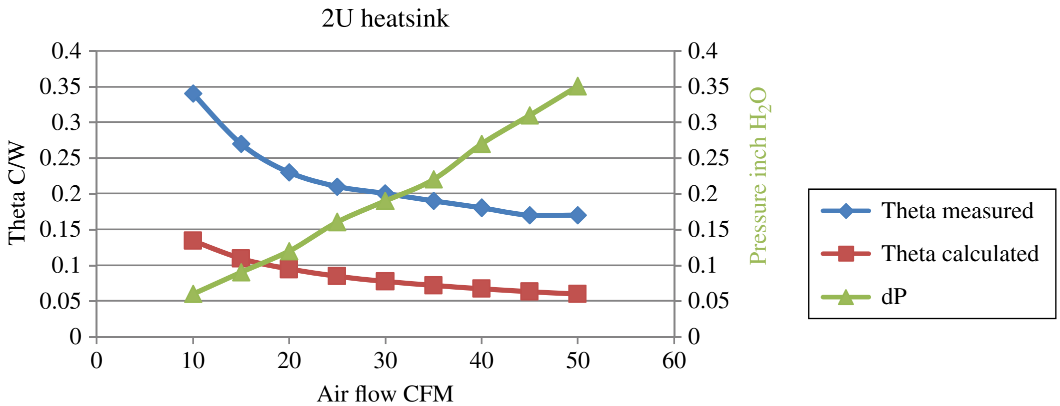


FIGURE 26.6 Heat sink characterization.

Conclusion: Fluorinert 3283 or similar with high Grashof numbers should work quite well in systems where a system board is immersed in fluid and has a cooling surface a few millimeters away. Iceotope is the only company with such a solution currently.

26.5.5 Forced Convection

Both gases and liquids can be used in forced convection systems. We will only discuss air and water in this context. Air is generally ducted to where its cooling effect is required, but water is tightly constrained in piping and heat exchangers.

26.5.5.1 Air While the number of parameters required to derive the heat transfer coefficient grows to about 18, there are some simplifications that can be used for sanity checks. One of the simplest for air at standard temperature and pressure that can be used for heat sinks is as follows: Theta (Θ) = $916^*(L/V)^{0.5}/A$ in W/°C

where L = heat sink length in inches, V = air velocity in ft/min, and A = total surface area in square inches.

However, this and even more sophisticated models are no substitute for in situ measurement. Figure 26.6 shows a representative example of the (very significant) difference between the datasheet values and those derived using the heat sink calculator provided on some manufacturers' websites. The derived values use exactly the same values as the aforementioned formula. Note the approximately 2× difference between curves; dP shows the pressure drop required to achieve the stated airflow.

Adding to the complexity is the variation between servers. The same heat sink will perform differently as the externalities vary. These include ducting, positioning of each CPU (if more than one), DRAM, and VRM layouts.

The other significant factor is the fans' specifications. They must be capable of providing sufficient volume and pressure to drive the air through the heat sink(s) and not consume too much power in doing so. To establish the operating

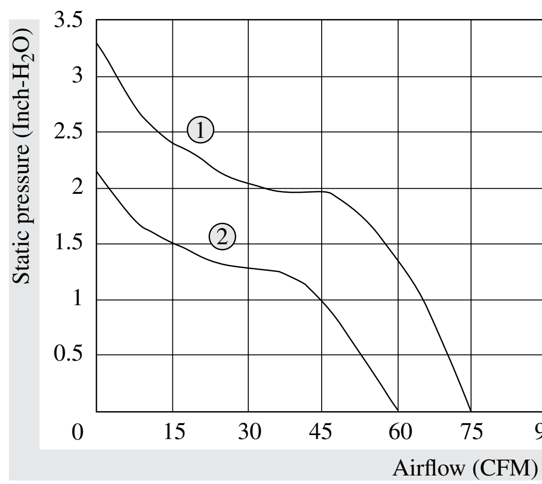


FIGURE 26.7 Fan curves 60 mm × 60 mm.

requirements, we need to look at the maximum allowable CPU lid temperature and CPU power. Typically, 70°C has been the allowable maximum, but excursions up to 95°C may be allowed in the future. Maximum power for high-performance CPUs is commonly 130 W (even though most servers may be equipped with only 95 W CPUs).

Assuming that the maximum operating inlet temperature is 45°C, we'd have a margin of 25°C. Thus, the allowable thermal resistance would be 25/135 = 0.185°C/W. As can be seen from the graph (Fig. 26.6), that is the maximum capability of the heat sink. At that point, the fans must deliver 50 CFM with a static pressure of 0.35 in. of water.

Figure 26.7 shows a typical set of operating curves for two fans. When operating at maximum power, they should be at the inflection point, delivering 30–40 CFM.

Typically, a 2U server heat sink is about 3.5 in. wide and 2.5 in. tall. Banks of DRAMs will be deployed on one or both sides of the CPU (Fig. 26.8). In the case of the half-width board on the left, there is room for only two fans, mandating the choice of the more powerful fan 1. These will draw 60 W.

FIGURE 26.8 Intel memory cooled blade and rack CDU rear. Courtesy of Intel Corporation and Asetek Inc.

Further, at least 50% of the air will bypass the heat sinks, producing borderline performance in normal operation. A fan failure will cause the CPU to throttle in order to stay within the thermal envelope and thus lose performance. The system on the right is a little more forgiving, but a fan failure still has the potential to affect performance. Potentially, its fans could draw up to 150 W—an additional 30% load.

As the power consumed by a fan is proportional to the volume of airflow (CFM) cubed, from an energy-efficiency point of view, it is better to have as many fans as possible. For example, if one fan could produce adequate airflow for cooling at 32 W, two of the same fans sharing the load would only consume 8 W. Note that the energy of the fans adds slightly to the air temperature, but is usually low enough (<1°C) so as not to be a significant factor.

After the heat is exhausted from the server, it is either sucked into a cooling unit, which is itself cooled by water or pumped refrigerant, and then recirculated to the server inlets or exhausted to the atmosphere. In the latter case, fresh outside air is directed to the server inlets. For a rack with 80 server motherboards (left motherboard) drawing 450 W each, for a component load of 36 kW and a typical fan load of 6 kW (75 W/server), approximately 445,000 ft^3 of air (12,600 m^3) needs to be recirculated with its fans to maintain a 10°C air temperature rise at the server exits.

It should be noted that the external environment can also affect fan performance. Passive rear-door heat exchangers and cabling are the two biggest problems. They can block server exhaust and reduce cooling efficiency.

26.5.5.2 Water Water is much easier to handle than air. It is piped exactly to where you want it to go. Most systems consist of three components, in-server, in-rack, and exhaust. In all known systems, the in-server component connects to the in-rack distribution system via two quick connects.

They also come in two flavors, IBM and everybody else. The IBM version is very solidly engineered with all cooling components connected with brazed copper tubing. In Figure 26.9, it

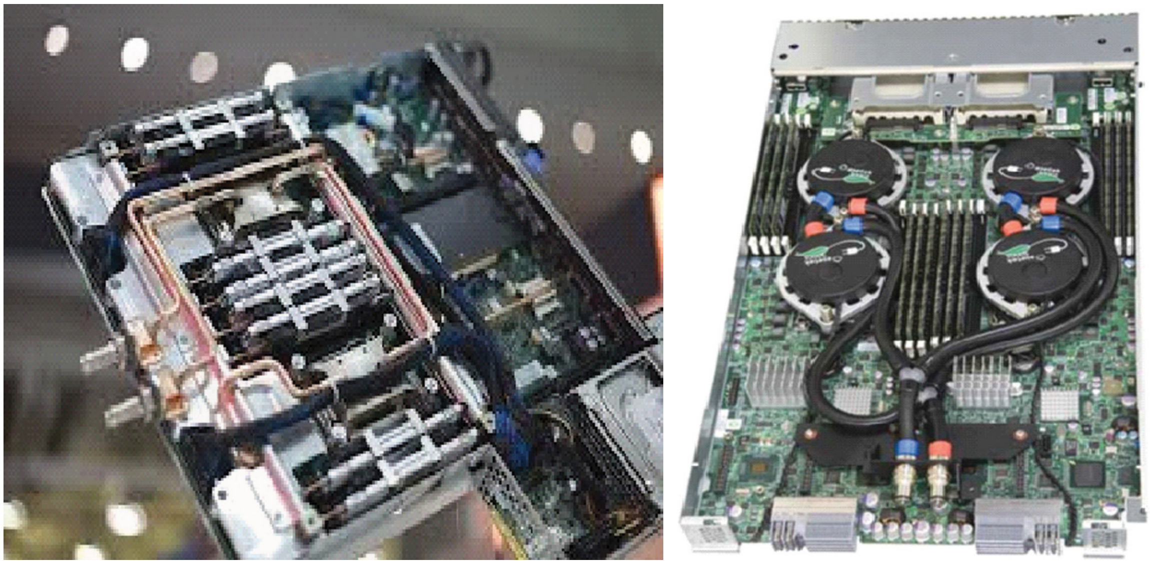


FIGURE 26.9 Water-cooled servers. Courtesy of IBM Corporation and Asetek Inc.

FIGURE 26.10 Water pipes and manifolds. Courtesy of Asetek.

can be seen that each hot component has an individual cooling block. Very little if any air cooling is required.

The representative of the “others” cools only the CPUs and is interconnected with flexible tubing and plastic connectors. Air cooling is still required for all other components including DIMMs.

Figure 26.10 shows the rack-level plumbing of a typical water-cooled system.

Most of these systems are advertised as having the ability to be cooled with hot water, and they do remove heat quite efficiently. The block in contact with the CPU or other hot body is usually copper with a conductivity of around 400 W/m-K, so the temperature drop across it is negligible. If the water is pumped slowly enough, reducing pumping power, flow is laminar. Because water is not a very good conductor of heat, a temperature drop of around 5°C can be expected across the water–copper interface. This is usually negligible but, if necessary, can be reduced by forcing turbulent flow by increasing flow rate. This could be an expensive waste of energy.

Both server types have two CPUs plumbed in series. The maximum power consumption of a CPU is around 130 W. If we assume that the maximum lid temperature is 70°C and the inlet water is 40°C, each CPU could heat the water 10°C while accommodating the thermal resistance of the water film and the cold block itself. For a rack with 40 servers, 160 CPUs (21 kW), about 1.8 m^3 of water per hour would be required. Pump energy would be around 80 W. Of course, another 15 kW (450 W total per server) remains to be removed by fans. Clearly, the racks cannot be deployed at maximum density, resulting in a power density of around 600 W/ft^2, without special provision such as rear-door heat exchangers.

While the physics of the system are workable, the statistics may not be. Let’s be very optimistic and assume that the mean time between failure (MTBF) of a liquid connector is 10^7 h and the service life is 3 years, that is, 26,280 h. The probability of survival is $e^{-(26,280/10-7)} = 0.9974$, or there is a 0.26% probability that it would fail. If there were 1000 servers, 2000 connectors, about 5 would fail. This calculation would be reasonable for the IBM system where all the connectors are brazed to the piping. Where flexible tubing and plastic connectors are in the mix together with the vibration of fans, then the probability of failures goes up.

Finally, water chemistry can be difficult. Described as the "universal solvent," it can eat through metals and plastic if it has not been pretreated properly. Another concern could be algae growth. A closed secondary loop to the components is essential to reliably manage such issues. A leak in such a loop might bring the entire loop and its associated servers down.

26.5.5.3 Oil Light mineral oil has been applied in a couple of instances for cooling. In one case, multiple servers are immersed in an oil bath, and in the second, servers are put into individual sealed cases. In both cases, the oil is forced through the individual server containers using circulation pumps. Heat is removed from the oil by passing it through a heat exchanger on a water loop.

Typical parameters for light oil are (water in parentheses) 800 kg/m^3 (1000); viscosity, 0.0345 N-s/m^2 (0.000798); specific heat, 1100 J/kg-K (4186); thermal conductivity, 0.15 J/s/m-K (0.000615); thermal expansion coefficient, 0.00064 K^{-1}(0.000291); and Grashof number, 1.01×10^5 (1.34×10^8).

This scheme is more energy efficient than air but suffers from two disadvantages. Servicability can be a problem when the system boards are covered in an oil film and more energy is required to drive the circulation pumps than a water-based system due to the lower specific heat of the oil and higher viscosity. Ride through might also be an issue as the oil has a fairly low Grashof number and specific heat so there would be little natural circulation if a pump failed. This may cause overheating.

26.5.6 Phase Change

Phase change-based systems use the latent heat of evaporation to absorb heat and remove it from the hot objects.

In one case, servers are placed in a bath open to the atmosphere and filled with cooling fluid with a relatively low boiling point; in the other, the coolant is delivered to the server through a cold plate in a sealed system.

26.5.7 Bath

A coil with coolant, usually water or water and glycol, circulating through it is mounted in the lid of the bath. In operation, the liquid boils, the gas rises and is recondensed by the cooling coil, and the liquid drops back into the bath.

Originally designed for single-phase sealed systems, fluids such as 3M's Novec 7000, boiling point of 34°C, and Novec 649, 49°C, at normal atmospheric pressure are being proposed for nonsealed systems. While Novec 7000 has the best physical characteristics, such as latent heat of evaporation and boiling point, it has a significant global warming potential (GWP), which may be a problem in some jurisdictions. On the other hand, Novec 649 has an uncomfortably high boiling point, which may compromise reliability of some components, but has a very low GWP.

These and other similar fluids have been used for cleaning for years with no apparent harm to operators as the liquids are always below boiling point. Presumably, most would have evacuation hoods over the cleaning baths so inspiration is minimized. Precautions will be needed where operatives run the risk of continuous exposure to the additional vapors released by the boiling fluids until long-term effects are understood. Additional precautions might be necessary to guard against failure of the cooling loop, which could cause the room atmosphere to become saturated with coolant vapor.

26.5.8 Sealed System

In the open system, the fluid is directly in contact with the hot objects and is insensitive to system topology and component height. In a sealed system with flat, minimally flexible cold plates, heat must be brought up to a single plane. While convection is adequate for low-wattage components, a conductive path is required for high-power devices.

In the implementation available from Clustered Systems, heat is conducted to a single plane by a series of heat risers placed atop each component that generates a significant amount of heat. In all cases, these include CPUs, VRMs, DIMMs, and system glue, plus, if merited, networking and other components generating over approximately 2 W. The heat risers can be seen at the top of Figure 26.11. For clarity, only the bottom server is shown covered by a cold plate. The cold plates are a chassis component and are all permanently soldered into refrigerant distribution manifolds. This completely eliminates the probability of leakage from connectors.

Liquid (R134A) is pumped through cold plates placed upon heat risers attached to CPUs, DIMMs, VRMs, etc. The heat causes the liquid to boil, absorbing 93 times as much heat as the same weight of water.

The liquid and gas mix is then passed to a heat exchanger where it is reconverted to 100% liquid. Unlike air-cooled

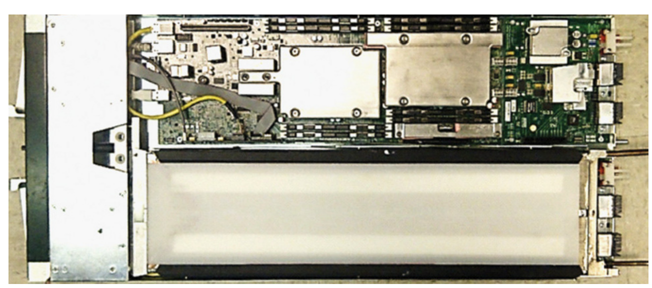


FIGURE 26.11 Clustered systems' phase change-cooled blade with two half-width servers. Courtesy of Clustered Systems, Inc.

systems, the thermal resistance between heat source and liquid is so small that high coolant temperatures can be tolerated. No chiller is required in most cases. The only energy required is for circulation pumps and external fans in a dry or adiabatic cooler. The cooling PUE can be as low as 1.03.

Figure 26.12 shows the front of the chassis. The cold plates can be seen at the right of each nonpopulated slot. They slip into the blade and contact the heat risers when the blade is inserted.

FIGURE 26.12 Sixteen-blade chassis. Courtesy of Clustered Systems, Inc.

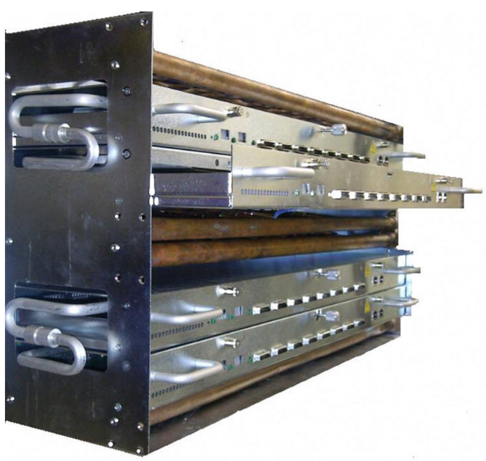


FIGURE 26.13 Chassis rear. Courtesy of Clustered Systems, Inc.

Figure 26.13 shows the four rear switch blades and a partial view of the distribution manifolds.

The maximum power consumption of a CPU is around 130W, and we assume that the maximum lid temperature is 70°C. As the system is isothermal, the cold plate is the same temperature virtually everywhere. Heat input just causes liquid to change to gas with no temperature rise. Assuming that the inlet refrigerant was 40°C and having established by measurement that the thermal resistance from CPU lid to refrigerant is <0.2°C/W, the CPU lid would reach 66°C ($40+130\times0.2$). Because the gasification causes bubble formation, hence turbulence, laminar flow film formation is not a problem.

For a whole rack with 160 servers (72kW at 450W per server), about 0.66 m^3 of refrigerant per hour would be required. In practice, with viscosity of refrigerant being 25% and fluid flow being 10% that of a water-based system, pump energy is very low, about 30W.

The benefits of such an efficient phase change cooling system are striking:

- Very high-power densities can be achieved.
  - 100kW racks enable data center density of 4000W/ft^2.
- Rack floor space for a 10MW data center can be reduced from 50,000 ft^2 to about 2,500 ft^2.
- Data center construction and facility costs drop approximately 50%.

26.6 CONCLUSIONS AND FUTURE TRENDS

Whatever liquid cooling technology is chosen, it will always be more efficient than air for two reasons. The first and most important is that the amount of energy required to move air will always be several times greater than that to move a liquid for the same amount of cooling.

Table 26.1 illustrates some typical numbers. While the move to water reduces the energy by 50%, going to refrigerant cuts it by 90%.

TABLE 26.1 Cooling analysis

	Density (lb/cu ft)	Specific heat (BTU/lb)	States	ΔT (°F)	lb/min/ton	CFM/ton	Static press (PSI)	Req. Watts	Fan/Pump efficiency (%)	Total (W)	% of load
Air (std day)	0.075	0.205	Gas–gas	18	54.11	722	0.036	84.6	30	282	8.0
Water @ 50°F	62.45	0.998	Liquid–liquid	9	22.27	0.36	35	40.6	30	135	3.8
R134a @ 95°F	72.94	72	Liquid–gas 30%	0	9.26	0.13	20	8.3	30	28	0.8

REFERENCES

[1] Koomey JG. Growth in data center electricity use 2005 to 2010. A report by Analytics press, completed at the request of The New York Times; August 1, 2011. Available at http://www.analyticspress.com/datacenters.html. Accessed on May 23, 2014.

[2] Rasmussen N. Guidelines for specification of data center power density. West Kingston: Schneider Electric ITB; 2005. APC White Paper #120.

[3] ASHRAE Technical Committee 9.9. Thermal guidelines for data processing environments–expanded data center classes and usage guidance. Whitepaper, 2011. Available at http://tc99.ashraetcs.org/. Accessed on July 28, 2014.

[4] Coolcentric. Frequently asked questions about rear door heat exchangers. Available at http://www.coolcentric.com/info_center/frequently-asked-questions.php. Accessed on May 23, 2014.

[5] Bell GC. Data center airflow management retrofit technology case study bulletin. Berkeley: Lawrence Berkeley National Laboratory; September 2010. Available at http://hightech.lbl.gov/documents/data_centers/airflow-doe-femp.pdf. Accessed on May 23, 2014.

[6] Moss D. Data center operating temperature: what does Dell recommend? Round Rock: Dell Data Center Infrastructure; 2009.

[7] Coles HC, Greenberg SE, Vita C. Demonstration of intelligent control and fan improvements in computer room air handlers. Berkeley: Lawrence Berkeley National Laboratory, LBNL-6007E; November 2012. Available at http://eetd.lbl.gov/sites/all/files/publications/control-and-fan-improvements-in-crahs.pdf. Accessed on July 28, 2014.

[8] Greenberg S. Variable-speed fan retrofits for computer-room air conditioners. Berkeley: The U.S. Department of Energy Federal Energy Management Program, Lawrence Berkeley National Laboratory; September 2013. Available at http://www1.eere.energy.gov/femp/pdfs/dc_fancasestudy.pdf. Accessed on May 23, 2014.

FURTHER READING

Aquasar. Available at http://en.wikipedia.org/wiki/Aquasar. Accessed on June 17, 2014.

ASETEK. Faster, Denser, Greener, Quieter Servers & Data Centers. Available at http://www.asetek.com/data-center/data-center-coolers.aspx. Accessed on June 17, 2014.

ASHRAE Technical Committee 9.9. Mission critical facilities, technology spaces, and electronic equipment. Available at http://tc99.ashraetcs.org/. Accessed July 28, 2014.

Bright Hub Engineering. Natural Convection Heat Transfer Coefficient Estimation Calculations. Available at http://www.brighthubengineering.com/hvac/92660-natural-convection-heat-transfer-coefficient-estimation-calculations. Accessed on June 17, 2014.

Clustered Systems Company, Inc. Available at www.clusteredsystems.com. Accessed on June 17, 2014.

Hewitt GF, Shires GL, Bott TR. *Process Heat Transfer*. Boca Raton: CRC Press; 1994. Made in IBM Labs: IBM hot water-cooled supercomputer goes live at ETH Zurich.

Stefan–Boltzmann Law. Available at http://en.wikipedia.org/wiki/Stefan-Boltzmann_Law. Accessed on June 17, 2014.

27

UNINTERRUPTIBLE POWER SUPPLY SYSTEM

Chris Loeffler and Ed Spears
Eaton, Raleigh, NC, USA

27.1 INTRODUCTION

Uninterruptible power supplies (UPS) are an extremely important part of the electrical infrastructure where high levels of power quality and reliability are required. In this chapter, we will discuss basics of UPS designs, typical applications where UPS are most commonly used, considerations for UPS selection, and other components or options that are an important part of purchasing and deploying a UPS system.

27.1.1 What Is a UPS?

Put simply, a UPS is a device that provides backup power when utility power fails or becomes unusable by devices requiring regulated electricity to operate. The UPS can either provide electricity long enough for critical equipment to shut down gracefully so that no data is lost or no process is interrupted or long enough to keep required loads operational until another electrical generating source (typically a generator) comes online. Some of the different UPS topologies also provide conditioning to incoming power so that all-too-common sags and surges don't damage sensitive electrical and electronic gear. UPS systems are designed to integrate easily into the standard electrical infrastructure, so that means smaller power requirements are typically single-phase designs, with larger power requirements being handled by three-phase systems. In North America, the typical single-phase UPS design is smaller than 25 kVA, while three-phase systems start around 8 kVA and go up into the MVAs. In some counties in Europe, all systems larger than 8 kVA must have a three-phase input to make sure the utility mains' electrical system stays balanced. Single UPS systems come in sizes ranging from 300 VA (enough power for a typical PC and monitor) to over 2 MVA (enough power for 175 homes), with larger systems being able to be installed in parallel for power levels as high as +20 MVA (enough power for a small town).

27.1.2 Why a UPS?

In this age of critical computing systems and the Internet, business continuity requires that you protect your IT infrastructure from all the hidden threats of the typical facility environment. Even in today's manufacturing environments, power disruptions can cost businesses thousands of dollars in lost revenue, not including the lost productivity of their workforce. Every business, no matter how small or large, is at risk from internally or externally generated power abnormalities.

You may only notice power disturbances when the lights flicker or go out, but your compute, storage, network, and process equipment can be damaged by many other power anomalies that are invisible to the human eye, which can lead to degraded equipment performance or premature failure over time.

So you can worry now or worry later. One choice is proactive, while the other potentially painful. IT systems are at risk even in the largest data centers. Of the 450 Fortune 1000 companies surveyed, each site suffered an average of nine IT failures each year. About 28% of these incidents were caused by power problems.

According to Price Waterhouse research, after a power outage disrupts IT systems:

- 33 + % of companies take more than a day to recover.
- 10% of companies take more than a week to fully recover.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

- It can take up to 48 h to reconfigure a network.
- It can take days or weeks to reenter lost data.
- 90% of companies that experience a computer disaster and don't have a survival plan go out of business within 18 months.

Downtime is costly. Your IT hardware may be insured, but what about the potential loss of goodwill, reputation, and sales from downtime? Consider the number of transactions or processes handled per hour, and multiply that by the value of each one and the duration of an anticipated power incident. Add the delays that inevitably occur when rebooting locked-up equipment, restoring damaged files, and rerunning processes that were interrupted. Then add the cost of lost revenue from being disconnected from your suppliers, business partners, and customers.

Could your business absorb the cost of an extended power outage or IT failure? According to the U.S. Department of Energy, when a power failure disrupts IT systems:

- 33% of companies lose $20,000–$500,000.
- 20% of companies lose $500,000–$2 million.
- 15% of companies lose more than $2 million.

27.2 PRINCIPLE OF UPS AND APPLICATION

27.2.1 UPS Basics

UPS designs get a base classification by the actual energy storage/delivery method used. There are two general classifications: static and rotary. The most popular design in the IT industry, the static UPS, uses some type of electronic switching components that take the stored energy (battery typical) and convert the direct current (DC) to alternating current (AC), at the correct voltage to be used by the downstream critical equipment. The rotary UPS uses a rotating device (generator) that is typically powered by the Utility AC through some type of motor system. The rotary generator, sometimes labeled Diesel Rotary UPS (DRUPS), typically uses a heavyweight flywheel assembly to store energy, which allows the generator section time to start, and then provide power when AC utility power is lost.

UPS systems come in different input and output voltages based on application and on the countries where they are deployed. In North America (United States and Canada), single-phase UPS systems designed to plug into standard wall receptacles come in 120V input. Systems deployed where the electrical contractor pulls a specialty receptacle can come in 120, 208, or 240V. In some of the Caribbean islands and other countries where Europe influenced the electrical system infrastructure, the standard single-phase voltage is 220, 230, or 240V. Mexico and other parts of Central America use 127V for their standard single-phase distributed voltage. UPS for three-phase applications are typically manufactured for 208 Y/120, 220 Y/127, 480 Y/277, and 600 Y/347V for North America and 380 Y/220, 400 Y/230, and 415 Y/240V for the rest of the world. However, some North American data centers are now deploying 415 or 400V products to operate the IT loads close to their maximum value, driving up the efficiency of the power supplies. In addition, the entire data center benefits as this eliminates the needs for 480 or 600V to 208V transformers, typically gaining another 1–3% of efficiency. Since almost all IT applications require an input voltage less than 250V, systems with higher output voltages use transformers to reduce the voltage to an acceptable level. There are a few latest-generation IT power supplies that can handle 277V, which may become more standard as data center power use has become a large expense and there is tremendous focus on more efficient systems.

27.2.1.1 UPS Components and Subsystems

UPS include a number of different individual subsystems based on the system type. This section will cover most of them with some basic information on their function in the system.

Inverter All static UPS systems include an inverter, which uses the DC or backup energy source and creates an AC waveform for the connected load equipment. Inverter designs vary greatly based on the type of system, typically based on criticality of the system and its cost. Small low-cost systems may use power transistors or MOSFETs, which typically output a basic square wave or modified sine wave. Care must be taken when applying these lower-cost inverter designs, as the nonsinusoidal output may cause a negative interaction between the UPS and load power supply, which could result in an inoperable system. Higher-cost systems typically use devices called insulated gate bipolar transistors (IGBTs), which are used with an inverter output filter to create an almost perfect sine wave output. These inverters typically switch the IGBTs on and off thousands of times per second, in a sequence called pulse width modulation (PWM, Fig. 27.1). As seen in the figure, the pulses at the beginning and end of each half-cycle have very short "on," times and longer "off" times, with the "on" time increasing in duration to the peak of the sine wave and then again decreasing as the waveform decreases. The longer the device stays "on," the more the energy delivered to the filter network that is used to create the sine wave output. Advances in high-power IGBTs allow switching frequencies of inverters typically less than 50 kVA in size to be above the human audible range (18 kHz), therefore reducing UPS operational noise. Latest-generation double-conversion UPS systems may use a three-level inverter design, which doubles the number of IGBTs per phase to allow for lower-voltage-rated devices to be used in a series relationship. This helps to raise the efficiency of the inverter by typically 1–3% over typical two-level inverter designs, therefore reducing the overall cost of operating the UPS.

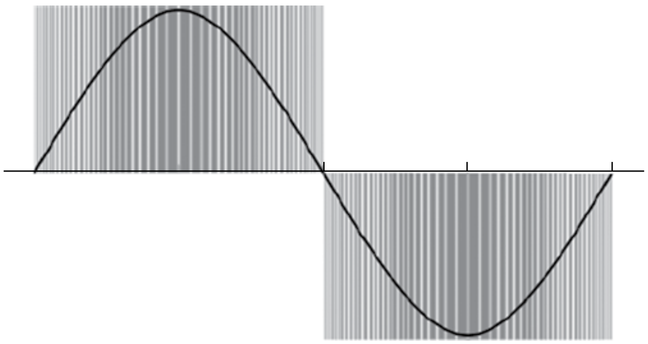


FIGURE 27.1 PWM switching pattern is filtered into clean sine wave AC output waveform. Courtesy of Eaton.

Rectifier Double-conversion and multimode UPS systems include a subsystem known as a rectifier. The rectifier is the first device that sees the input voltage from the utility mains, and it converts the available AC power to a DC voltage for use to power the inverter and charge the batteries. Typically, large UPS used Silicon-Controlled Rectifier (SCR) semiconductors to perform this AC to DC conversion. These devices were rugged and easy to control electronically. However, today, many UPS use IGBT transistors in place of SCRs. This allows the UPS rectifier to be controlled such that its input current is sinusoidal and its input Power Factor is near 0.99. Both of these traits allow for easy interface with an on-site generator, and they eliminate the large costly "low-harmonic LC" input filter required for SCR-based designs.

UPS Logic Control Every UPS system has a main controller, which takes in multiple inputs and makes adjustments or changes operational modes when required. All large UPS systems have eliminated most analog control functions and use digital control algorithms coded into the systems firmware. The logic control may also contain external communication capability to talk to externally connected devices or may be responsible to send information to other communication-only controllers that are externally connected for purposes of monitoring and control.

DC-to-DC Converter The purpose of the DC-to-DC converter is to provide proper battery charging voltage for systems with either a higher or lower DC link ("DC Link" is the physical link between the rectifier output and the inverter input) voltage than that of the battery or backup energy source. Many of today's transformerless UPS designs rectify the incoming AC voltage to a level that is advantageous to ensuring the highest total system efficiency. This voltage level may not be a level that works well with the DC energy storage, so the DC-to-DC converter converts the DC link voltage to the levels needed to charge the battery, and when AC power is lost and the battery is called on, it converts the battery voltage to the levels needed for the inverter to provide the correct output.

Static Bypass Switch The static bypass is typically found on most double-conversion UPS systems, some higher-powered single-conversion systems, and rotary systems. The switch's job is to supply a direct path for the utility mains to power the connected load equipment; however, the mains voltage must be within a certain tolerance range, typically ±10%. There are usually several reasons why the switch is used. First, it may be used for maintenance of the system, activating the switch to ensure that the utility and UPS output voltages are synchronized so a maintenance bypass breaker can be closed. The second is to quickly transfer power to the utility source if for some reason there is a large output overload or internal fault in the UPS itself. Larger UPS systems include a separate power supply and fans for the static bypass to raise the fault tolerance of the system. Another way the static switch is used is to be the main power path for high-efficiency modes in multimode UPS systems. Static switches are built using SCRs, as they are much faster (1–10 ms) to close than mechanical-type contactors (50–100 ms). They need to be able to close within a few milliseconds (ms) to make sure the load equipment does not see a large disruption in mains power. Some UPS designs use both SCRs and a mechanical contactor in parallel, as the SCRs make the quick transition and the contactor closes to then provide unlimited bypassing. This type of static bypass is known as a "momentary static switch," and most times, it is not rated to operate continuously, and if it had to operate under full load, it would fail. Latest-generation multimode systems must use fully rated SCRs as they need to be able to react by either turning on or off very quickly to make sure the load does not see any interruption in power, as utility mains voltages fluctuate or fail.

Maintenance Bypass The maintenance bypass is installed to make sure the UPS can be taken off-line to perform service work including repairs, preventive maintenance, and upgrades. There are a number of ways that electrical designs can be deployed to ensure proper service can be done on the UPS. The most basic and typical for a single UPS installation is with a stand-alone bypass cabinet, typically called a "wraparound" maintenance bypass switch. Some UPS systems do come with "an internal" maintenance bypass switch, but you need to discuss the capability of this switch with the manufacturer to see how much of the UPS is serviceable with only this switch installed. Smaller UPS designs (<50 kW) may use a wall-mounted rotary switch, with larger systems using molded case switches or breakers and very large UPS using high-power rack out breakers in electrical switchgear. Large multimodule parallel systems will typically include many of the same devices as mentioned in the following but in a highly custom-configured system. At a minimum, the bypass for a single UPS should include a maintenance wraparound switch and a UPS output disconnect. This would be typically known as a two-breaker or switch bypass. Some bypass switches include a third breaker as a UPS input breaker (three breaker), and when deploying a UPS with what is known as "dual feed" (static bypass and rectifier have separate input breakers), the switch becomes a four-breaker bypass. Adding UPS input breakers to the bypass is not a necessity, but it does help ensure a more safe installation, as the source breaker is within sight of the UPS and the UPS technician can more safely control the service environment. Another option for maintenance bypass switches is a load bank breaker, or connection point, so the UPS can be tested under load before bringing it online with the critical loads. This can help eliminate a problem that may not manifest itself until load is applied to the UPS. However, caution must be exercised when designing the electrical infrastructure when using a load bank breaker to make sure it does not overload an upstream breaker, causing it to trip off-line and drop the still-operating loads. Sometimes, the maintenance bypass switch may be included in a system cabinet that could include a transformer and/or power distribution panels. This type of design does typically save space and cost, but you must take into consideration that an issue with one of the packaged components may require you to power down the entire cabinet.

Backup Energy Source The primary stored energy source for most static UPS systems deployed today is still the lead acid battery. These batteries have been around for many decades providing a low-cost energy storage medium that is highly predictable. In addition, lead acid batteries are one of the most highly recycled components in production today, with the recycled lead and plastic used to build new batteries. There have been some advancements in lead acid battery technology since the first UPS systems were deployed, and these include making the battery easier and safer to transport as well as maintain. In the past, all the lead acid batteries were a type of "wet or flooded cell" that required that the battery be shipped dry and filled with electrolyte (sulfuric acid) upon installation and then periodically monitored for electrolyte level when in use. Large UPS installations still typically use this energy storage medium as they come in high-capacity sizes in the thousands of amp hours. Since the lead and sulfuric acid are toxic materials and as the battery is recharged hydrogen is released, the typical large battery installation includes a separate room, with separate backup-powered ventilation fans, acid spill containment, and acid neutralizer. This special infrastructure and maintenance can be quite costly, so about 30 years ago, a new type of battery known as the valve-regulated lead acid (VRLA) was released. This battery was originally released in smaller amp hour ratings, but today, many larger sizes are also available. This battery does not include the typical battery fill caps, as the electrolyte is in a gelled form, and may be impregnated in a fiberglass mat between the battery's positive and negative plates (absorbed glass mat (AGM)). This change in electrolyte and the way that it is contained in the battery earned the battery the reputation of being nonspillable. This allowed the battery to be shipped full of electrolyte. Also the battery design uses a pressure valve to utilize the properties of recombination of the emitted gas back into the electrolyte before it leaves the battery case, so a fully sealed battery with low maintenance was the outcome. These advancements in battery design did come with a couple shortfalls. First, the design life is typically shorter than that of the wet cell designs: 10 years for VRLA, while 20 years for flooded cells (some large VRLA batteries are available in 20-year design). VRLA batteries are also subject to higher cell failure rates if overcharged or operated in higher ambient temperatures, and they typically cannot be recharged as quickly as flooded cells.

Other energy storage types are also briefly mentioned in later sections; however, most are only now starting to show some promise to even be thought about for use. Rising energy costs and internal company environmental directives are slowly cutting away on the lead acid battery's dominance in the market; however, it will be a number of years before these other energy sources become economically feasible for mainstream deployment.

27.2.2 General Classifications of UPS Systems

27.2.2.1 Static UPS Systems The most basic static-type UPS designs include a battery, some type of electronic switching devices to convert DC voltage at the battery to an AC voltage (inverter) that can be used by the connected equipment (information technology equipment (ITE)), and another electronic or electromechanical device (switch) that switches between the utility mains (power company) to the battery backup if the incoming power is lost or out of tolerance (Fig. 27.2). The basic UPS also has to have a way to keep its battery charged and recharge the battery if used to supply power during an outage, so it contains some type of battery charger. More complex static systems contain additional components such as automated bypassing devices (static switch) and devices that convert incoming AC to DC (rectifier or power converter). We will discuss these different subsystems in more detail throughout this chapter.

27.2.2.2 Rotary UPS Systems Since rotary systems use rotating mass for energy storage, they have more mechanical moving pieces than a static UPS. The spinning flywheel of a rotary system can typically only provide AC power for 5–15 s at full load, so a backup source such as a diesel generator may be coupled to the generator to provide longer runtimes (Fig. 27.3). Rotary UPS are typically very large power designs, which are not well suited for smaller data center power requirements; however, they are well suited for large manufacturing or process control applications. Data centers wanting to move away from lead acid batteries to an

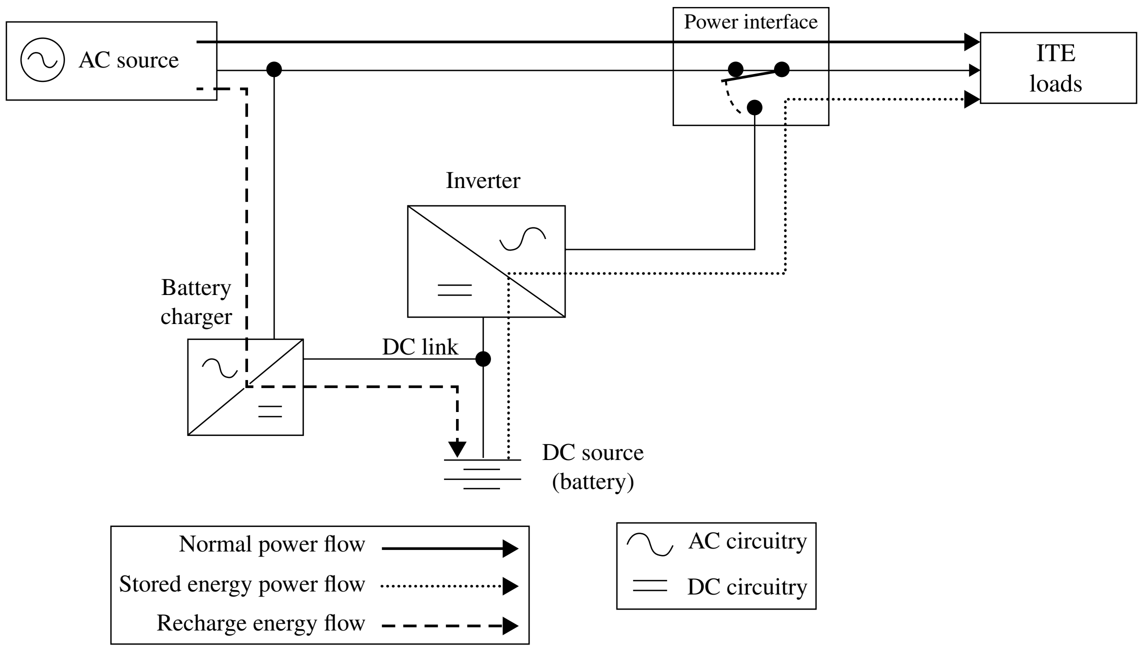


FIGURE 27.2 Simplified power flow diagram for typical small single-phase UPS. Courtesy of Eaton.

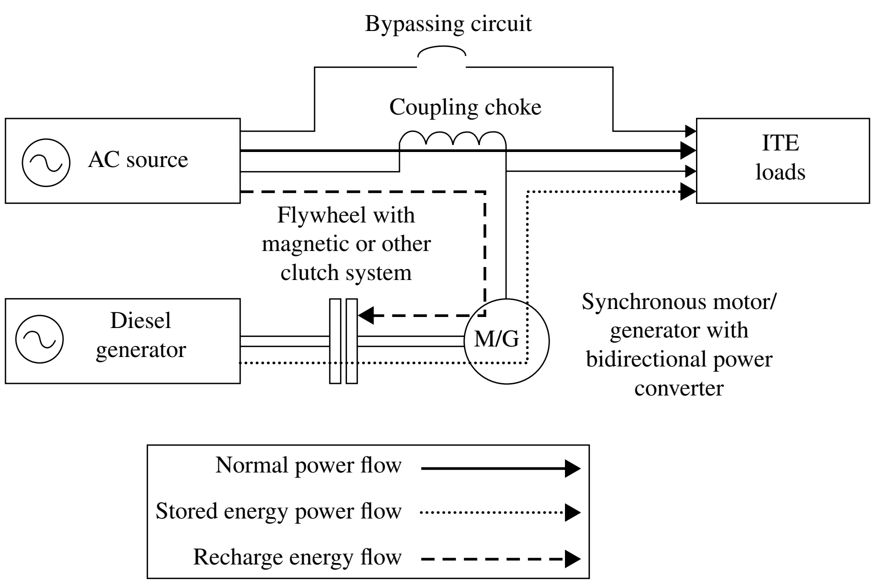


FIGURE 27.3 Rotary UPS system combined with diesel generator. Courtesy of Eaton.

alternate stored energy medium, such as the flywheel, are however deploying variations of the rotary design. These deployments use a rotating flywheel for DC storage, supplying the UPS DC energy similar to what the battery does. When AC mains are lost, the flywheel generates the needed DC energy for the UPS, which is then converted to AC for the protected equipment. Many of the static UPS manufacturers have adapted their UPS designs to allow their systems to use a flywheel for energy; however, the short standard runtimes (15–30 s) are still a consideration when taking this path. In addition, the UPS rectifier, or bidirectional converter in the rotary systems, usually supplies the electrical power to again spin up the flywheel to the correct operating speed (up to 40,000 RPM on some designs). So if you are looking to deploy an energy storage system using a rotating flywheel, you should first check with the manufacturers of the separate components to make sure they are compatible.

27.2.3 UPS Topologies

As mentioned earlier, the static UPS comes in a number of different topologies; however, they fit into two different classifications per the IEEE, either single conversion or double conversion (Fig. 27.4).

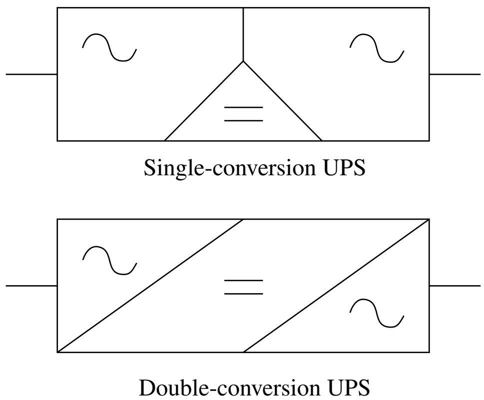


FIGURE 27.4 IEEE accepted drawings for different static UPS designs. Courtesy of Eaton.

27.2.3.1 Single Conversion In a single-conversion design, the incoming AC power is used to directly power the connected critical equipment; however, there may be some voltage regulation done inside the UPS to either "buck" (lower) or "boost" (raise) the voltage to a level that is acceptable to the loads. Some single-conversion UPS designs supply power to critical loads through a series inductor or a linear or ferroresonant transformer, therefore giving "line to load" isolation and transient protection. Single-conversion UPS typically require a separate battery charging circuit to make sure the battery stays charged properly to support the critical load when called upon. The two most popular single-conversion static UPS designs are the standby UPS and line-interactive (LI) UPS (Fig. 27.2).

Standby UPS Standby UPS are most often found in small power applications such as desktop computer or home office or home theater protection. For these applications, some transient protection and short battery backup are the main drivers, as well as low cost. Standby systems operate by passing utility power from input direct to the output, through only a switching device, such as a static switch or small relay. When AC input voltage is within the tolerance level of the UPS output voltage specification (typically −15 to +10%), the UPS stays in this mode, as a very efficient pass-through device. If there is a power outage or large voltage swing, the UPS inverter turns on and the internal switch switches to the backup source (battery) for power. Standby designs all wait until AC input voltage is lost, and then they switch to battery operation, so there is always a "transfer" or switching time where output voltage goes to 0 V. The amount of time is typically based on the system design and the fault condition; however, most manufacturers publish times ranging from 6 to 12 ms, which is usually fine for PC operation, but some slower transfer times could affect operation of larger servers and network equipment. Many low-cost standby UPS have an output voltage on battery that may look almost square in nature, rather than sinusoidal like the utility AC provides. This is due to the lack of output filtering circuit to reshape the output voltage to a sinusoidal shape. While this is okay for many simple PC power supplies, it can create issues with higher-end server or network equipment power supplies, which react differently to the abnormal voltage waveforms. In addition, you will need a true RMS voltmeter if you are checking the output voltage reading of one of these "modified waveform" type of systems, as they will appear to have a large voltage drop when they switch to battery if using an averaging type meter.

LI UPS The LI UPS is found in many applications, from home office to network closets, to the factory floor backing up IT, and even to some processing equipment. LI systems come in a number of different basic power flow architectures; however, they all follow the same principle—that they will regulate the incoming AC voltage to a certain output voltage specification, allowing a wider input voltage range than the standby system design. In doing this, they are slightly less efficient than standby systems, but they offer the benefit of using the backup battery less and therefore extending operation during possible brownout or overvoltage conditions. The different ways that LI systems regulate voltage range from continuously operating the inverter in parallel with the utility to having a simple tap-switching transformer that switches to a buck or boost mode, based on the incoming voltage. Systems that operate their inverter to regulate the voltage will be less efficient than the tap-switching models; however, they typically provide an output voltage that is more tightly regulated. Many LI designs also have a break in output voltage while the inverter is switched on, ranging from 4 to 10 ms; however, systems that keep the inverter operating continuously should have no or a very short break in output power in most power failure conditions. Most LI and standby UPS designs build in some type of transient (surge) protection circuits by using metal oxide varistors (MOVs) or similar devices to clamp or shunt high-voltage and frequency surges or spikes, thereby protecting the connected loads.

Most ferroresonant UPS designs fall under the LI category, utilizing a ferroresonant transformer as the regulating device in normal conditions, with a separate winding on the same transformer, connected to the inverter, for operating on the backup energy (battery). A ferroresonant transformer differs from standard linear transformers as it has the capability to regulate voltage by using the properties of ferroresonance in combination with an output winding circuit known as the "tank" circuit, which also stores energy and provides enough ride-through during an outage to provide a no-break AC waveform. The input winding of a ferroresonant transformer is operated in saturation, with the tank circuit creating a high-voltage resonant winding that creates the very constant output voltage. Ferroresonant UPS do suffer from typically large sizes and lower than typical efficiency, particularly at light loads; however, they are a very rugged design due to the use of very few active devices. These types of UPS are frequently used in poor power and environmental locations inherent in industrial, shipboard, and military applications.

Some rotary UPS designs may also use single-conversion design traits to help raise efficiency by not loading the rotating mass (generator) with the load continuously, but bypassing this to a regulation circuit using inline inductors. If AC power is lost, the generator takes over in supplying power to the loads, until they can be shut down or another source like a motor generator comes online to supply power to the system.

27.2.3.2 Double Conversion A double-conversion UPS (Fig. 27.5) differs as it has the capability to take the input AC voltage and rectify it to create a DC voltage, and then that DC voltage is used to create a new AC voltage waveform; therefore, it converts the energy twice. This newly regenerated output waveform is entirely and constantly controlled by the UPS inverter. However, it must be noted that even a double-conversion UPS will typically track the utility frequency in order to keep itself in synchronization with the source frequency. This is important as if the UPS does need to make an emergency transfer to bypass, it can do so with a minimal break in power, ensuring the loads don't see the interruption. Many double-conversion UPS are programmed to operate up to 3 Hz (hertz) above or below the standard utility frequency to ensure this "lock" to the bypass source.

The double-conversion UPS control logic regulates the voltage, the frequency, and the wave shape of the UPS output at all times (except if on bypass due to overload or failure). These systems are sometimes referenced as rectifier–inverter systems [1]. Double conversion is one of the oldest topologies, having been available for more than 40 years, typically used in highly critical or poor AC power quality environments. Since the system isolates the incoming utility AC from the newly generated AC from the systems inverter, they were always considered the ultimate in protection. However, that level of protection came with a price: higher cost and lower operational efficiency.

Transformer-Based Double-Conversion Designs Some of the older UPS designs still on the market today use large transformers and inductors in their design to operate properly. These systems usually contain more components than the single-conversion systems, as most systems include a rectifier (AC to DC), battery charger (either the rectifier or a DC-to-DC converter), and inverter (DC to AC), and many systems include an emergency bypass mechanism, typically referred to as a static switch. The static switch is used to directly supply the loads with utility AC in case of a severe overload on the UPS output or a failure of the UPS internal systems. Older technology three-phase transformer-based systems typically use six SCRs in the UPS rectifier section to convert the incoming AC to DC to supply DC energy for the inverter and the battery. These systems are sometimes referenced as having a "six-pulse" rectifier. This name comes from the fact that for short periods of time during every AC cycle, there is a period where two different SCRs are on at the same time creating a short circuit of two of the three phases, therefore putting what is called a "notch" in the waveform. If you were to look at the input waveforms with an oscilloscope, you would see six distinctive notches on the input of the UPS. This phenomenon also creates an issue for the incoming utility known as input current distortion or

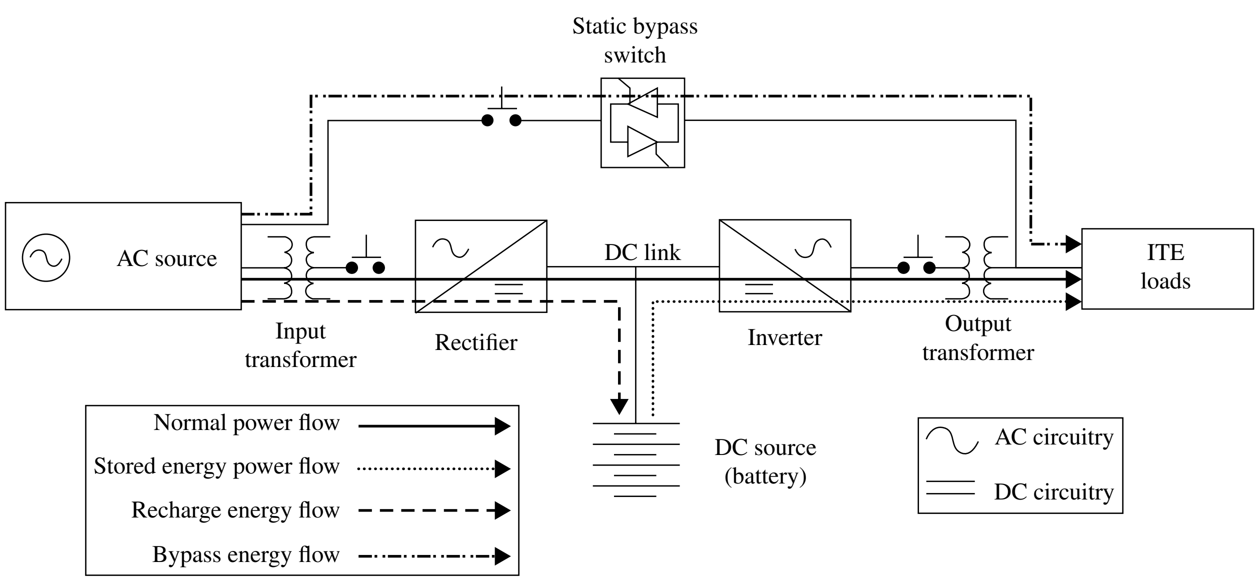


FIGURE 27.5 Typical transformer-based double-conversion UPS system. Courtesy of Eaton.

current (I), total harmonic distortion (THD or iTHD). Most UPS manufacturers recommend installation of an input harmonic filter in the UPS, which is installed on the utility side of the UPS rectifier. The filter consists of capacitors and inductors, which can reduce the current distortion from a typical six-pulse rectifier of 33% to less than 10%. Another means to reduce the input distortion is by adding a large input transformer with dual output windings, known as a 12-pulse rectifier, to the system design. The dual windings of the transformer are phase displaced by 30° to allow use of 2 sets of SCRs, or 12, rather than 6. Doing so eliminates the fifth and seventh harmonics, which are the largest contributors to distortion in the six-pulse design. So THD levels of 12–13% are normal without additional input filters, which when added can reduce the THD below 5%; however, this means more space, higher cost, and lower operating efficiency. By adding input filters, there are also some concerns of operation with generators, as the large input filter capacitors can create a negative effect on the generators voltage regulator, known as a leading input power factor. So most manufacturers offer some type of disconnect circuit to remove the entire or most of the UPS input filter while operating on generator backup, therefore avoiding costly downtime due to negative interaction between systems.

The inverter of the transformer-based designs has gone through a number of different generations, with earliest designs being what is known as “six” step (Fig. 27.6), with higher-power rating designs using a “12”-step architecture. Most of the earlier designs used SCRs, where six of these switches were used between the positive and negative DC points, switching together in a way that made each waveform look like positive and negative steps per each cycle of power. These steps were then formed into a sine wave by using large inductive and capacitive filters on the output of the inverter.

An SCR is a device that cannot be commanded to switch off, but turns off by a process called “commutation,” which happens when current flow through the device goes to zero. Later-designed 12-step inverters used devices known as IGBTs, which could be switched off, but they still only switched on and off at a 60 Hz rate.

There are several different reasons that these UPS system designs needed the transformer, the main one being the inability to “boost” the incoming voltage high enough to regulate the output at the correct voltage (nominal ±1%). In addition, the input also had to charge the battery at the correct voltage level as the battery was directly on the rectified DC “link.” So the input and output transformers would usually be designed to either buck or boost the voltage to meet the needs of the battery charging and the output voltage to the downstream loads or PDUs.

As newer generations of transformer-based UPS systems were released, inverter IGBT’s switching changed to start using PWM switching technology (Fig. 27.1). This was possible due to advances in IGBT designs, as higher-power devices were capable of operating at higher frequencies. PWM switching changed the requirements in the design of the inverter output filter, as filters became smaller, and it also allowed the UPS to change the inverter output faster in case of load changes on the UPS output.

While transformer-based systems are slowly being phased out in many smaller kVA designs, they are still available in some large system designs, where new designs such as transformer-free (transformerless) UPS systems are slower to be adopted and availability of high-speed, high-power IGBTs is still fairly new.

Transformer-Free Systems Starting in the 1990s, manufacturers started making higher-powered UPS systems using a transformer-free design (Fig. 27.7). Some of the benefits of these designs were better dynamic response, smaller physical size and weight, slightly better cost, and, in some manufacturers’ designs, higher efficiency. Transformer-free designs were only possible due to the availability of PWM inverters and transistorized or “active” rectifiers, which replaced the SCRs with IGBT transistors. These designs feature low iTHD (<4%), achieved without the need for the

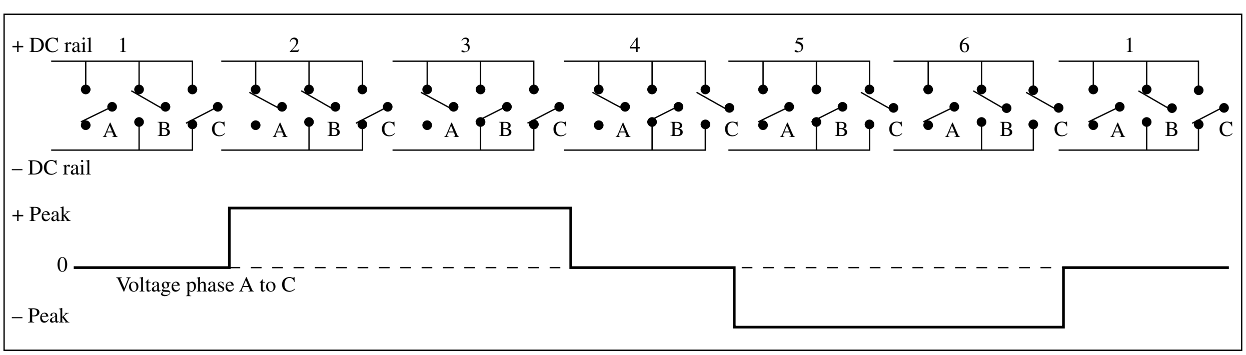


FIGURE 27.6 The six-step inverter switching sequence and voltage waveform for a three-phase UPS. The voltage waveform is for phases A to C of the inverter output. Courtesy of Eaton.

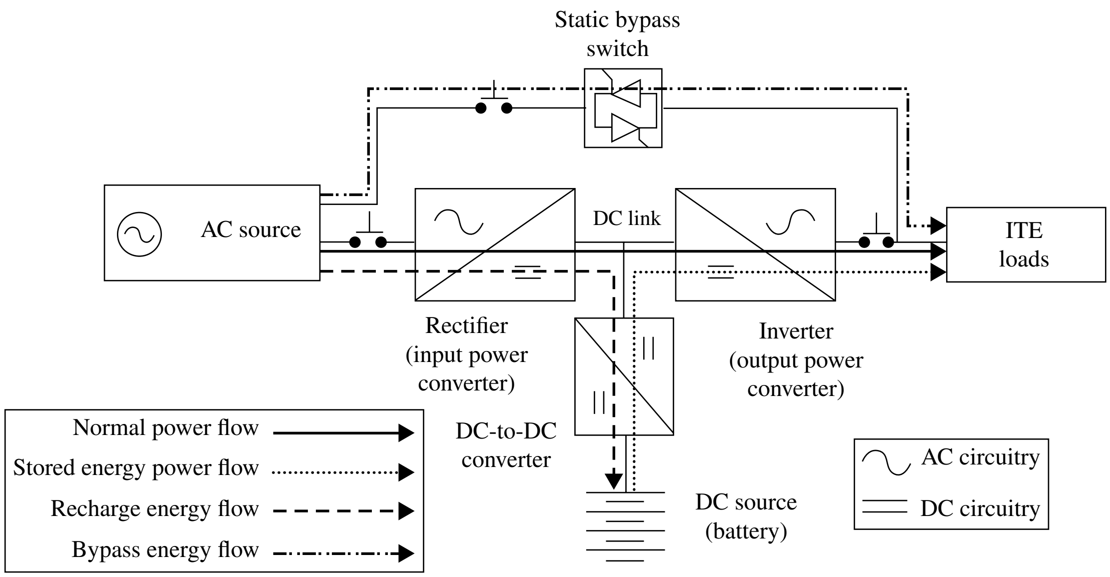


FIGURE 27.7 Basic transformerless double-conversion UPS power flow drawing. Courtesy of Eaton.

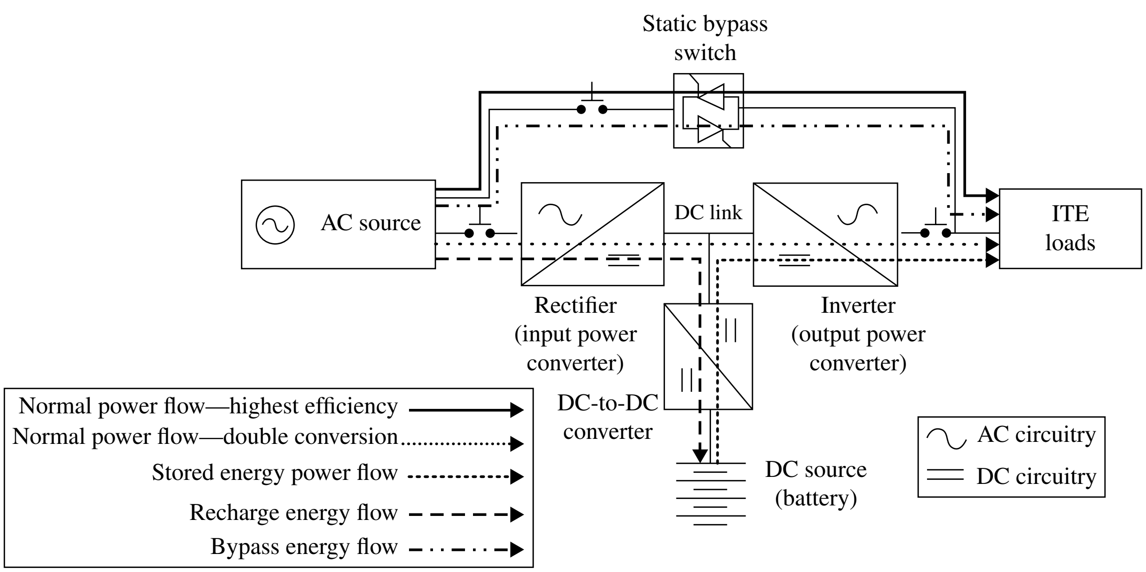


FIGURE 27.8 Multimode UPS uses multiple power paths for highest efficiency and protection. Courtesy of Eaton.

input transformer, and the low-harmonic input filter, saving significantly on cost and complexity. Additionally, the faster switching rates and higher current ratings available with modern power IGBTs allow the UPS inverter and rectifier to respond instantly to transients and faults. This means that the "buffer impedance" provided by input and output transformers in older UPS designs is no longer required. Again, the user benefits from lower weight, smaller footprint, and better operating efficiency.

Multimode UPS The latest generation of UPS, the Multimode UPS (Fig. 27.8), has come about due to the concerns with the high costs of AC mains power and the need to raise the operating efficiency of the system while still providing highly reliable backup power. The multimode UPS uses operating modes found in the LI and standby UPS, as well as double-conversion UPS. The reason the multimode UPS does this is to greatly improve operational efficiency. In the normal mode of operation, when the utility is within the output voltage tolerance of the loads, the static switch remains closed, feeding the loads from the utility. When a voltage aberration outside of the acceptable voltage limit occurs, the UPS immediately opens the static switch and turns on the power devices to create a new AC waveform from the UPS DC source, either the rectifier if utility voltage is still present or the UPS battery.

The most advanced of these systems returns more than 99% efficiency while giving transfer times of less than 2 ms when utility power is lost.

High-efficiency modes on double-conversion systems are not something new. In the past, some double-conversion systems included a high-efficiency mode; however, they were typically subject to several performance issues so the feature was not used. The issues with high-efficiency operation included inconsistent transfer times to battery if utility was lost, no filtering of high-frequency transients, and requirement to operate the inverter to keep the output transformer energized so the transfer to battery could be made in time to support the load. Another issue occurred when the data center design called for large downstream static switches (Fig. 27.21 for system designs) on the output of the UPS, as they saw long breaks in power, which would transfer the loads to the secondary power source every time utility power was lost to the upstream UPS.

Let's look at each of these issues and see how newer designs are handling this. Remember that the static bypass switch in the UPS, which is based on SCRs, requires that the current goes to zero before if it can be switched off. If this device is on and the utility fails upstream of the UPS, the system would wait until the static switch shut off before turning on the inverter. If you didn't wait, all the loads upstream of the UPS (on the same utility feed) may appear as loads on the UPS inverter, causing a severe overload of the inverter until the static switch can be shut off. Therefore, the UPS output voltage momentarily dips so low that the connected loads shut down. Some manufacturers have found ways to force the static switch off in microseconds, rather than the milliseconds that it took in the past. Leading UPS manufacturers are using predictive algorithms to predict what the incoming waveform should look like, and if it deviates from normal, it immediately forces off the static switch and changes the UPS to its double conversion or battery modes of operation to provide the highest protection. In addition, these leading UPS designs can determine if the faulted condition is on the input or output of the UPS, which is extremely important when determining if the static switch should remain on or be forced off.

In the second case, the filtering of the AC is now possible due to the transformerless designs of the higher-powered UPS systems. Surge suppression is accomplished by keeping the rectifier and inverter high-frequency filters attached to the utility source during high-efficiency operation. The capacitors in the filters do the job of greatly reducing any very fast rise time transients, such as lightning, from thousands of volts to just a few volts. Some UPS designs may not, or cannot, use the filters for this purpose, so you need to check the manufacturer's specification for transient suppression. If little or none is provided, install other surge protection devices in front of the UPS somewhere in your electrical system design.

If the UPS design uses transformers and the transformer stays energized in the high-efficiency operating mode, the system will be less efficient than a transformerless UPS design. The latest generation of transformerless UPS designs now ranges in sizes above 1000 kVA, so the benefits of the highest operating efficiency are now available for data centers of any size.

The last matter with downstream static switches is still sometimes an issue but easily overcome with some timing reprogramming of the static switch. However, not all high-efficiency UPS are equal and this is typically only accomplished if the UPS internal transfer times are below 2 ms. The reason is that most sites would like a maximum break in AC power of 4 ms from switching between two different AC sources. If one of the sources, the UPS, takes more than 4 ms to make a transfer itself, the static switch will have already moved the load to another source. Many static switch designs are programmed from the factory to switch if they see as little as a 1 ms break; however, they can be reprogrammed to allow a 2 ms or longer break before switching sources. If you are using downstream static switches, make sure the UPS can make consistent transfers between high efficiency and other modes in 2 ms or less. This will typically ensure a trouble-free operation between the different systems.

The highest- efficiency multimode UPS designs were not possible without the advancements in the industry of transformer-free large UPS system designs, high-frequency sensing and control systems, and advanced predictive control algorithms used in latest-generation designs.

27.3 CONSIDERATIONS IN SELECTING UPS

As with any significant purchase, the designer and the user must evaluate the benefits provided by a product against the costs. UPS systems are similar to other critical data center infrastructure items, like power distribution and HVAC systems, in that a balance must be struck between required performance and reasonable cost. The following are key components of a typical evaluation of UPS characteristics.

27.3.1 UPS Response Time

The ability of a UPS to correct power anomalies by responding quickly to regulate its output is the subject of much variation and debate. The lesser-cost topologies like standby systems described earlier may have a "switching time" from utility fed to battery fed of as little as 2 ms to as much as 10 ms. Larger systems may have a response time of 4 ms to as much as 16 ms (1 power cycle at 60 Hz), and this response time must be evaluated against the tolerance of the IT load devices and possibly against the reaction time of downstream static switches in an A/B bus architecture. It is important to note that LI and double-conversion systems are designed with an inverter that operates constantly and can transition to and from battery operation with no interruption in output. These UPS systems also typically utilize a "make before break" overlapping transition when transferring the critical load to/from inverter to utility bypass, again allowing no break in output voltage continuity. How fast is fast enough? In general, as prescribed by the ITIC/CBEMA and IEC guidelines (Fig. 27.9), IT equipment will not operate reliably if their input power is interrupted for more than 20 ms, and some devices can fail with only a 10 ms outage. Thus, it is a goal of every UPS design to provide response times that are as brief as technically possible and, ideally, make transitions with no loss of output or zero response time. For a large fraction of the mega datacenter market, 0–3 ms response time is preferred.

27.3.2 Efficiency

UPS efficiency is one of the most aggressively advertised and competitively debated features of any UPS. This is true, in part, because the user will readily appreciate that higher efficiency saves money. These savings include both a reduction in the cost to power the UPS and a reduction in the cost to cool the UPS environment. In larger UPS systems, these savings can be quite significant, and a UPS that is a few percent more efficient can often justify a higher initial cost versus a cheaper alternative, or even pay for itself when evaluated against an existing legacy UPS system. Additionally, the more efficient product will provide tangible benefits to the user as they make their case for a more environmentally friendly data center, which is less of a drain on local community resources.

For modern UPS products, a typical efficiency at full load is 92–94%. The best double-conversion products may provide up to 96.5%, with multimode UPS systems reaching an amazing 99+% efficiency. Keep in mind that this full-load efficiency may drop dramatically at light loads and many if not most UPS are loaded at less than 50%, especially in highly redundant facilities, where two separate power systems are provided in complete redundancy. The multimode UPS excel at maintaining high efficiency at loads as low as 15–20% (Fig. 27.10), with many of the systems only losing 2–3% at most. Keep in mind that an older UPS that has been in service for the past 10–15 years may be 5–15% less efficient than a new product. Return on investment (ROI) calculations are easily done comparing legacy performance against multimode performance, with some evaluations showing that the entire cost of a new UPS system may be recouped in 2–3 years, making the choice to upgrade very attractive.

27.3.3 Environmental and Safety

While it is tempting to select the UPS based on technical performance alone, users have a responsibility to evaluate its environmental performance and its impact on the safety of their employees and service personnel.

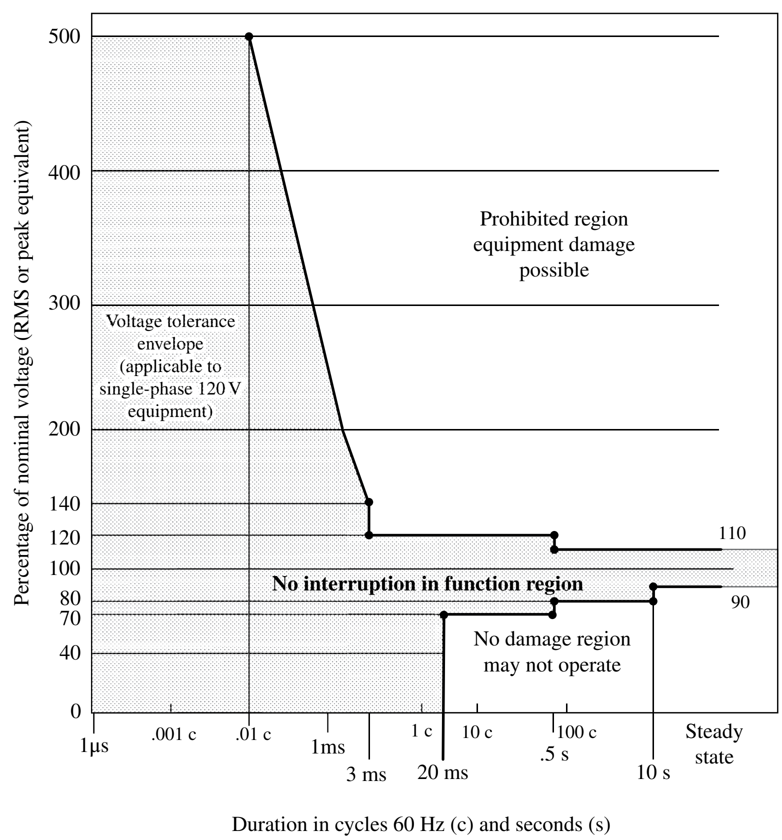


FIGURE 27.9 The ITIC PSU performance envelope was developed before globalized power supplies were readily available; however, it is the only IT equipment PSU specification agreed upon by manufacturers. Courtesy of Eaton.

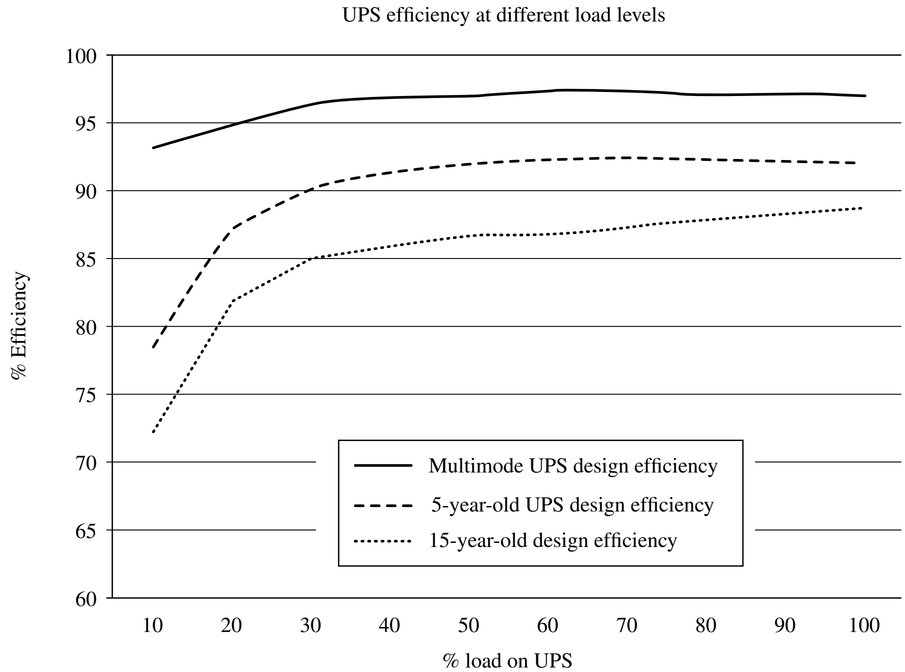


FIGURE 27.10 Different UPS topologies have varying efficiencies under differing loads. Courtesy of Eaton.

27.3.3.1 Sustainability Users that emphasize "sustainability" as a key component of their enterprise will want to evaluate the UPS based on its use of sustainable materials and sustainable manufacturing practices. A "cradle-to-grave" life cycle analysis may be requested as a part of a Leadership in Energy and Environment Design (LEED) compliance process, for example. One key component is the environmental cost of production, which includes the cost to procure and ship raw materials, and the power cost required to construct and test these large electrical systems. There are guidelines and even legal requirements enforcing the mandate to avoid certain hazardous materials in the UPS, and its internal battery, and still more laws regarding disposal of the UPS at the end of its service life.

The battery associated with the UPS is often scrutinized as a hazardous material and subject to special rules for servicing and disposal. In addition, the sheer weights involved will require careful handling during installation, service, and maintenance activities. VRLA, or "sealed" batteries, still contain lead and sulfuric acid, but because they are "sealed," they may be considered less of a hazard than the larger, heavier, flooded electrolyte batteries. Flooded batteries usually require a specialized room with provisions for seismic bracing, containment of spilled electrolyte, and ventilation of hydrogen gas. Smaller UPS, up to 500 kVA in size, are more likely to utilize VRLA batteries due to size and lower cost, with larger UPS, greater than 500 kVA, more likely to use flooded electrolyte battery systems due to their longer service life. Note that there is considerable overlap here, and it is entirely possible that any size UPS could use either battery technology. Either way, operational safety procedures as well as disposal requirements are often legally mandated and strictly enforced.

27.3.3.2 Serviceability While smaller UPS systems may feature safe user serviceability via "field replaceable units" or FRUs, the bulk of larger systems are not user serviceable beyond the external air filters covering the air intakes. Even so, the user should be vigilant and careful whenever one is in close proximity to the UPS or the power distribution system, keeping in mind that a significant amount of stored energy is contained within the UPS, even when utility power is absent. Due to concerns about arc flash, the aptly named "dead front" covers internal to the UPS should never be removed by the user. When considering the UPS from a safety perspective, the selection of the UPS should include verification of UL listing, or local international safety certifications, along with the use of a certified, experienced installation contractor that will observe proper wiring and grounding requirements as per local codes and the National Electrical Code (NEC).

27.3.4 Cost

As with most purchases, better performance and quality internal components associate directly with higher cost. But that's not the whole story. There is also the need to consider the total cost of ownership, or TCO, for these systems. TCO includes the upfront cost to procure, manufacture, factory test, and ship the system, along with the following:

- Cost to perform the electrical installation and testing
- Floor space cost, per year of operation
- Power cost to operate over the UPS service life (efficiency)
- Cooling power cost (also affected by UPS efficiency)
- Cost to maintain and repair the UPS over the service life
- Cost to maintain the system battery and planned cost to replace VRLA batteries every 5–6 years during the life of the UPS
- Disposal cost for the UPS and batteries at the end of its service life

Then there is the significant cost if the user chooses a redundant UPS system or a 2N or "dual-bus" architecture. These systems add the "extra" redundant UPS and, importantly, the extra battery system for that UPS. This may double many of the costs listed earlier, affecting footprint, testing, maintenance, power cost, and cooling capacity cost. Some modern UPS systems feature internal or inherent redundancy, due to their modular and scalable construction. In many cases, this can allow the user to have the reliability benefits of an $N+X$ redundant system, without the traditional penalty in capital cost expenses and increased footprint.

27.4 RELIABILITY AND REDUNDANCY

Historically, mean time between failures (MTBF) has been a key metric that UPS manufacturers use to measure and express reliability. In truth, however, MTBF is generally a poor means of predicting UPS availability.

To understand why, consider a UPS with an MTBF of 200,000 h. A layperson might expect such a device to experience one failure in 200,000 h—or 23 years—of operation. In reality, however, UPS manufacturers can't and don't test their products for 23 years. Instead, they calculate an initial MTBF based on the projected life span of the UPS's components. Then, after they've shipped a statistically meaningful number of units, they replace those preliminary estimates with new ones based on actual performance in the field. Those revised numbers can be misleading though. For example, if 2500 UPS perform flawlessly over a 5-year study period, the result will be an impressively high MTBF rating. But if those systems contain a component with a 6-year life span, 90% of them could fail in the year following the study period.

In addition, there is no universal standard for measuring MTBF. For years, most government agencies have required manufacturers to provide calculations based on the latest revision of the MIL-HDBK-217F handbook, while many commercial customers have adopted the Telcordia (Bellcore) SR-332 process. However, like many differing standards bodies, these two standards will give markedly different results. More recently, the technology industry has concluded that these measurements, while helpful, should not be the only way manufacturers grade a product's reliability. As a result, manufacturers today increasingly focus on Design for Reliability (DFR) as well. Unlike past standards, which concentrate on individual electronic components and their relationship to the circuits used in the product's design, DFR methodologies pay greater attention to a product's intended and expected use under varying conditions.

Still, at the end of the day, there remains no one standard for measuring how a UPS performs its mission, which is keeping connected loads powered. As a result, it's nearly impossible to compare one UPS manufacturer's MTBF figures to another's. Availability offers a somewhat more realistic measure of critical power backup systems. Given the vital role that UPS play in the data center, the ability to replace aging or failed parts rapidly is crucial. Availability combines MTBF with a second metric called mean time to repair (MTTR) that measures the time required to acknowledge a problem, respond to it, and complete a repair:

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}$$

Availability is typically expressed as a number of "nines" representing the percentage of time over a year's worth of use that a given system is operational. For example, a UPS with an MTBF of 500,000 h and an MTTR of 4 h would have an availability of 0.999992, or 99.9992% (500,000/500,004). That translates to an expected downtime of 4.2 min/year.

Still, though it's a better gauge of reliability than MTBF numbers alone, availability is flawed in important respects. In particular, it fails to account for time spent on routine service functions. If a system has to be taken down once per year for inspection, recalibration, or general maintenance, its actual operational availability will be lower than the aforementioned formula suggests. Therefore, the MTTR number should also include the time off-line to service the equipment throughout the year to get a better gauge on total system availability.

27.4.1 Strategies for Increasing the Availability of UPS Power Paths

One of the most common ways to increase availability is to increase the number of power paths for AC power to be delivered to the connected load. This can be internal to the UPS or could consist of paralleling multiple UPS systems. When internal to the UPS, there may be multiple independent subsystems that can support themselves, or it may be separate, replaceable, or repairable modules that can be quickly serviced. Most data center grade UPS systems include an automated bypass, the static switch, which gives the UPS two internal paths (Fig. 27.11) for getting power to the load. As the diagram shows, a path through a maintenance bypass is also typically specified for most data center applications.

Some latest-generation UPS have been designed with multiple modular power components that can be fully isolated from the live power bus and repaired or tested (Fig. 27.12). These types of designs will typically have the main power components, the rectifier and inverter, replicated so they can give some internal redundancy in case of a failure of either of these components. However, these systems still have some common components that may require the entire load be bypassed or shut down in case of a failure of one of these components.

When evaluating the UPS reliability, a simple diagram is typically used to help understand the relationship of the components to the entire systems reliability. For this, we will consider that any one of the three redundant components in Figure 27.13 can be shut down and the two remaining devices can support the output requirement. In the upper portion of the figure, the subsystems in series (A, C, D) from the input to the output are considered a failure point that will jeopardize total system reliability. Subsystem B is redundant and one module could be replaced while the other systems support the load. The lower diagram in Figure 27.13 shows a typical parallel redundant configuration using three separate modules, each containing all subsystems needed to operate on its own. So a failure in any one of the single systems (1, 2, or 3) will not affect the entire system reliability.

One thing that must be considered when paralleling multiple systems or subsystems in redundancy is how many is too many. As you can imagine, the more components that you add to the system, the more components there are to fail, so you may end up in a situation where you have diminishing returns. This can be avoided by selecting the proper building block upfront and limiting the number of parallel systems. It is recommended that you try to start with power blocks where you can handle the full-load capacity with four to six parallel systems, therefore reducing the number of components being deployed. However, there could be a situation that your application is so large that you are using the largest block available, therefore increasing the number of modules needed and increasing the possibility of replacing failed parts in the future.

Statistics have shown that the most common failure point in any static UPS system is the battery. Even with most new UPS designs completing multiple battery tests on a monthly or quicker schedule, a single internal or battery connection problem could cause a failure. A simple way to increase the battery reliability is to add a parallel battery string.

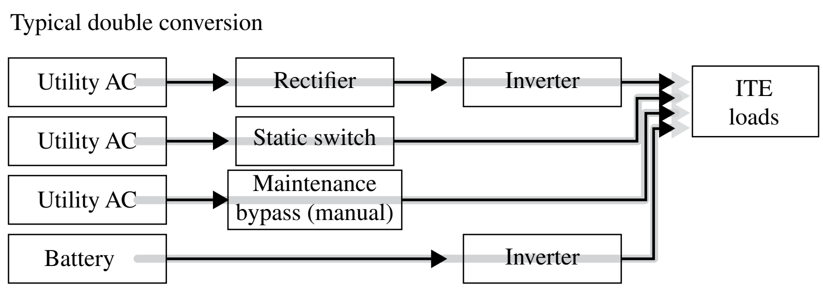


FIGURE 27.11 Multiple power delivery paths to get power to the connected loads. Courtesy of Eaton.

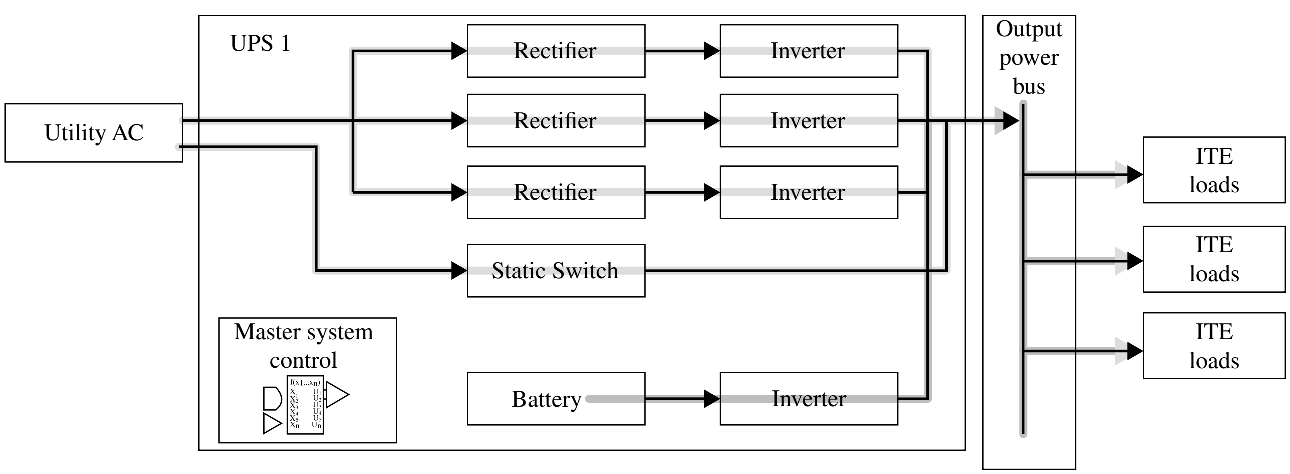


FIGURE 27.12 Single UPS system with some internal redundancy provided for rectifier and inverter. Courtesy of Eaton.

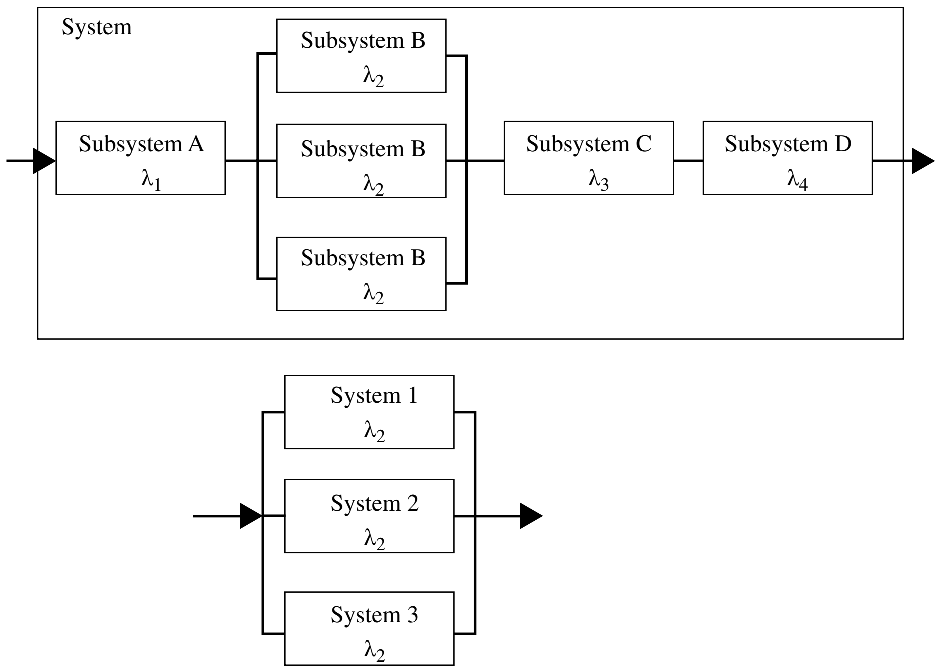


FIGURE 27.13 Upper drawing shows subsystem redundancy, while lower drawing shows parallel redundancy. Courtesy of Eaton.

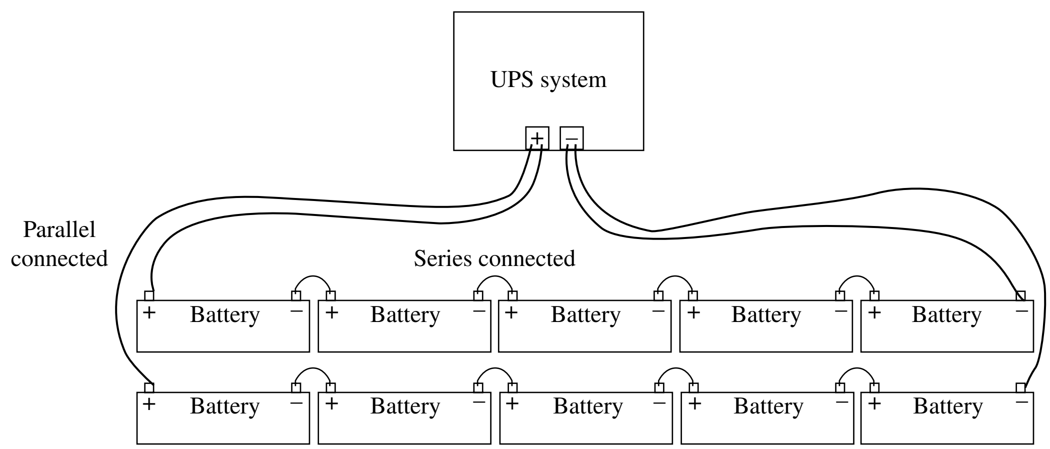


FIGURE 27.14 A way to increase UPS backup energy (DC) reliability is to parallel multiple battery strings. Courtesy of Eaton.

Equipping a UPS with a single string of series-connected batteries can dramatically increase risk of load loss. Say, for example, that a large UPS has 40 batteries connected in a series (+ of the first battery to − of the next). If a problem occurs in any of those batteries, the entire string will probably fail, causing the UPS itself to fail. Adding another 40 batteries and then tying the most positive and most negative points together give you two parallel strings of batteries (Fig. 27.14). If either string fails, the UPS can typically run for a limited time on the other string until either a backup generator comes online or load equipment is shut down gracefully.

Increasingly, organizations are finding that the risk of running off straight utility power—even briefly—is too great to ignore. So they deploy redundant UPS modules to ensure conditioned power even if one UPS module fails. In *paralleling*, two or more UPS are electrically and possibly mechanically connected to form a unified system with one output—either for extra capacity or redundancy. In an $N+1$ redundant configuration, you would have at least one more UPS module than needed to support the load. As a conjoined system, each UPS stands ready to take over the load from another UPS whenever necessary, without disrupting protected loads. Let's take a closer look at parallel UPS architectures—how they work, what challenges must be overcome in establishing parallel configurations, how modern paralleling technology enhances availability, and what difference it makes in your power protection scheme.

Redundant UPS configurations were once relatively rare. Organizations balked at the expense of buying two UPS to do the work of one. Only the most substantial organizations—or those with the most critical power requirements—made the investment. That has changed. Data center managers and facilities managers have come to the conclusion that running off raw utility power, even briefly, represents unacceptable risk. The cost of downtime is now so high that even small data centers can justify the cost of redundant UPS. In fact, redundancy is a requirement of data centers that attempt to meet reliability levels defined by some industry experts, such as the Uptime Institute.

The Uptime Institute requires a minimum of $N+1$ redundancy in the power systems at as low as a Tier II redundancy level, with greater electrical system redundancies at higher levels. As a result, parallel UPS configurations are becoming commonplace. At least 50–60% of large UPS systems (300 kVA and up) are configured as parallel systems. Ten years ago, it was uncommon to parallel smaller systems (in the neighborhood of 10 kVA), but now up to 40% of these smaller systems are paralleled—particularly in Europe and Asia.

27.4.2 How Do Parallel UPS Configurations Work?

On the surface of it, the concept of paralleling UPS for redundancy is simple enough. Multiple UPS modules are linked to perform in unison (like one big UPS), sharing the critical load among them via a common output, with each module ready to take over for any other module if necessary. In an $N+1$ configuration (a typical redundancy arrangement), there would be sufficient spare capacity to support the load if any one module became unavailable. For example, you could protect a 500 kVA load by deploying two 500 kVA UPS systems (Fig. 27.15) or an 800 kVA load by deploying three 400 kVA UPS modules. During normal operation, the three 400 kVA modules would each carry one-third of the total 800 kVA load. If one module went off-line, the remaining two modules would have sufficient capacity to support the load. Figure 27.15 shows a typical parallel configuration with two UPS modules. In normal operation, AC power flows from the utility source to each UPS. Each UPS has two inputs, or what is known as dual feed, where one input goes into the rectifier and one into the internal bypass (static switch). The UPS converts incoming AC power to DC and then back to AC and then sends this clean power to a tie cabinet, where outputs from both UPS are merged into a single output to protected loads.

Should a failure of any kind occur with either UPS module (Fig. 27.16), the critical load is still UPS protected. Internal diagnostics immediately isolate the faulty UPS module from the critical bus while the other UPS assumes the full load, remaining in normal operation, not needing to activate the internal static switch to go into a bypass mode. When the UPS installed in a parallel configuration retain their own internal static switches, the installation is said to have a "distributed bypass."

During a utility failure, each UPS module is supported by its battery system and can continue operating for minutes or hours, depending on how much battery runtime has been provisioned.

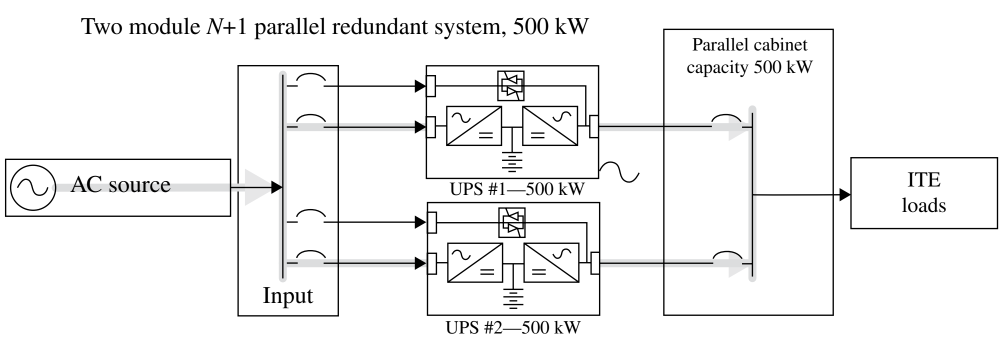


FIGURE 27.15 In normal parallel operation, both UPS modules contribute equally to shared output. Courtesy of Eaton.

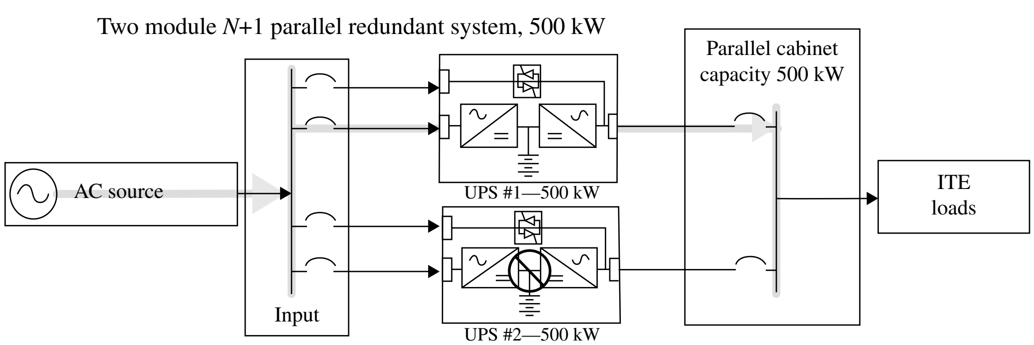


FIGURE 27.16 If either UPS module becomes unavailable, the remaining module assumes the load. Courtesy of Eaton.

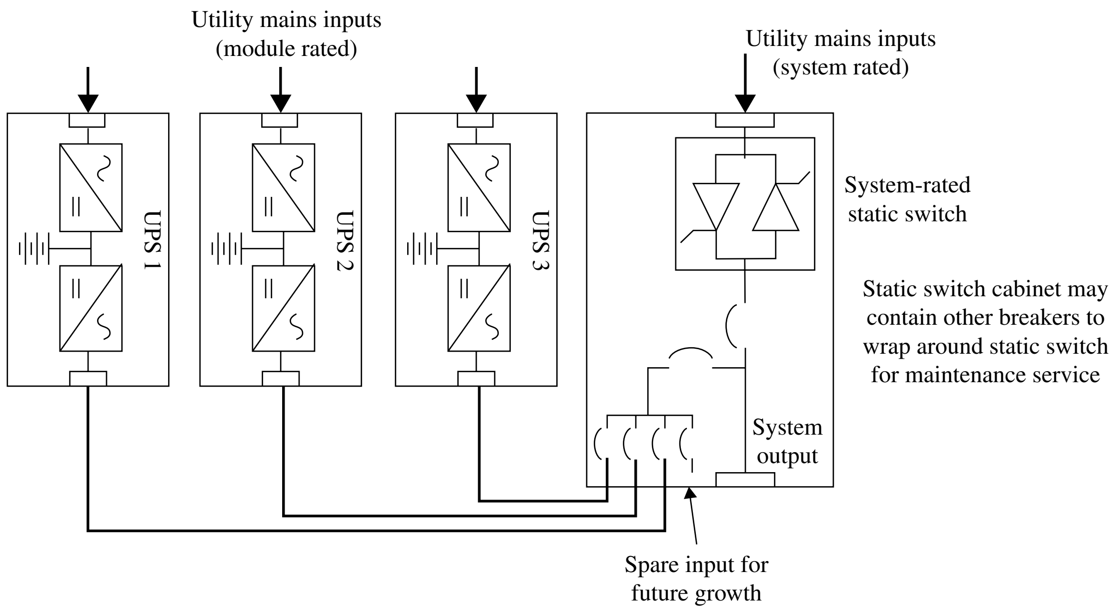


FIGURE 27.17 In a centralized bypass system, power flows to critical loads, even if all three UPS modules were off-line. Courtesy of Eaton.

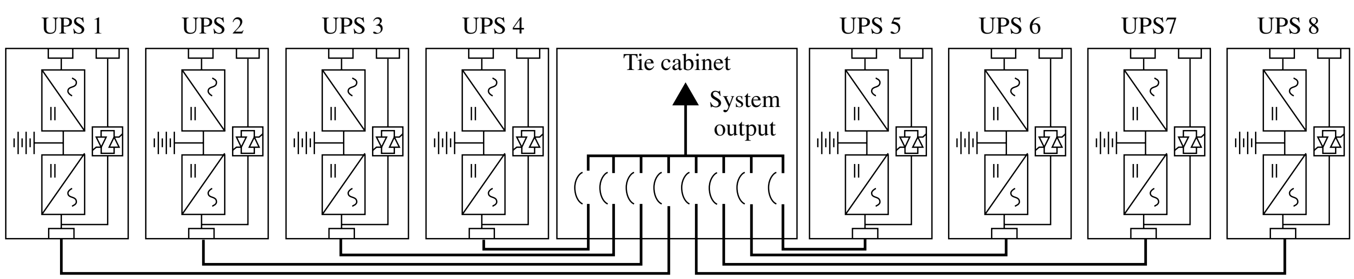


FIGURE 27.18 This distributed bypass configuration has eight UPS modules paralleled into a single system. Courtesy of Eaton.

You can (and should) provision separate battery backup for each UPS, for even greater backup protection and a higher level of redundancy; however, sometimes, that may not be economically feasible. A parallel UPS configuration is not limited to two UPS modules, as it frequently includes up to four modules and some installations may contain eight modules or more. With the newest rack-mounted UPS designed for high-density server environments, no freestanding tie cabinet is required. Paralleling is accomplished using a plug-and-play bus structure that mounts easily in the back of a standard IT rack, where the UPS modules also are installed.

The configuration shown in Figure 27.17 has a single bypass cabinet rather than the standard tie cabinet, which is known as a centralized bypass configuration (the static switch is "centralized" in an external cabinet). The tie cabinet with a separate bypass, its own full system-rated static switch, provides an alternate route for power during a failure—an automatic and instant wraparound bypass. Such an event would be rare; however, it may be activated during service or repair instances. The wraparound bypass would be activated only if the connected UPS were unable to support the load in normal operation. Perhaps, a short circuit caused an extraordinary overload that exceeded the capacity of all three modules together. The system would identify a failure on the critical bus and transfer to bypass mode with virtually no interruption.

Another alternative is known as a distributed bypass parallel system (Fig. 27.18). In this system, each UPS retains its own internal static switch, and they all operate in unison when a transition to or from the bypass source is required. In this type of system, when many UPS are linked in parallel, the load they collectively support will exceed the capacity of the internal static switch and bypass circuit in any one UPS. So there is a need to ensure that all UPS modules equally share the load, even when powered by the internal static switches in each UPS. The more UPS modules that are tied together, the more important the electrical cabling that connects the modules from the utility power to the output tie cabinet becomes. The individual static switches in each UPS have no way of controlling load sharing, so it is only the electrical paths impedance that controls how much load each UPS carries when in bypass. If the total cabling route is very long to one system and very short to another, the system with the shortest route will have the lowest impedance, so it will assume more load than any of the other systems when in bypass. If it is too unbalanced, that UPS may become overloaded and either take itself off-line due to overheating or a breaker may trip off-line, creating a cascading effect that could cause the rest of the systems to fail.

27.4.3 Four Key Challenges in Parallel UPS Systems

As soon as you connect multiple AC power sources into a unified, parallel system, there are four key challenges to address:

1. Controlling how the separate UPS should cooperate as a unified system
2. Synchronizing the output of each UPS so it can flow into a shared output
3. Balancing the load equally among all UPS in the parallel configuration
4. If trouble occurs, identifying and temporarily decommissioning the UPS with the problem

These issues can be complex, and they must be managed in a way that does not compromise the high reliability for which UPS are paralleled in the first place.

27.4.4 Parallel Systems for Added Capacity

Most organizations plan to grow, but when and how much? How much power will you consume next year or in 5 years? You don't want to overbuild the power system today for future demands that may or may not materialize. Even if you could justify the cost, the power infrastructure would operate far below capacity and be very inefficient as a result. And you certainly don't want to rip out and replace today's UPS just because next year's moves, adds, and changes suddenly double the need for power.

Paralleling provides an excellent solution for matching growth while extending the value of existing UPS. The architecture to parallel for capacity looks very similar to paralleling for redundancy. Hardware components are the same; there are just small differences in operation. A system paralleled for capacity allows you to add load until it reaches capacity and then notifies you to add another module. In contrast, a redundant parallel system constantly ensures that there are enough modules to take over the total load if one drops off ($N+1$). For example, if the parallel system has five 100 kVA modules, the system would issue an alarm if the load exceeded 400 kVA—the load that four of those five modules could support.

27.4.5 Customization Options for Large Parallel Systems

In practice, large customers need one-of-a-kind, specialized configurations that match their unique needs for availability and manageability. Many options are available for parallel UPS, such as the following:

- Wraparound maintenance bypass, to allow loads to keep running (off straight utility power) even if the parallel system is unavailable, such as during a natural disaster
- Redundant breakers in the tie cabinet, to permit maintenance of the primary breakers without turning the system off
- Separate load bank breakers in the switchgear, to enable use of a load bank to test the UPS system under load while it is isolated from protected loads
- Communication cards and a monitoring system for remote monitoring

27.4.6 Other Options for Establishing Redundant UPS Protection

Redundancy doesn't always require paralleling. There are other options for deploying multiple UPS modules—separate rather than paralleled—to provide an added layer of assurance in the power protection architecture. For example, separate UPS can be set up to provide serial redundancy, where even if the primary UPS is off-line, its bypass path is protected by another UPS system upstream of the failed UPS module. Or a data center could be divided into separate zones served by separate UPS, thereby minimizing the impact of any single UPS failure. Or separate UPS could serve either side of dual-corded loads—or source power from different utility substations. Furthermore, any of these options can be set up for duplicate redundancy. However, each option presents some compromises, compared with peer-to-peer configurations described earlier.

27.4.7 Total System Installation

There is a wealth of other considerations as the facility UPS system and its ancillary equipment are being defined and as the entire critical infrastructure is designed. These include items like alternative energy sources, maintenance bypass capabilities, and even high-end ultraredundant architectures like A/B or dual-bus architecture.

27.5 ALTERNATE ENERGY SOURCES: AC AND DC

27.5.1 Alternate AC Energy Sources

There are choices for the AC input to the UPS system. While the utility power grid is by far the most common source for any UPS, there are a few other possibilities. When designing for mobile installations, or other harsh environments, where a stable utility source may not be readily available, one will need to evaluate other possibilities, which include the following:

- Wind or solar power (or wave power for marine installations)
- Diesel or turbine generators
- Fuel cells

27.5.1.1 Alternate DC Energy Storage Sources

Given the lead acid battery's many flaws, it's no surprise that data center managers have long been clamoring for alternatives. At present, five such technologies show particular promise. Though none is in widespread use today and a limited number of existing UPS models are equipped to support them, all are likely to gain increased traction over the years ahead.

Flywheels A flywheel is a mechanical device typically built around a large metal disk. During normal operation, electrical power, via a motor, spins the disk rapidly. When a power outage occurs, the disk continues to spin on its own, generating DC power that a UPS can use as an emergency energy source. As the UPS consumes that power, the disk gradually loses momentum, producing less and less energy until eventually it stops moving altogether. Due to the cost and requirement to parallel multiple flywheels for longer runtimes, most are deployed with a backup time of typically 15 s to about a minute maximum, with the average time being 30 s. This is typically plenty of time to ensure that the backup generator(s) comes online to support the critical power requirements.

Ultracapacitors Also known as supercapacitors, ultracapacitors are specialized, extremely high-density batteries. They typically contain nontoxic, carbon-based materials such as activated carbon and graphene. Available runtimes are very short, typically less than 30 s, so operation in instances where there is a backup generator is typically a must.

Fuel Cells Unlike batteries, fuel cells generate power rather than store it. A fuel cell is basically an electrochemical device that converts fuel (typically hydrogen) into energy. However, unlike an internal combustion engine, which also converts fuel into energy, a hydrogen-powered fuel cell's only exhaust product is water. As a result, everyone from automakers to electrical utilities to UPS manufacturers are presently either introducing fuel cells to their product lines or investigating their use.

Lithium Ion Batteries Most cell phones and laptops use lithium ion batteries, which have grown steadily smaller, lighter, and denser over the last decade. Though they're rarely used today in industrial settings or data centers, lithium ion batteries are capable of performing most of the same functions as lead acid batteries.

Nickel Sodium Batteries These batteries offer high capacity, rugged construction, good power density, extreme temperature operation, and excellent environmental characteristics, with no toxic material or by-products of disposal. Nickel sodium battery systems are already being used for outdoor applications like utility power and telecommunications facilities, where long runtimes are required. Smaller systems are already being offered for UPS applications as well.

27.5.2 Dual-Bus or 2*N* Architecture for Dual-Corded Data Center Equipment

In this arrangement, the UPS modules feed separate distribution panels that support separate power supplies (PSU) within every piece of IT equipment. UPS A supports one power path and one of the power supplies in the IT equipment, and UPS B supports the other power supply (Figs. 27.19 and 27.20).

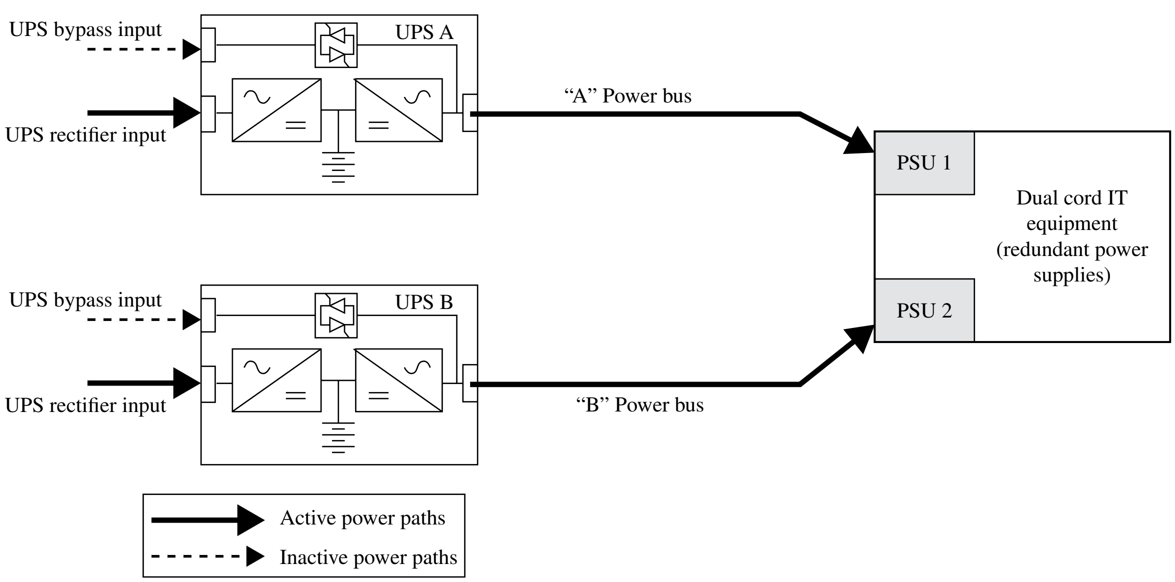


FIGURE 27.19 Typical dual-bus power system with separate UPS systems feeding the different power distribution buses. This type of deployment puts the power failure point at the individual IT device, ensuring the highest levels of availability. Courtesy of Eaton.

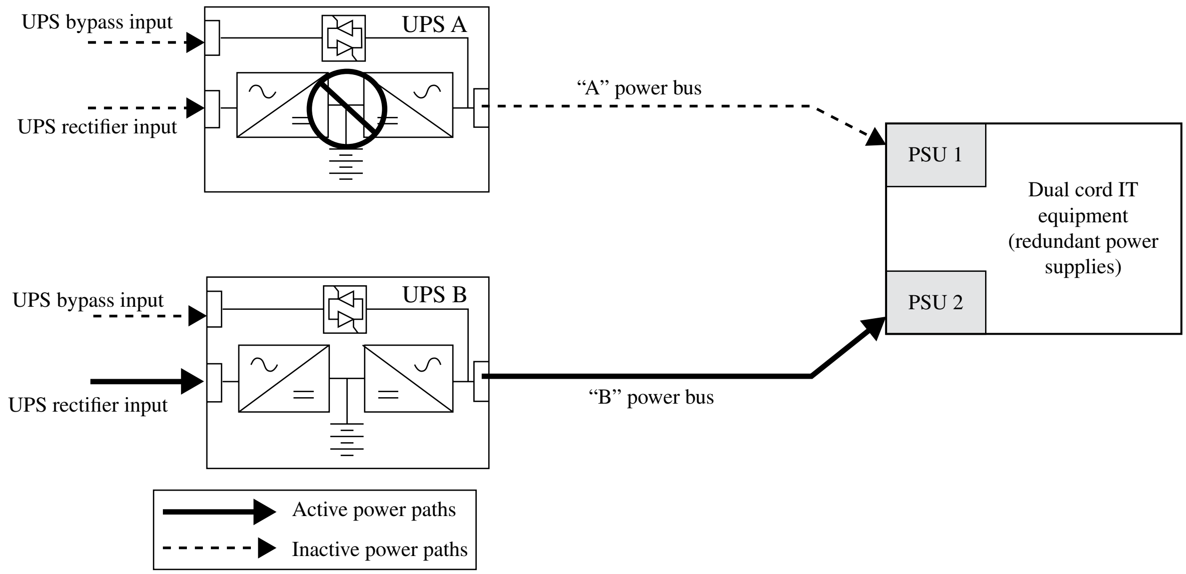


FIGURE 27.20 Failure of one of the power paths or UPS forces the other UPS to assume the entire load; therefore, designs like this should be sized properly and continually monitored to reduce the chance of a UPS or power distribution system overload. Courtesy of Eaton.

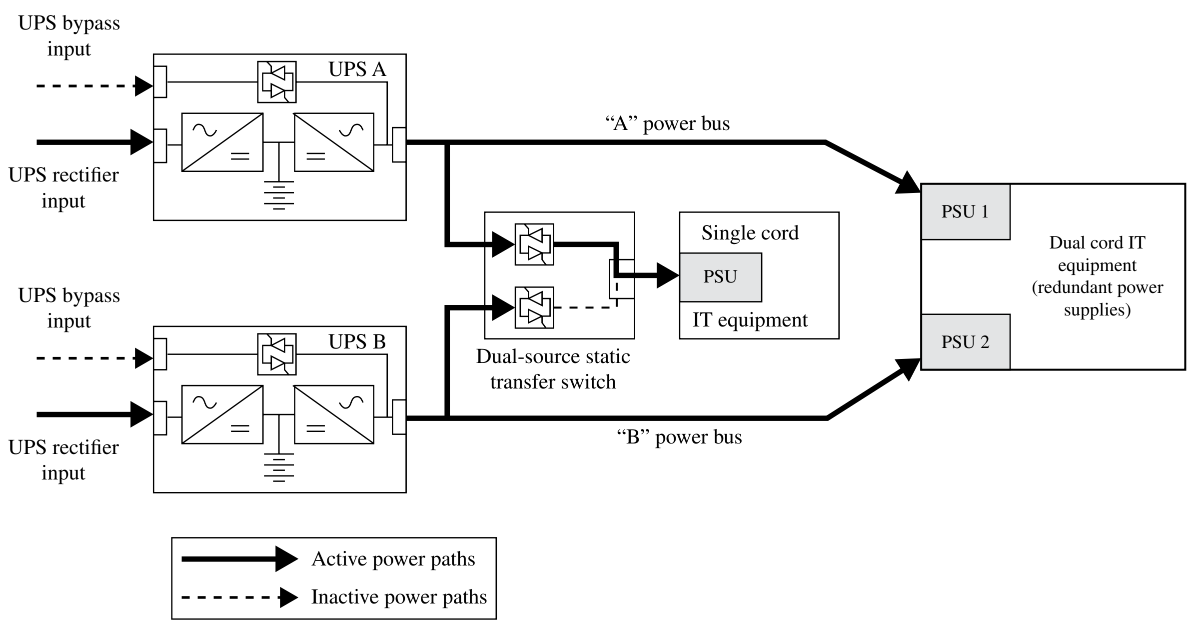


FIGURE 27.21 The dual-source static switch above is connecting the "A bus" to the single power cord loads. Courtesy of Eaton.

This configuration offers a lot of flexibility, because the UPS modules do not have to be equivalent. They can be of different sizes, carry vastly different loads, and even come from different manufacturers. But it only makes sense for a data center that *exclusively uses dual-corded IT equipment*. Most of today's data centers still have legacy equipment that uses a single power supply, such as modems and other communications gear.

How do you provide redundancy for those single-corded loads? Some sort of static switch arrangement would be required to switch single-corded loads from one UPS to the other, in the event of a failure of the primary UPS. In Figure 27.21, a static switch serves the single-corded equipment in the data center. Alternately, you could use a small, relay-based dual-source transfer switch mounted in the rack to feed any single-corded equipment in that rack. Whatever type of dual-source switch is deployed, the switch would transfer the load from a failed power source (UPS in this case) to the available UPS in milliseconds, without disrupting the protected load (Fig. 27.22).

This arrangement adds complexity to the power distribution architecture. The more components in the power

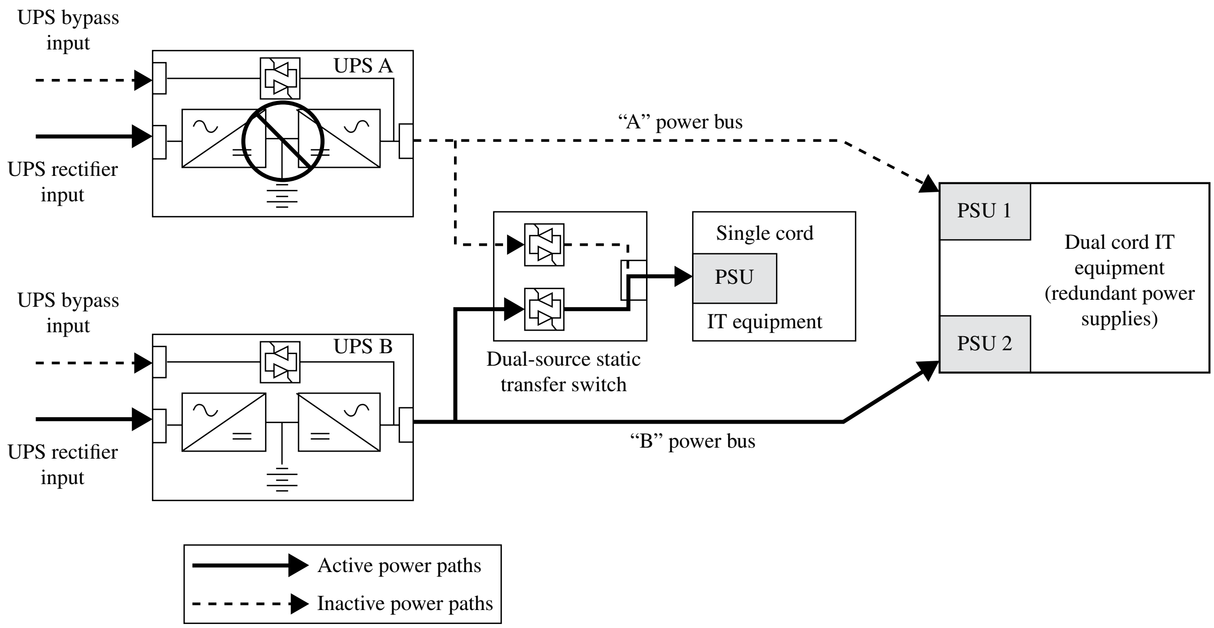


FIGURE 27.22 Failure mode: the static switch must instantly switch the single power cord loads to the "B bus." Courtesy of Eaton.

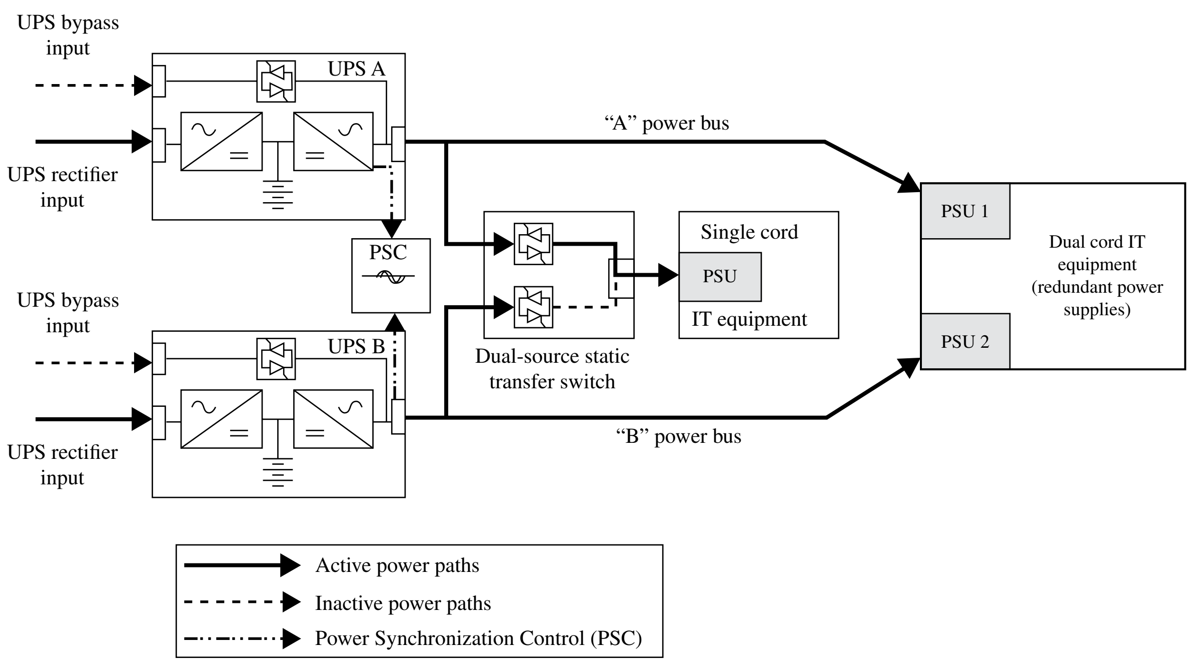


FIGURE 27.23 Power sync control (PSC) is necessary to ensure that the static switch can change sources without disturbing the single power cord loads. Courtesy of Eaton.

delivery chain, the more points to monitor, maintain, and troubleshoot, and the more possible points of failure. However, an even more troubling issue is synchronization. If the UPS are not in sync with each other, the rapid switch of power from one to another via the static switch could introduce a voltage transient that could shut down or possibly damage or destroy those single-corded loads.

So now we have a situation where, even though the UPS are not providing all the benefits of paralleling, their outputs still must be synchronized. This can be accomplished with an external power synchronization control (PSC) unit, which sets up a master–slave arrangement. Now the availability of those single-corded systems rests on the reliability of the static switch and synchronization controller. For that reason, this arrangement is best used as a stopgap measure as single-corded loads are phased out in favor of dual-corded devices (Fig. 27.23).

27.5.3 Separate UPS with Multilevel Redundancy

Higher levels of redundancy can be achieved with a dual-bus system, especially if each bus gets its power from a different utility substation. In the diagram, each side has two UPS modules (a primary and standby UPS, for $N+1$ protection), a system bypass module (SBM) to transmit power from the UPS modules or utility source, and its own backup batteries and diesel generator. Under normal operation, one bus feeds power to distribution panels serving one power supply in the dual-corded IT equipment; the other bus feeds distribution panels serving the other power supplies (Fig. 27.24).

If any UPS drops off-line, the standby UPS on that side goes into action. Even if both UPS on a side became unavailable, the IT equipment would still be powered from the other side. If a substation went out, the power would still remain up, because the other side is served from a different utility source.

For its high availability, this is a widely used arrangement. But "redundant redundancy" is expensive. And there's still the issue of what to do with single-corded loads. You can add a PSC, which resolves the synchronization issue described earlier, and simply accept a small point of vulnerability. In the arrangement shown in Figure 27.25, those redundant UPS systems are linked via a hot tie cabinet. The hot tie cabinet has breakers that can isolate either side from the power chain entirely or link them together in parallel.

In normal operation, the breaker in the middle would be open, isolating the two redundant UPS systems from each other. The UPS system on the left feeds its output to the left-side bus. During a failure condition or routine maintenance of, say, the left side, the breaker in the middle would be closed and the left breaker open. Then the right-side UPS is powering both the A bus and B bus. The loads see no change in the voltage, frequency, or quality of the power they receive.

27.6 UPS PREVENTIVE MAINTENANCE REQUIREMENTS

With proper servicing and a stable environment, a well-made UPS can operate safely and reliably for as long as 20 years. Without proper servicing, even the best UPS is significantly more likely to fail when you can least afford it. Companies in the market for UPS hardware, therefore, should also choose an appropriate UPS service plan from a service provider with the experience, know-how, and resources to provide comprehensive, high-quality support—and do it safely and quickly!

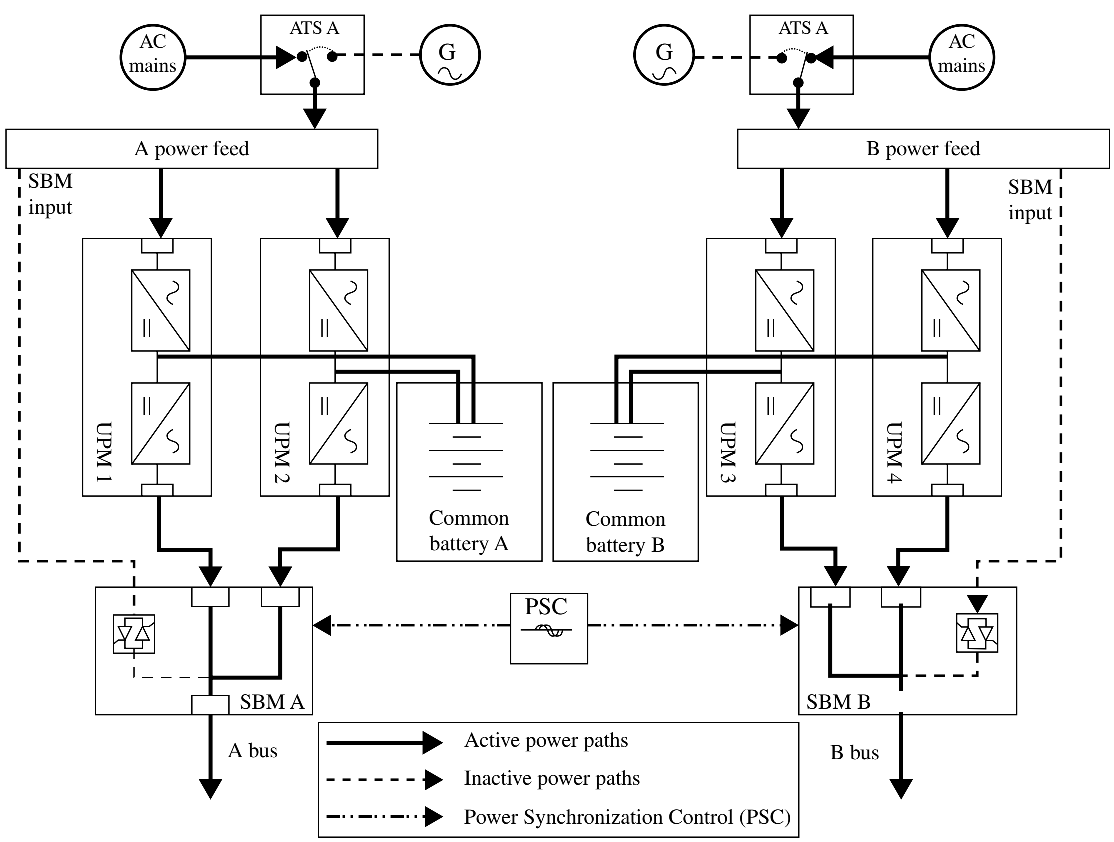


FIGURE 27.24 Multiple $N+1$ UPS systems can be used in a dual-bus configuration. Each bus has its own system bypass module (SBM), and upstream ATS with generator. A common battery is used for each group of uninterruptible power modules (UPM). Courtesy of Eaton.

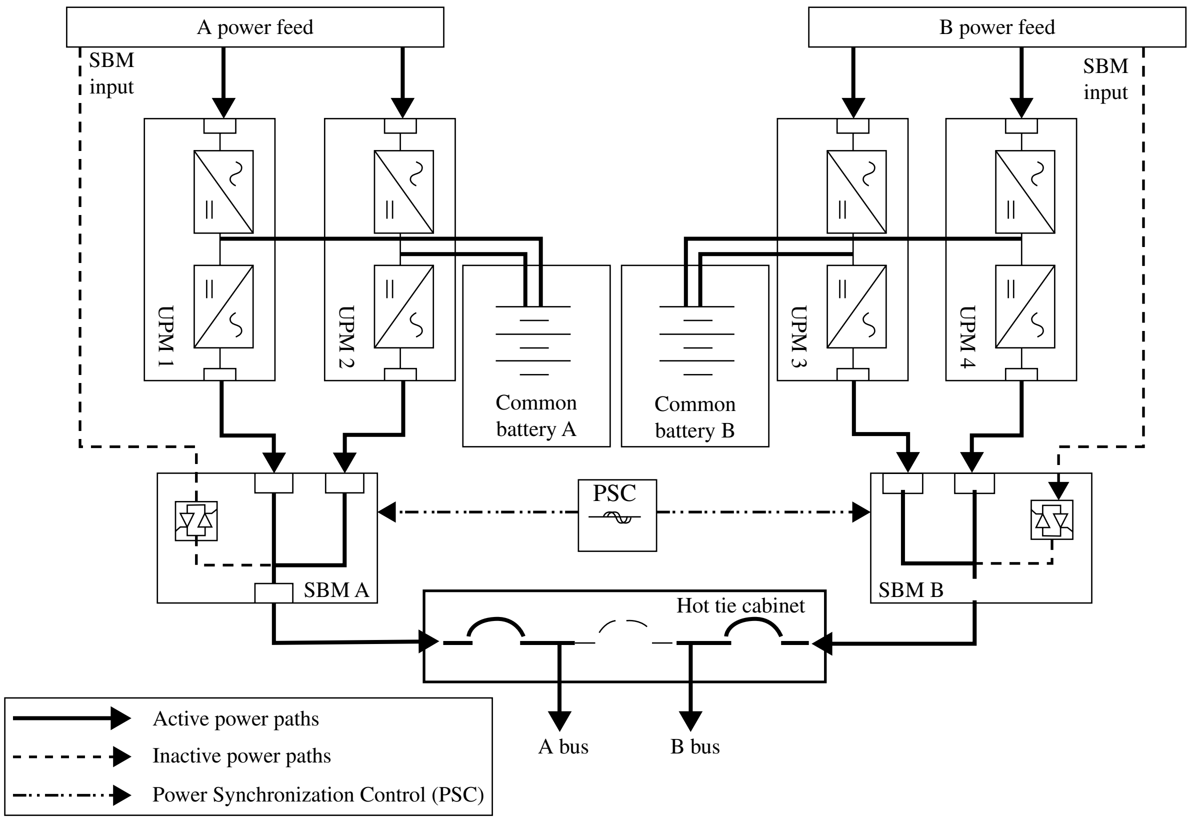


FIGURE 27.25 The hot tie cabinet (HTC) allows concurrent maintenance of the upstream A or B bus while the downstream loads stay powered by both buses. Courtesy of Eaton.

Research indicates that regular preventive maintenance—which affords the opportunity to detect and repair potential problems before they become significant and costly issues—is crucial in order to achieve maximum performance from your equipment. In fact, studies show that routine preventive maintenance appreciably reduces the likelihood that a UPS will succumb to downtime.

Selecting a service provider for your UPS can be a complex decision. Some customers simply purchase a service contract or extended warranty from the UPS manufacturer, while others prefer to contract with an independent service provider. A handful of companies employ internal engineers who are capable of maintaining all or certain parts of the power equipment. Still others choose to engage in UPS service only when something goes wrong. All of these options have advantages and disadvantages, with no one choice being the best solution for every organization.

27.6.1 Common Questions for Choosing a Service Provider and Plan

1. If my UPS fails to provide reliable backup power, what is the cost of downtime to my organization?
2. How critical is continuous power to my application? Is it simply an inconvenience or do I lose customer sales, destroy products, or shut down a network of critical servers?
3. How long can I wait to obtain an emergency repair on my UPS? A week, a day, or an hour?
4. What's my position in the "priority list" during an environmental disaster?
5. How many trained field technicians for my specific UPS model are within 100 miles of my site, and do they carry the correct parts?
6. Do I have budget or cost constraints for UPS service?
7. How much scheduled preventive service do I need and what can I afford?
8. What level of service is recommended by the manufacturer?
9. Have I budgeted for battery, capacitor, or other unplanned part replacement costs?
10. Do I have competent electrical staff resources to do some or all necessary maintenance?
11. What is my risk tolerance for a UPS failure, and what happens to me personally if this UPS fails?

Regardless of the exact course of action you implement, an effective preventive maintenance plan saves time and money by minimizing business interruption and the costs of downtime, as well as enhancing your overall ROI by extending the life span of your critical power equipment.

27.6.2 Option 1: UPS Manufacturer's Internal Service Organization

Engaging in a service contract with the manufacturer of your UPS affords a number of benefits. To begin with, customers receive the extensive knowledge, capabilities, and expertise of factory-trained field technicians who receive ongoing and in-depth training on the manufacturer's specific UPS products. As a result, technicians are armed with the most up-to-date and comprehensive information pertaining to the functionality of the UPS, as well as access to the latest firmware and upgrade kits to maintain the highest level of performance from the UPS. Furthermore, the advanced troubleshooting capabilities of technicians translate to a reduced MTTR. When performing service on a UPS, the day-to-day familiarity and knowledge that comes from being brand specific cannot be underscored enough.

In addition to offering a deep support infrastructure of design engineers, technical support personnel, and other experts to back up its field technicians, UPS manufacturers generally possess the greatest number of field personnel and back-office resources. Furthermore, manufacturers most often have in place risk mitigation programs that are frequently overlooked by customers, such as appropriate safety programs and proper levels of insurance.

Another significant advantage to manufacturer-provided service is that technicians have spare parts readily available either from a stocked van or from a central location, ensuring that UPS problems are quickly resolved, most often on the initial service call. Furthermore, many service plans include discounts on part kits and product upgrades, which can significantly reduce the overall cost of maintenance.

To meet the varied needs of customers, UPS manufacturers offer a wide variety of service plans, including standard warranty, extended warranty, preventive maintenance, numerous service contract levels, and time and material (T&M) billing. Many also feature value-added support such as remote monitoring. Even more, most manufacturers offer service contracts that include options such as 7 × 24 coverage, with response times ranging from 2 to 8 h or next-day response—an especially appealing benefit for customers in mission-critical environments. While the price of service may be slightly higher from a manufacturer compared to an independent service provider, the advantages that only a UPS manufacturer can offer may outweigh any additional costs.

27.6.3 Option 2: Independent Service Provider

An independent service provider is a third-party organization that often offers a range of services for UPS or power quality equipment, such as professional maintenance, consulting, start-up, installation, and emergency service. Although independent service providers are frequently priced lower than a UPS manufacturer, they also generally have fewer resources available and may not be comprehensively trained on your particular UPS model.

While an independent service provider's field technicians have usually received training on either a specific UPS product or brand and may or may not be certified by a UPS manufacturer, it is virtually impossible to fully train a technician on every UPS model from every manufacturer. Furthermore, because UPS products are continually being updated and changed, if a technician has not been recently trained by the manufacturer, he or she may not have the knowledge to adequately service the UPS.

When it comes to having access to repair parts, some technicians may carry the appropriate parts with them or have them available from a central location. However, it is difficult to carry a local supply of adequate parts for all brands. Generally, independent service providers will access a UPS manufacturer's deep support infrastructure of design engineers, technical support, and experts to back up their own field team, as the depth of their own resources can be limited. Insurance and safety records may or may not be maintained at an acceptable level. While independent service providers generally do not deliver a factory warranty unless contracted by a manufacturer, they do offer preventive service, a variety of service contract levels, and T&M billing. Some may offer value-added support such as remote monitoring.

27.6.4 Option 3: Self-Maintenance

If an organization has an internal resource that possesses sufficient electrical and safety skills, it may make economic sense to perform self-maintenance on a UPS. The most important aspect of self-maintenance is to have an efficient plan in place, in which routine scheduled maintenance is performed and common wear items such as batteries and capacitors are proactively addressed.

First responder training enables a skilled person to understand the operation, safety, environmental concerns, and basics of preventive maintenance on a specific UPS. This person must also understand the various alarm conditions and responses required for specific events, as well as the steps to start and stop a UPS correctly in various applications.

A spare parts kit obtained from the UPS manufacturer can supplement those who choose to self-maintain their UPS equipment. However, it is important that an organization also has access to a professional service provider for more critical repairs, upgrades, or routine maintenance that may be required to supplement a self-maintenance resource.

27.6.4.1 Questions to Ask When Considering Self-Maintenance Before opting to perform self-maintenance on a UPS, consider the following questions:

1. Is there an internal resource within your company that possesses basic UPS knowledge and electrical skills? If so, does this individual have time that can be designated to UPS maintenance?
2. Has your organization developed a specific plan for self-maintenance, including a schedule for replacing common wear and tear items?
3. Has a spare parts kit been purchased from the UPS manufacturer?
4. Has an external service resource been identified for more critical repairs?

27.6.5 Option 4: T&M

Paying as you go is a common UPS maintenance approach that can be appropriate in certain situations, primarily for very old UPS models where no service contract is available. However, this tactic does not make good economic sense for complex, multimodule, or redundant UPS configurations.

Available at any time to all customers, T&M is typically charged per hour of labor, often with a minimum number of hours required. Charges are also generally more for after-hours and weekends, compared to normal business hours. Response time for T&M is typically "best effort" with no guarantee of arrival, as customers with existing service agreements are always given priority over T&M customers.

Another downside to T&M is that replacement parts are usually very expensive. For example, the average board for a common three-phase 80 kVA UPS costs more than $5,200, while power modules that integrate several components exceed $10,000 each, with larger models containing several pairs of modules.

The uncertainty of response time during an emergency and financial exposure to unplanned repairs may make T&M less attractive to more mission-critical organizations. On the other hand, T&M may be appropriate for a self-maintainer, in situations where a UPS is not fully utilized or where preventive maintenance is being performed by a manufacturer or independent provider and the insurance portion of a service contract (parts and labor coverage and emergency response) is deemed unnecessary by either self-insuring or other reasons.

27.6.5.1 Questions to Ask When Considering T&M If you are considering the pay-as-you-go approach, it is important to first consider the following questions:

1. Is there a service plan available for your particular UPS?
2. How complex is your organization's UPS?
3. Is your UPS utilized regularly or occasionally?
4. Is your UPS supporting mission-critical applications?
5. In the event of a UPS failure, can your organization afford an uncertain amount of downtime until a technician is able to schedule a service call?
6. Does your company have sufficient funds allocated for T&M service, parts, and repairs?

27.7 UPS MANAGEMENT AND CONTROL

Even with a UPS, your IT system could still go down in case of an extended power failure or if the UPS gets overloaded for too long. Communication software can not only provide real-time notification of UPS status but also let you assign automatic actions to perform in case of a power event. This is extremely useful if your system operates continuously without users being present to manually shut down affected equipment.

For the past 20 years, most UPS systems have come with software that would signal one or more servers that AC power was lost and that the UPS was on battery. If AC did not return and the battery energy was near depletion, the software would close all open applications to prevent any data loss. When AC power was restored, the system would automatically reboot, bringing the system back to its previous state. This solution was initially implemented on small PC servers protected by a single UPS then moved to larger systems with an array of operating systems, many of which were proprietary to the IT equipment manufacturer. Communication was established through an RS232 serial port or via relays to a simplistic control port.

As IT systems grew bigger in size and numbers, serial communication (be it RS232 or through a USB port) was replaced by network-based communication to enable communication between the UPS and multiple servers. In this type of installation, the UPS is assigned its own IP address on the network and could be accessed remotely by all servers being powered by that UPS, so each server could be programmed to shut down or monitor the UPS for power issues.

As networks and UPS communication hardware and software became more complex, other automatic features were developed through power management software, including remote notification via email, pager, or SMS; data accumulation allowing report generation and trending analysis; complex script programming to shut down a database or a program before stopping the server operating system; and much more. Even with all of these advancements, the typical installation involved servers with single operating systems and with a single application running on each server.

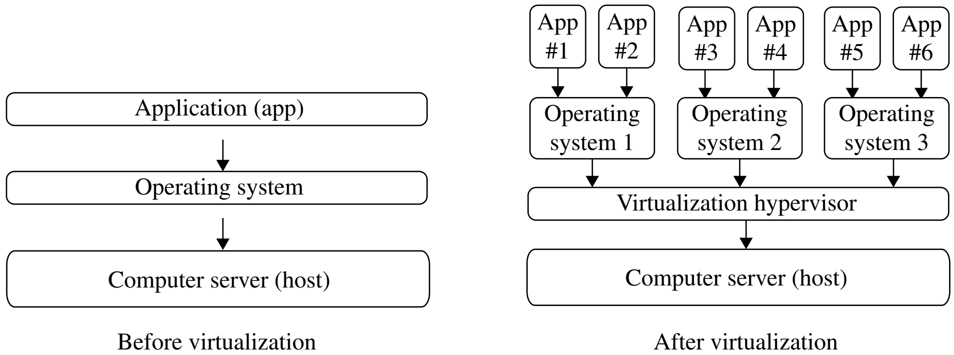


FIGURE 27.26 Advancements in computer hardware and software have led to the need to more efficiently use the compute power of each server, so the technology of virtualization was brought down from the mainframe architecture to the individual server level. Courtesy of Eaton.

Virtualization is now bringing a new set of complexities, as the bond between operating system and physical hardware is no longer the standard. Some suppliers of UPS software must ensure that shutdown software agents are installed on each virtual machine as well as on the host machine. This can be quite tedious if the number of virtual machines is large, which is becoming the standard in many virtualized environments. Leading-edge UPS manufacturers have developed new software platforms that reduce this management complexity by integrating their software into virtualization management platforms like VMware's vCenter® or Citrix XenCenter®. In these environments, one single software installation can control and shut down any cluster of servers. Another benefit is the enablement of automatic live migration of virtual machines in case of a power outage, as you are no longer limited to the option of shutting down the servers and stopping operations. Business continuity is now possible through this integration, which is not only available on vCenter® but also on Microsoft SCVMM or Citrix XenCenter® (Fig. 27.26).

To summarize, logical and complete power management applications can help companies:

- Monitor and administer their UPS from any location with Internet access
- Automatically notify key personnel of alarms or alerts
- Perform orderly, unattended shutdowns of connected equipment or, better, work with virtualization software to move virtual machines so as to maximize availability of key applications and hardware
- Selectively shut down noncritical systems to conserve runtime
- Analyze and graph trends, to predict and prevent problems before they happen
- Integrate with existing network and management systems via open standards and platforms

27.8 CONCLUSION AND TRENDS

Businesses today invest large sums of money in their IT infrastructure, as well as the power required to keep it functioning. They count on this investment to keep them productive and competitive. Leaving that infrastructure defenseless against electrical dips, spikes, and interruptions, therefore, is a bad idea.

A well-built power protection solution, featuring high-quality, highly efficient UPS hardware, can help keep your business applications available, your power costs manageable, and your data safe. By familiarizing themselves with the basics of what a UPS does and how to choose the right one for their needs, data center operators can ensure that mission-critical systems always have the clean, reliable electricity they need to drive long-term success.

As IT solutions progress, there are always industry leaders that are looking to challenge the ways of the past and deploy equipment in new and somewhat unproven configurations, pushing the economic envelope to their favor. With increasing reliability placed onto the IT software redundancy platforms, they are starting to look at ways to reduce the amount of equipment needed on the power redundancy side. While the economic impact of this can be shown on paper to be very attractive in the short term, the total business impact may not be known for years.

REFERENCE

[1] http://www.nema.org/Standards/Pages/Uninterruptible-Power-Systems-Specification-and-Performance-Verification.aspx. Rosslyn: National Electrical Manufacturers Association. NEMA PE 1-2003; 2004.

FURTHER READING

Corrigendum 1—Uninterruptible power systems (UPS)—Part 1: General and safety requirements for UPS. Geneva: International Electrotechnical Commission. IEC 62040-1 Ed. 1.0 b Cor.1:2008.

Crow LH. Achieving high reliability. J Reliabil Anal Center 2000;4:1–3.

Dod Guide for Achieving Reliability, Availability, and Maintainability. Washington, DC: Office of the Secretary of Defense; 2005.

IEEE guide for batteries for uninterruptible power supply systems. New York: Institute of Electrical and Electronics Engineers. IEEE Std 1184-2006 (Revision of IEEE 1184-1994).

LaCommare KH, Eto JH. *Understanding the Cost of Power Interruptions to U.S. Electricity Consumers*. Berkeley: Ernest Orlando Lawrence Berkeley National Laboratory; 2004.

Reliability prediction of electronic equipment; 1995. MIL-HDBK-217F.

28

USING DIRECT CURRENT NETWORK IN DATA CENTERS

Sofia Bergqvist
IBM Corporation, Stockholm, Sweden

28.1 INTRODUCTION

The amount of installed power usage in a data center is often in multimegawatts—large data centers can have a power usage of over 500 MW. The total power consumption of the data centers in the world is comparable to that of air traffic. The high energy costs and carbon dioxide emissions resulting from the operation of a data center call for alternative, more efficient solutions for power supply design. One proposed solution to decrease energy usage is to use a direct current (DC) system for all the servers in the data center and thereby reduce the number of conversions between alternating current (AC) and DC.

The DC solution has many advantages over the traditional AC design with regard to efficiency and reliability. Other aspects such as a more efficient integration to renewable energy sources are enabled with the DC technology.

This chapter will focus on the pros and cons of this technology, what studies have been made and what the current market trends are. Some examples from around the world of data centers operating on DC are presented in order to give real-life practical information from operators and data center owners. Furthermore, an introductory section where a background of how our power system has changed over the decades and what comes next can be found in the beginning of this chapter.

The goal is to take all possible parameters into consideration with regard to technology, economy, safety, and environmental aspects and leave the reader with a complete picture of what this new technology implies and why this change is happening now.

Other aspects of the DC technology in terms of design, integration with renewable energy supplies, reliability, and investments costs were also taken into account.

28.2 EDISON'S REVENGE

In the late nineteenth and early twentieth centuries, there were frequent debates on the issue of AC versus DC. Thomas Edison, the inventor of the light bulb, advocated the use of the DC technology, and Nikola Tesla was the champion of the AC technology. Gradually, during the remainder of the twentieth century, AC became the predominant technology. Most of the old DC grids were gradually converted to AC grids, and AC became the standard in our homes and buildings (subtransmission grid) as well as for long-distance transmission.

However, during the last couple of decades, a large number of areas have been identified in which local DC grids and long-distance DC power lines are proposed to be introduced instead of AC because of several advantages of the DC technology. In this section, the background of why AC came to defeat DC and why and in which areas the DC technology is being reintroduced will be discussed.

The fact that we are now finding new uses for DC grids is Thomas Edison being vindicated. Having lost the battle of the currents a century ago, he is now partly getting his revenge; which is why the new interest in DC is often referred to as Edison's revenge.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

28.2.1 Why AC Outpowered DC

In the early twentieth century, DC was the standard in homes and buildings empowering electric motors and light bulbs. However, difficulties in breaking the current and transforming between different voltage levels with the existing technology of the time led to a switch in favor of AC. This conflict in the early twentieth century of which technology to use is sometimes referred to as the "war of the currents."

28.2.2 DC for Long-Distance Distributions Lines

Large long-distance power distribution lines, for example, those connecting a large hydropower plant with a city, need to be highly efficient [1]. The demand for long-distance power distribution is increasing because of urbanization and the building of new hydro, nuclear, and large coal-fired plants.

AC has traditionally been used for long-distance power distribution. But because of the nature of AC, there are implications with power quality and efficiency in long-distance distribution lines. DC has proven to be more efficient and more economic for transmitting large amounts of power point-to-point over long distances.

AC comes with more control parameters such as frequency and phase, which makes synchronization between different grids more complicated compared to the DC alternative. DC, on the other hand, is more suitable for this purpose because of its "simpler" nature, which enables a more efficient distribution [2].

The use of DC instead of AC for long-distance power distribution also reduces the requirement for transmission of reactive power. Today, the power grid must transmit what is called reactive power. This power is of no use but requires "space" in the electricity network. The network provider sometimes has to install equipment to compensate for reactive power in the electricity [3].

28.2.3 DC in Buildings

AC is the standard for power distribution within buildings even though most of the equipment we use in our everyday life operates on DC, for example, computers, LED lighting, control systems, and robots. In Europe, according to new directions by the European Union, all buildings constructed after 2020 have to be practically energy neutral. This means that they have to produce as much energy as they use. This production will come from on-site renewable energy sources.

This calls for more efficient power distribution grids within the building. Since almost all of today's apparatus utilizes DC and the renewable energy sources generate DC, a network within the building distributing DC would be the most efficient one. Several large companies such as Siemens and ABB are investing in research in the area of low-voltage DC grids, sometimes referred to as "smart grids" [4].

A DC electrical distribution network integrates solar cells simply and efficiently on both large and small scales. Since the output from solar cells is DC, there is no need for DC/AC inverters, and therefore, the system will have higher efficiency. The power provided by solar cell arrays can be scaled up simply through simple parallel connection to a battery and the mains supply. Windmills can also be integrated to a DC network, and by doing this, there is no need for an additional DC/AC conversion step.

Turning to the end user side, almost all loads in the electricity network are already or will increasingly be electronic, which is due to the general technical development in the electronic field and the need for high energy efficiency in use of electricity.

The electricity network is not designed for electronic loads, since such load did not exist at the time when the electricity network was built. Breaking of DC to electronic loads is not associated with electric arcs and fire hazard. This is because all electronic loads contain an energy store in internal capacitors in the appliance, which means that no voltage surge that can cause an electric arc occurs across the break. The interface between the electricity network and the load, for example, in electrical installations in buildings, is not adapted to modern electronic loads and therefore not as efficient as they might be. This leads to unnecessary energy losses and costs, disturbances, electrical environment problems, harmonics, magnetic fields, higher appliance costs, etc.

Furthermore, electrical appliances can be made more efficient and cheaper if they only need to have DC/DC converters instead of AC/DC converters in their power supply units (PSUs) [3].

Data centers are perfect examples of part of a building where DC should be used. All servers, batteries, and LED lighting operate on DC, and that combined with renewable energy sources such as solar cells results in a need to switch from traditional AC to the more efficient DC alternative.

Another aspect of why DC should be used instead of AC in buildings is the lowered consumption of conducting material such as copper and aluminum in distribution cables. Philips in the Netherlands has made an evaluation of AC and DC in building electrical networks [5]. The evaluation shows that given the same standard voltage drop of 5% and thermal limits, the use of DC results in 37% of the AC copper area when comparing single-phase 230 V AC and 380 V DC and 44% of the AC copper area in a corresponding comparison between three-phase 380 V AC and two-phase 760 V DC, that is, ±380 V DC. Even if for practical reasons the use of conductive materials cannot be fully reduced, a reduction of the continuous energy losses in the distribution systems is achieved.

In addition to this, DC also delivers noise-free electricity. The DC voltage has zero frequency and cannot set off vibrations in lighting fixtures or other appliances. The AC voltage has a frequency of 50 Hz in Europe (60 Hz in North America), and in many installations and appliances, such as low-energy lamps or light tubes, the 50 Hz frequency causes vibrations, which in varying degrees give rise to a humming sound or more or less intense noise. Even if the noise is not loud, it is always present, is more or less audible, and creates hidden stress, which affects different people to different degrees.

A disadvantage of using DC in buildings is that it has no zero crossing and automatic electric arc extinguisher such as what AC has. This makes DC harder to break compared to AC and could result in electric arcs when breaking the current in inductive loads. In aircraft with extensive DC power grids, electric arc detection has long been in operation. In building installations, loose connections in junction boxes and distribution boxes can exist both in AC and DC systems. In both cases, electric arcs can be a problem and cause fire or disturbance. There are already products for dealing with the problem of arcs, and new ones are under development in many places, mainly for the installation of solar cell systems. Standards for electric arc protection in solar cell equipment are being drawn up in the International Electrical Commission (IEC) after a number of major fires in solar cell systems [3].

28.3 DATA CENTER POWER DESIGN

Computers are machines that process and store information in a logical manner so that the information can be retrieved as needed from the computers. To insure the quality of this information, computers need to be constantly operational, even in cases where the power is lost. In laptops, a battery that is included in the computer itself solves this, and the laptop is powered with DC via a rectifier in the cabling to the laptop computer.

Servers in data centers do not have this ability to assure uninterrupted operation. Therefore, data centers are normally powered via uninterruptible power supplies (UPSs), whose primary purpose is to ensure this constant operation by feeding the computers via a battery (or a flywheel; see in the following), which is always in operation and takes over operation during a certain amount of time so that the computers can be properly shut down or reserve power can be started.

This UPS serves as an extension of the power grid. From the UPS, current is fed to the computer's PSU. The PSU then supplies 12 and 5 V DC to the server loads.

Data centers are traditionally powered with AC.

28.3.1 Traditional Data Center Powered with AC

Figure 28.1 shows how a traditional data center is powered. The standard input in Sweden is three-phase 400 V AC from a transformer connected to the grid and 230 V AC to appliances, that is, servers in a data center. Large data centers are often connected directly to dedicated 12 kV high-voltage lines.

A centralized UPS powers all data equipment with AC where the incoming AC is converted to 12 or 5 V DC.

The energy storage in the UPS can be either a battery or a kinetic flywheel. A kinetic flywheel is an alternative to batteries for storing energy that is sometimes used in data centers. Most data centers also have diesel generators for backup power.

28.3.2 DC as an Alternative Solution

Since most of the IT equipment in the data center operates on DC, as well as the batteries used for energy storage in the UPS, a proposed alternative to the traditional AC design is to use DC throughout the data center. Thereby, several of the conversion steps between AC and DC can be removed, and this will result in higher system efficiency. The definition and concept of the DC solution are presented in the following.

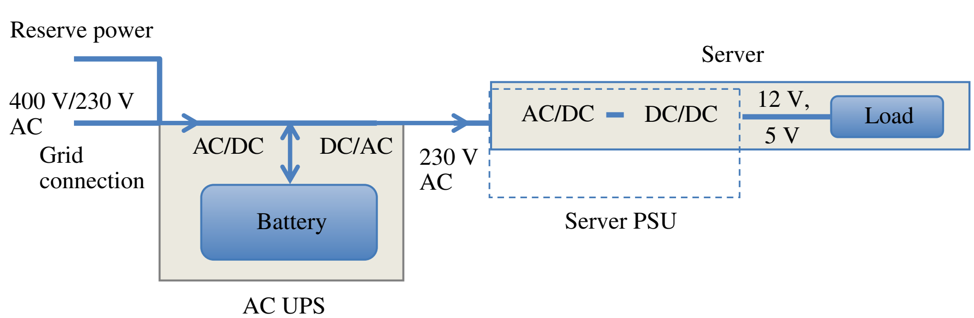


FIGURE 28.1 Schematic sketch of the power supply system in a traditional data center with UPS and PSUs powering the servers. Note: 230 V AC from the power grid is the standard in Sweden.

28.3.2.1 Definition of DC System Figure 28.2 shows the proposed solution of a DC-powered data center. The conversions marked with crosses can be removed with this alternative design. A data center normally fed from a three-phase line has an isolating transformer either connected to high voltage (10 kV) or low voltage (0.4 kV) feeding 230 V AC to rectifier modules. Thereafter, the current is passed through a UPS and after that directly fed into the servers through the PSU.

Figure 28.3 shows the concept of the PSU, fed with DC instead of AC. Using the DC solution can save the energy that is lost in the filter and/or preregulator in the corresponding AC PSU.

28.3.2.2 Why 380V DC Investigations on using DC in data centers sometimes refer to 48 V DC and sometimes 380 DC. The reason for using 48 V roots back to the telecommunication industry where 48 V has a long history of usage. Therefore, know-how and components for designing 48 V DC solutions are easier to access.

However, the 380V DC system has higher efficiency. This is because higher voltage compared to 48 V DC results in lower current and therefore lower losses and reduced cable area in the data center [6]. Therefore, in Europe, 380V DC is the current industry trend.

28.3.2.3 Rack-Level or Facility-Level Conversion There are several different approaches to system design in a DC-powered data center. The most common approach is to make the conversion at the entrance level, here referred to as the facility level, and thereafter distribute 380V DC to the servers.

The other alternative is to use a standard AC UPS combined with a rack-level AC/DC converter. Low-voltage DC is then fed to the servers. This approach however results in more conversions compared to the more efficient facility-level converter. The different approaches to system design are demonstrated in Figure 28.4 [7].

28.4 WHY USE THE DC SYSTEM IN DATA CENTERS

There are several parameters to take into consideration when comparing the traditional AC system design to the DC solution. Environmental and financial aspects as well as reliability and safety are among the most vital aspects that must be taken into consideration. The pros and cons of the DC technology will be discussed in this part taking these factors into account.

28.4.1 Efficiency

A system with high efficiency is key to keeping operating costs as low as possible and reducing the carbon dioxide emissions. Table 28.1 shows the results in efficiency of the two different systems from the study carried out by Netpower

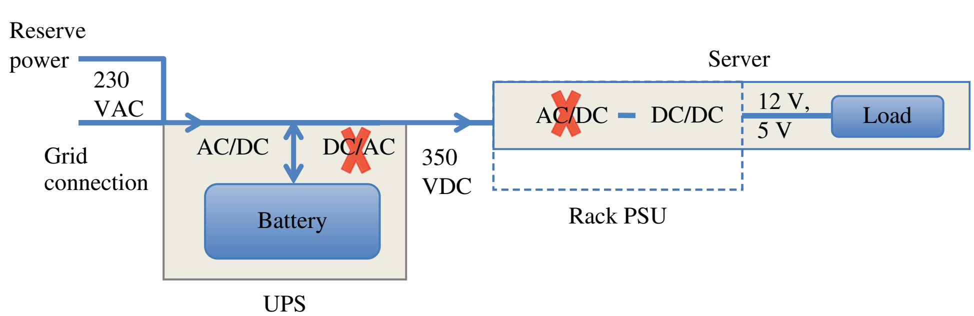


FIGURE 28.2 Schematic sketch of a data center powered with DC. The conversions marked with crosses are removed with this alternative design. Note: 230V AC from the power grid is the standard in Sweden.

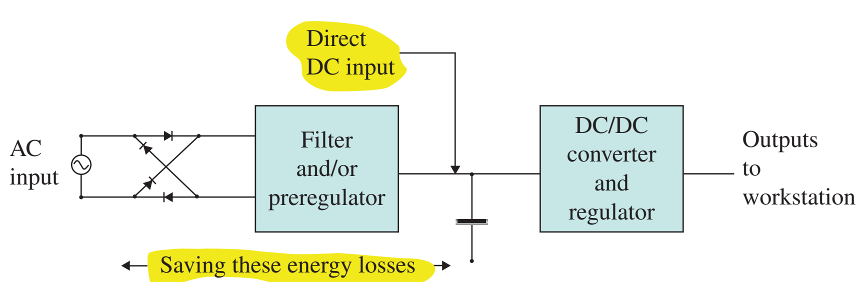


FIGURE 28.3 Schematic sketch showing which conversions can be eliminated in the DC PSU compared to the AC solution.

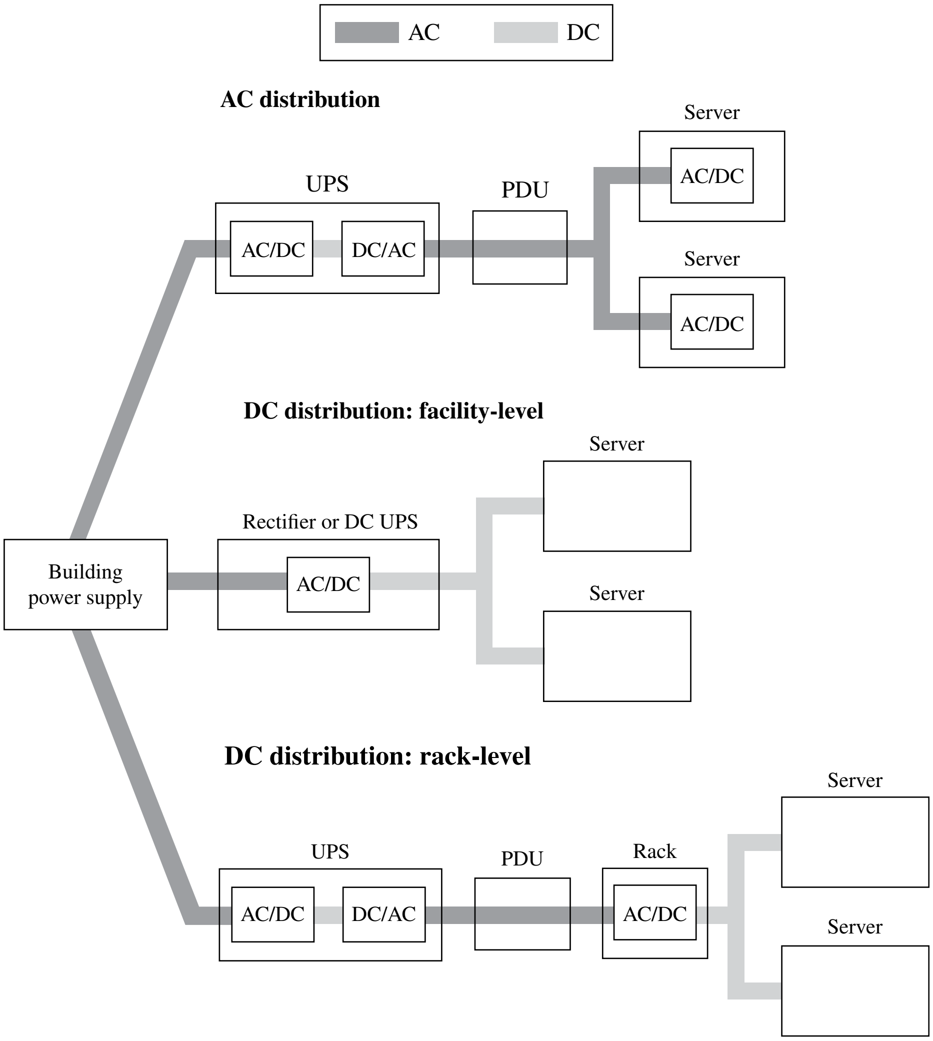


FIGURE 28.4 Data center power distribution systems [7].

TABLE 28.1 Results from measurements and calculations for the AC and DC systems [8]

Alternative	AC system	DC system
Efficiency UPS	0.91	0.97
Efficiency PSU	0.74	0.93
Total efficiency	0.67	0.9
Total power usage (W)	5,300	4,000
Electrical energy usage per year (kWh)	47,000	35,000

Labs, Uppsala University. The experiment was made on an IBM XIV Storage System where the energy efficiency for the traditional AC solution was compared to an alternative DC design. Measurements of power input and output at different loads were made after which the efficiency was calculated. The table shows the efficiency and the energy usage per year at normal operation.

As shown in Table 28.1, the efficiency for the DC system was 0.9 compared to 0.67 for the traditional AC approach in this study.

Table 28.2 shows a comparison of a traditional AC system to the DC alternative from a report published in 2008 by the Lawrence Berkeley National Laboratory. In this study, two AC systems, one referred to as a typical AC distribution system and the other as the most efficient AC system available, were compared to a DC distribution system. This project was a joint venture between the university and several industry experts. The results indicate an efficiency of 0.85 for the DC system compared to 0.61 for a typical AC system and 0.79 for the most efficient AC system. The efficiency of AC versus DC systems was measured during a year of operation [9].

The reduced electrical energy usage per year results in reduced cooling need for the facility. One kilowatt-hour used in the data center corresponds to 1 kWh cooling needed.

TABLE 28.2 Results from the Lawrence Berkeley National Laboratory [9]

System efficiency	UPS efficiency (%)	Transformer efficiency (%)	PS efficiency (%)	System efficiency (%)
AC typical distribution efficiency	85	98	73	61
DC distribution efficiency	92	100	92	85
Energy consumption	Compute load (W)	Input load (W)	Efficiency gain	
AC typical distribution efficiency	10,000	16,445		
DC distribution option (optimized)	10,000	11,815		
% energy consumption improvement versus typical AC distribution			28.2%	

TABLE 28.3 Control parameters for the AC solution and the DC solution

Traditional AC solution	DC solution
1. Voltage	1.Voltage
2. Frequency	2. None
3. Phase	3. None
4. Waveform	4. None
5. Electronic switch	5. None

28.4.2 Reliability

Reliability is one of the most important parameters for data center owners and is defined as the ability of a system to perform its required functions under stated conditions during a specified period of time [10]. Governments, organizations, and businesses rely on their IT systems to be operating 24/7.

DC has fewer control parameters compared to AC, which results in fewer components that can fail. Table 28.3 shows the difference in numbers of control parameters for the AC and DC systems.

AC is characterized by various control parameters such as voltage, phase, frequency, and waveform, which need to be considered. DC, on the other hand, is very simple with only voltage as control parameter. This leads to simplified design of the power supply solution of the data center and thereby safer and more reliable operation [11]. The probability of failure over 5 years is 6.72% for the DC solutions compared to 13.63% for the traditional AC solution according to reliability predictions made by Intel Labs in 2010 [12].

Turning to a practical measurement of reliability, Japanese NTT Facilities performed a study between 1995 and 2005 where they measured the number of outages of an AC UPS and a DC UPS in operation. The results were that the DC system had significantly higher reliability compared to the traditional AC UPS. The study was based on statistics of 23,000 DC systems and 10,000 AC systems. The DC systems had nine years of no operational failure, while the AC system had at least one failure each year.

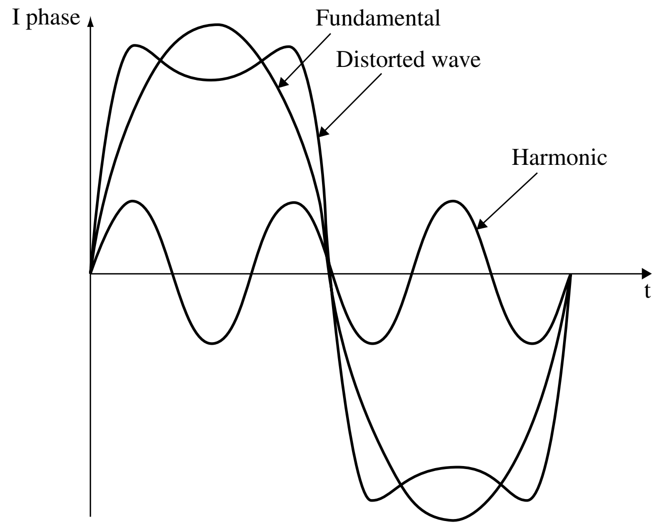


FIGURE 28.5 Example of electrical harmonics.

28.4.3 Redundancy

Redundancy is the duplication or in some cases triplication of components of functions in a system in order to increase reliability in case of component failure. It is easier to design and construct a redundant DC system compared to an AC system with the same redundancy. This is because the DC UPS can be connected directly in parallel and there is no need for phase synchronization. AC needs to be adjusted and synchronized when connecting several AC devices to one another.

28.4.4 Harmonics

Harmonics are an alteration of the normal sinusoidal waveform caused by nonlinear electrical loads such as computers, printers, and fluorescent lighting. An example of a nonsinusoidal harmonic distortion is shown in Figure 28.5.

Figure 28.5 shows an example of electrical harmonics. The harmonics current results in power quality problems and generates heat in a complex AC environment such as a data center powered by AC [11]. This problem is dealt with today using power factor-corrected equipment [13].

Since there is no sinusoidal wave in the corresponding DC system, this problem does not exist in this case.

28.4.5 Fault and Leakage Currents

A fault current is an abnormal current that occurs in an electric system and can cause disturbances. The system is protected from fault currents by electrical fuses. Fault currents are easier to break with AC compared to DC. Therefore, the DC system needs electrical fuses specially adapted for DC.

Leakage currents can occur in poorly isolated electrical equipment and can become hazardous if a person comes in contact with a conductive part of the equipment. Conductive parts will normally be connected to a protective conductor. As long as the protective conductor is without fault, it will eliminate the risk of leakage current. However, if the protective conductor fails, a hazardous leakage current may occur when a person touches the equipment. These currents sometimes occur via mains filters. Capacitors are often used in mains filters causing a leakage current path for AC, however not so for DC. The tendency of higher leakage currents, or rather protective conductor currents, in AC systems may in some cases also prevent a desired use of residual current devices (RCDs) [14].

28.4.6 Scalability

With the growing need for computer capacity comes a need for flexibility in terms of the installed base of servers in the data center. Both the DC and the AC systems can be built in a modular way that enables scalability. However, by using DC, the up- or downscaling is simplified since there is no need for synchronization and extra AC/DC conversion steps.

28.4.7 Standards

Because of the increased interest in the use of DC in data centers as well as buildings, so-called microgrids, there are industry associations such as EMerge Alliance working jointly for a common standard. This alliance consists of several large data center equipment manufactures as well as universities and governmental agencies.

The IEC, which is the global responsible standardization body for the use of electricity and all its relevant Technical Committees (TCs), is heavily involved in standardizing for the modern use of DC. Also, the European Telecommunications Standards Institute (ETSI) is standardizing for use of DC.

28.4.8 Safety

When it comes to personal safety, DC has several advantages over AC. A person touching an AC cable can be affected by cardiac muscle cramps and will have problems disconnecting from the cable. This is because of the frequency in AC. On the other hand, when touching a DC cable, a person will experience an electric shock and immediately move away from the cable [15].

One of the common arguments against the usage of DC is the problem of breaking the current. The first electric grids were all providing DC to the households. DC was then gradually replaced with AC because of troubles breaking the DC current, something that sometimes caused fire accidents. But the breaking of the current is only problematic when inductive loads (motors, heat radiators) are powered with DC. Most electric equipment in today's data centers are combined capacitive and resistive loads (computers, compact fluorescents lamps, etc.), also called electronic loads [8]. In electronic load, there is always energy stored in the capacitors. When breaking the connection to the grid, energy to the load will be taken from the capacitor storage and not from the grid. Therefore, when supply is switched off, there will be no flash or sparking in a switch or the plug and socket set.

28.4.9 Environmental Impacts

One very important aspect of this new solution for the data center power design is the environmental impact. To keep the data center eco-friendly with as low carbon footprint as possible is key to most data center owners of today.

The increased efficiency for the DC system compared to the AC system that is traditionally found in today's data centers results in lowered environmental impact. Depending on the energy mixture, for example, energy sources available, at the location of the data center, the reduction in carbon dioxide emissions will differ.

Energy production from fossil fuels such as coal, gas, and oil is however still the dominant energy source used for electricity production in the world. A more efficient use of this electricity implies lowered carbon footprint.

Furthermore, fewer components in the power solution design results in lowered environmental impact of the data center from a life cycle perspective.

Moreover, the environmental impacts for producing and operating a larger cooling apparatus are lower with the DC solution compared to the traditional AC system.

28.4.10 Cost Justifications

There are two different economical aspects to be considered when designing a data center: capital costs and operating costs.

28.4.10.1 Operating Costs Replacing the different conversion steps in a traditional AC design by one conversion (AC to DC) achieves the largest part of the power savings.

Other studies within the same field have concluded that financial saving from a 5.5 MW data center can be US$150,000 per year when converting from AC- to DC-based design [12].

TABLE 28.4 Operating costs per year [8]

Alternative	AC	DC
Costs per year including cooling (US$)	7000	5200

Several studies have been carried out measuring the energy usage on an AC versus a DC system for a data center. Table 28.4 shows a comparison between the two systems from the study on one IBM XIV Storage System carried out by Uppsala University. The price per kilowatt-hour refers to the Swedish electricity prices in 2011.

From Table 28.4, we can read that for one storage system, US$1800 per year in operating costs can be saved. This results in savings in operational costs of 25%. If the total amount of servers and storage servers in a large data center are added up, the total savings can be considerable.

28.4.10.2 Capital Costs The DC solution implies fewer components and thereby reduced production costs and lowered capital costs. According to a report released by Intel Labs in 2010, the usage of DC in data centers will result in 15% savings in electrical facility cost [12]. The reduction in the cooling needed will result in lowered investment costs for the cooling apparatus.

However, since AC is the standard today, there are many vendors supplying AC equipment. As for the DC-based UPS, several companies are investing in this technology and are offering solutions in the market. Several large server vendors are also investing in research and development of these new solutions and are now beginning to offer DC-based PSUs in the market.

An additional cost when converting to DC operation will be to educate and train the technicians working with operation and development of the system.

28.4.11 Space Savings

Another parameter that is also often referred to is the possible space savings with the DC solution compared to the traditional AC design. The saving in floor space is made possible due to fewer components and reduced cooling. The space savings can be as high as 33% for the DC design compared to the AC design, according to a study by Intel Labs [12].

28.4.12 Integration with Renewable Energy Sources

A strong argument in favor of using DC in data centers is the enhanced integration with renewable energy sources and fuel cells. There are several benefits using DC in the data center when integrating renewable energy sources to the power supply system. This is because many conversion steps can be removed (Fig. 28.6), compared to the traditional setup where AC is used in the data center [16].

Studying the image in Figure 28.6, with the renewable energy sources integrated to the data center power design, reminds one of a so-called smart grid. The smart grid has several definitions but can in a simplified way be described as a small-scale power grid with its own built-in intelligence. The intelligence refers to the grid's utilization of locally produced energy from renewable energy sources when they are available and the mains when the intermittent energy sources are unavailable. The utilization of the energy produced by

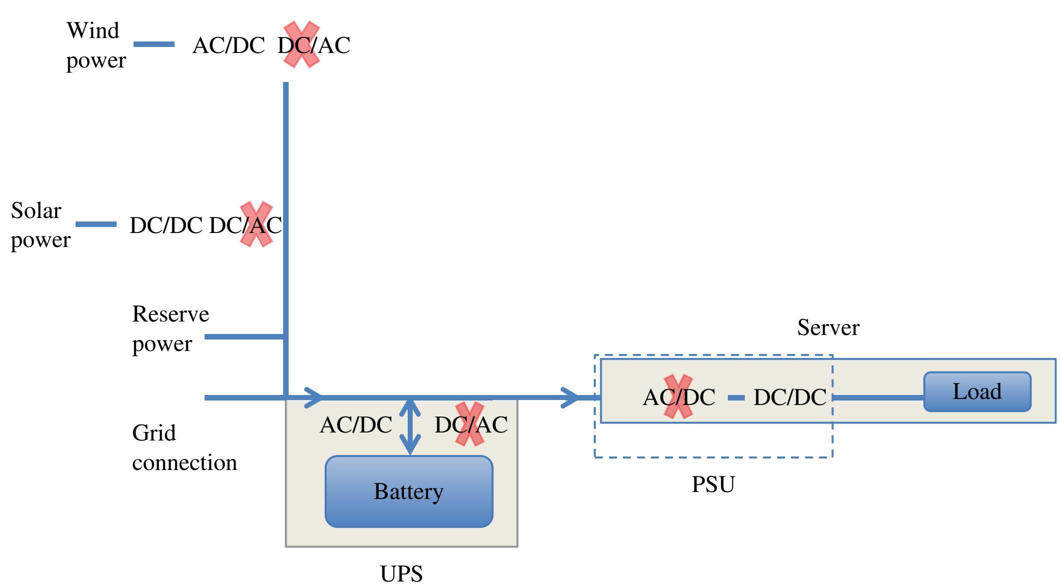


FIGURE 28.6 Schematic sketch of a DC-powered data center with renewable energy sources integrated into the design. The conversions marked with crosses are removed with this alternative design.

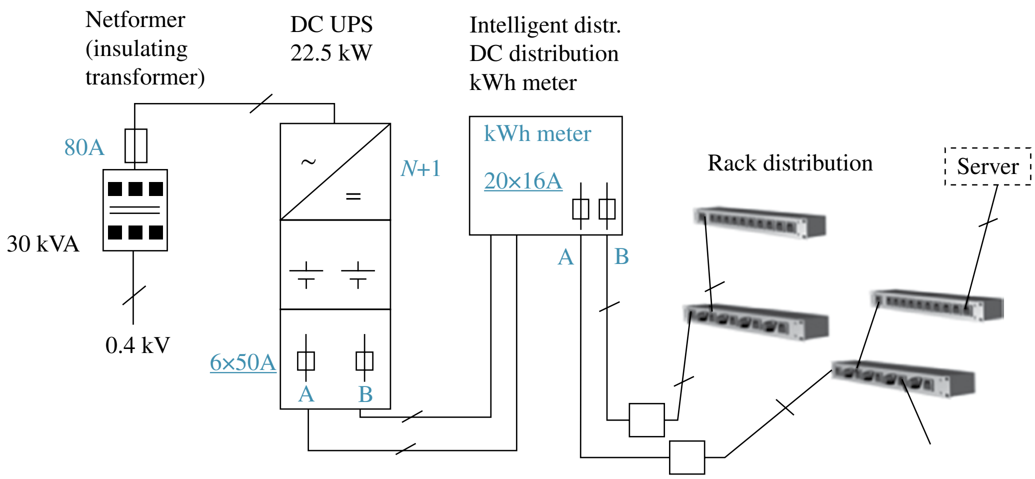


FIGURE 28.7 System power design at the Swedish Energy Agency.

the renewable energy sources locally, where it is produced, is key to keeping the losses as low as possible.

A data center is ideal for the implementation of a small-scale smart grid operating on DC combined with renewable energy sources since the load is relatively stable 24/7 and all the devices operate on DC. The solar cells or wind turbines can also be attached directly to the facility, and thereby, the losses that occur in longer-distance distribution are eliminated. There are several examples of where solar cells have been integrated into a data center. This has resulted in lowered operating costs and reduced environmental impact [3].

28.5 EXAMPLES OF DC DATA CENTERS IN OPERATION

More and more companies, organizations, and governments around the world are installing the DC solution. The approach whether it is 380 or 48 V DC and the details of where the power conversion steps are being made may vary, but the aim is still to reduce energy consumption through simplified design. The following are three examples of DC installations in operation.

28.5.1 The Swedish Energy Agency

"Within the Energy Agency we have a vision of a sustainable energy system, and that we shall live as we learn and be the most energy efficient government agency in Sweden" (Bjorn Lundqvist, CIO).

The Swedish Energy Agency is located in Eskilstuna, south of Stockholm. The DC UPS system was installed in 2010. All the servers and storage servers in the data center are powered with 380 V DC. Furthermore, they have installed solar panels with a maximum output of 12 kW powering the data center during summertime. The DC UPS and solar panels will in the near future power part of the office LED lighting and equipment. The combination of DC, solar panels, virtualization, and blade server technology has enabled a 45% reduction of energy usage at the Energy Agency [8].

Figure 28.7 illustrates the schematic layout of the power design in the data center at the Swedish Energy Agency.

28.5.2 SAP Data Center, Palo Alto, United States

In the design of the SAP Data Center in Palo Alto, several innovative solutions have been applied in order to reduce energy consumption and thereby the environmental impact from the business and data center operation. One of them is to let the data center run on 380 V DC. By implementing DC throughout the data center, the estimated savings are between 15 and 20%. Furthermore, there are rooftop solar panels that are connected to the data center, which results in savings in energy consumption by 30–40%.[1]

28.5.3 NTT Group, Japan

The Japanese telecom company NTT has been active in research and development of solutions for DC-powered data centers for several years and are pioneers in this field. Their own measurements indicate savings in energy usage of 15%. NTT has five data centers in the Tokyo region operating on 380 V DC.[2]

[1] www.greentechmedia.com/articles/read/the-worlds-best-green-technology, November 28, 2012

[2] http://www.ntt.co.jp/kankyo/e/protect/greenbyict/index.html, December 1, 2012

28.6 FUTURE TRENDS AND CONCLUSIONS

Global warming as a result of our emission of carbon dioxide and the rising electricity cost combined with the reliability that we demand from our information technology systems will leave us with the need for more efficient data centers also in the future. The focus areas up to now have been through concentrating of computing power through virtualization and new and more efficient cooling methods. But after those measures have been taken, there is now a need to redesign the power supply into the most efficient possible solution.

The possible savings with DC compared to AC fluctuate with different investigations and of course depending on which AC system you are comparing with which DC system. The highest numbers are between 25 and 30% in reduction in energy usage, while the lowest are around 5%. Even though the savings might be 5% in some cases, one must remember that this is a considerable number in an industry that uses as much energy as the air traffic and where the electricity bill is said to be the second highest cost after personnel costs.

Having said this, there are still some obstacles to be met before DC can become the standard for data centers around the world. The most obvious one is if or when more of the server vendors will start to offer DC-fed servers at large scale. There are indications that some large server vendors are now beginning to offer 380 V DC-fed servers.

Another perhaps smaller obstacle is that the technicians working in installation and operation of the data center have to be trained in using DC.

As for safety-related concerns with DC that are often raised, there is no fire hazard with using DC in data centers, and DC can cause less harm to the human body compared to AC.

A very obvious benefit of the DC design for the data center is the integration to renewable energy sources, LED lighting, and other devices that generate or operate on DC.

Finally, achieving acceptance for new ideas and technologies is always a time-consuming process, especially ideas that challenge something that has been in practice for decades and that concerns something as important to us as our base for sharing and storing information. But nevertheless, the era of DC for the data center will come.

ACKNOWLEDGMENTS

Special thanks to John Akerlund M.Sc. E.E. for being my mentor during the writing of my Master thesis and this chapter and to Ben Amaba, PhD, PE, CPIM, LEED® AP BD+C, IBM Complex Systems—Rational World Wide Executive for mentoring and guidance.

REFERENCES

[1] Schneider D. Edison's final revenge. *American Scientist*; April 2008.

[2] Padiyar KR. *HVDC Power Transmission System*. New Delhi: New Age International; 2005.

[3] Akerlund J. Investigation of a micro DC power grid in Glava Hillringsberg—a smart grid. Stockholm: UPN; 2012. p. 10.

[4] Revolutionary power: direct current in buildings. Siemens Innovation News, IN 2012.07.3e; 2012. Available at http://www.siemens.com/innovation/en/news/2012/e_inno_1222_1.htm. Accessed on September 19, 2014.

[5] Boeke U, Wendt M. Comparison of low voltage AC and DC power grids. Eindhoven: Philips Research; 2010.

[6] Tomm A, Annabelle P, Pavan K, Dwight D, Guy A. Evaluating 400 V DC distribution in telco and data centers to improve energy efficiency. Hillsboro: Intel Labs; 2007.

[7] Power distribution cuts data center energy use. California Energy Commission; 2008.

[8] Bergqvist S. *Energy Efficiency Using DC in Data Centres*. Uppsala: Uppsala University; 2011. pp. 9, 12, 15.

[9] Brian F, My T, William T. DC power for improved data center efficiency. Berkeley: Lawrence Berkley National Laboratories; 2008. p. 4.

[10] Dunn WR. *Practical Design of Safety-Critical Computer Systems*. Solvang: Reliability Press; 2002. p. 233.

[11] Murrill S. Evaluating the opportunity for DC power in the data center. Stamford: Emerson Network Power; 2010. p. 6.

[12] Albridge T, Kumar A, Dupy D, AiLee G. Evaluating 400 V direct-current for data centers. Hillsboro: Intel Labs; 2010. pp. 10, 11.

[13] Spitaels J. Hazards of harmonics and neutral overloads. Palatine: Schneider Electric; 2011. p. 5.

[14] Mannikoff A. Advances in Measurement Systems. Croatia: In-Tech, www.intechweb.org; 2010.

[15] Akerlund J, Boije af Gennas C, Olsson G, Rosin D. One year operation of a 9 kW HVDC UPS 350 V at Gnesta Municipality Data Center. INTELEC 2007; Rome; 2007. p. 6.

[16] Savage P, Nordhaus RR, Jamieson SP. DC grids: benefits and barriers. New Heaven: YALE School of Forestry and Environmental Studies; 2010. p. 54.

FURTHER READING

EMerge Alliance. Available at www.emergealliance.org. Accessed on May 22, 2014.

Innovations in DC technology by ABB. Available at http://www.youtube.com/watch?v=_KeMgVNlSPQ. Accessed on May 22, 2014.

Rasmussen N. AC vs. DC power distribution for data centers. Boston: Schneider Electric; 2012.

The world's best green technology? Available at greentechmedia.com. Accessed on December 9, 2010.

29

RACK PDU FOR GREEN DATA CENTERS

Ching-I Hsu
Raritan, Inc., Somerset, NJ, USA

29.1 INTRODUCTION

The rack power distribution unit (PDU) is emerging from obscurity. As the last link of the elaborate data center power chain, the traditional role of the rack PDU has been to deliver stable, reliable, and adequate power to all the devices in the rack or cabinet—servers, storage, network equipment, etc., which are plugged into it. And while it provides the electrical heartbeat to all the systems that run the critical applications that support the operation of the business (or that, in some cases, are the business), it was often considered a simple commodity—just a power strip. Typically, IT merely told facilities how much power was needed, based on device nameplate specs and often with redundancy, so there was plenty of headroom and minimal risk of downtime. Little thought was given to efficiency or what other value a rack PDU could provide.

That was yesterday. Over the past few years, system availability has become a "given," and now, data center management attention is being focused on operational costs, efficiency improvements, and resource optimization. With the annual expenditure for powering the average data center surpassing the cost to purchase the equipment (ITE) itself, the use (and waste) of energy is now targeted as a priority. And beyond the actual cost to power the data center, there are the related issues that impact both current operations and future expansion—for example, physical space and utility power availability, CO_2 footprint, and potential government regulation. Since almost all of the power delivered from the utility to the data center is consumed either directly by the devices plugged into rack PDUs or indirectly by the infrastructure to bring power to the rack and cool the devices, the once obscure rack PDUs have become visible on the data center management radar.

Not surprisingly, many of the major strategies to address the above issues and improve overall data center efficiency depend on new capabilities not available in commodity outlet strips:

- To maximize the use of data center space and other resources, there has been a trend to deploy racks densely packed with 1U servers or power-hungry blade servers. So today's rack PDUs typically handle loads of 5–10 kW with 20 outlets compared to 2–3 kW with 8–12 outlets of a few years ago; and there are PDUs now designed to support 20+ kW and 40+ outlets.
- To increase IT staff productivity and conserve energy by employing lights-out and/or remote data center operation, some rack PDUs provide real-time monitoring, reporting and alerts, as well as secure, reliable outlet switching.
- To identify ghost (no function), underutilized, or grossly inefficient servers for elimination, replacement, consolidation, or virtualization, rack PDUs provide individual outlet monitoring.
- To create individual awareness, accountability, and/or charge-back for power usage and CO_2 footprint, some rack PDUs are equipped with highly accurate, real-time power measurement capability at the PDU and outlet levels.
- To optimize IT workload and make informed decisions for infrastructure capacity planning, IT and facilities management need rack PDU management software that continually collects data on power consumption, analyzes trends, and correlates with IT workload data.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

These are but a few of the reasons that the selection of rack PDUs has become important.

A wide variety of rack PDU configurations is available based on parameters such as the number of phases, voltage, total ampere, branch circuits, number of outlets, socket type, plug type, rack units consumed, and physical dimensions. Beyond the functions of the *basic* rack PDU, additional capabilities are available in rack PDU categories, or types, we call *metered*, *switched*, and *intelligent*. Furthermore, if you cannot find an off-the-shelf rack PDU that matches your specific requirement, some vendors will assemble or even design a custom rack PDU (also called BTO/ETO: built-to-order/engineered-to-order).

	Basic	Metered	Switched	Intelligent
Rack power distribution	●	●	●	●
Rack PDU-level metering of current	○	●	◎	●
Remote outlet switching	○	○	●	●
Outlet-level metering of current, active power, apparent power, kWh	○	○	○	●
Encryption and secure access	○	○	◎	●
Directory services and user mgmt	○	○	◎	●
IP and SNMP accessibility	○	◎	●	●
Environment sensor support	○	◎	◎	●

● Almost always
◎ Sometimes
○ Almost never

FIGURE 29.1 Types of rack PDUs. Courtesy of Raritan, Inc.

29.2 FUNDAMENTALS AND PRINCIPLES

ITE is normally mounted in racks or cabinets with provisions for all necessary cables, ventilation, cooling, and convenient access. Previous chapters of this handbook have discussed the large data center PDUs that are used earlier in the power chain and take the form of panel boards mounted on walls or freestanding pedestals. In this chapter, we're discussing only the rack PDU, at the end of the chain, which supplies power to the ITE in the rack. *Unless otherwise stated, any reference to "PDU" for the remainder of this chapter means "rack PDU."*

Rack PDUs come in many configurations with respect to the number and type of receptacles, voltage, load capacity, and physical mounting (horizontal or vertical). A unit may perform no function other than providing power to the devices plugged into it; or it may also provide additional functions—for example, turning power off and on remotely, monitoring power consumption, and sensing the temperature in the ITE rack.

29.2.1 Overview and Class of Devices

A rack PDU is mounted in an ITE rack and provides electrical power to various IT devices such as servers, networking, and storage equipment. Today, rack PDUs are available in a number of configurations. We describe in the following the basic characteristics of four types of rack PDUs using Frost and Sullivan's classifications as a general guide (Fig. 29.1). In Section 29.4.3, we will discuss the strengths and weaknesses of each PDU type as well as their typical applications.

29.2.1.1 Types of Rack PDUs

Basic PDUs Basic PDUs are power strips constructed out of high-quality components for use in critical environments such as data centers. They distribute correct voltage and current to multiple outlets.

Metered PDUs Metered PDUs measure the current draw (load) at the PDU level and display the data locally. More sophisticated models offer user-defined alarm functions and remote access to the data over a serial or network port.

Switched PDUs Switched PDUs offer the features of metered PDUs and also provide controlled on/off switching of individual outlets and load metering (see Metered PDUs) at the PDU level. They enable authorized users to securely power cycle devices remotely; and they may also provide a power sequencing delay as well as some outlet use management.

Intelligent PDUs Intelligent PDUs can be controlled remotely via a Web browser or command line interface (CLI). They may or may not be switched. They meter power at the PDU and individual outlet levels; support alerts based on user-defined thresholds; provide security in the form of strong passwords, authentication, authorization, and encryption; and incorporate environmental management capabilities. Some models are customizable; support industry standard-based protocols like Simple Network Management Protocol (SNMP) TRAPs/SETs/GETs, IPMI, and SMASH CLP; and integrate seamlessly to existing corporate infrastructures like Lightweight Directory Access Protocol (LDAP), Active Directory®, RADIUS, and NFS servers.

29.2.2 Electrical Power Distribution to the Rack

29.2.2.1 Branch Circuits Power is distributed to the rack over one or more electrical branch circuits. Branch circuits are power feeds that originate from a panel, switch, or distribution board and terminate into an electrical receptacle mounted in a junction box near the ITE rack. Branch circuit wiring can be overhead, underneath a raised floor, or both. The rack PDU itself could have multiple branch circuits. See Section 29.2.5 for details regarding branch circuit protection requirements.

29.2.2.2 Branch Circuit Load Capacity The power that can be delivered by a branch circuit depends on the electrical characteristics of the circuit. A key factor in delivering power to a rack is whether the power is single phase or three phases. The amount of electricity delivered to a rack is often referred to as the load capacity and is the product of the rated voltage and the rated current and is presented as volt-amps (VA) or kVA (VA × 1000). Given the rated voltage and current, the load capacity that can be delivered by a branch circuit is determined using these formulas:

- Single-phase: load capacity = rated voltage × rated current
- Three-phase: load capacity = $\sqrt{3}$ × rated voltage × rated current

29.2.2.3 Branch Circuits: Rated Voltage The rated voltage of a branch circuit specifies both its magnitude (volts) and number of phase conductors (Table 29.1). Single-phase wiring is straightforward and consists of two wires (plus safety ground) where the AC voltage is a single sinusoidal wave as measured across the two wires.

Three-phase wiring is more complicated and consists of either three (three-phase conductors) or four (three-phase and one neutral) wires, plus safety ground (Fig. 29.2). Three-phase branch circuits deliver more power but require a rack PDU specially designed for three-phase branch circuits. Internally, a three-phase rack PDU divides the three or four branch circuit wires into pairs of single-phase circuits—and these single-phase circuits are wired to the rack PDU's single-phase outlet receptacles.

The three-phase conductors have the same voltage magnitude, but the sinusoidal AC waveforms are out of phase with each other by 120°. Regardless of the number of wires, the rated voltage of three-phase wiring is always the measured voltage difference between any two-phase conductor wires—not the difference between a phase wire and neutral. Just as with single-phase power described earlier, connecting across one 120 V hot line and the neutral provides 120 V AC. But connecting across any two 120 V hot lines, say, L1 and L2, provides 208 V AC, not 240 V AC. Why? Because the phase of L1 is offset 120° from L2, the voltage is not 240 V (120 V × 2), as it is for single-phase, but is 120 V × square root 3 or 120 V × 1.732 = 208 V. A three-phase PDU can deliver three circuits of 208 V each. Some rack PDUs take advantage of a neutral wire to provide three circuits of both 120 and 208 V. But as mentioned in the preceding paragraph, regardless of the number of wires, or whether or not both higher and lower voltages are supplied as outputs, a three-phase rack PDU is rated at the voltage

TABLE 29.1 Branch circuit rated voltage and wire requirements

Rated voltage	Location	# of wires	Outlet voltage(s)
120 V	North America	2 (phase + neutral)	120 V
208 V	North America	2 (phase + phase)	208 V
230 V	International	2 (phase + neutral)	230 V
208 V 3Ø	North America	3 (three-phase lines)	208 V
208 V 3Ø	North America	4 (three-phase + neutral)	Mixed 120 and 208 V
400 V 3Ø	International	4 (three-phase + neutral)	230 V

Source: Courtesy of Raritan.

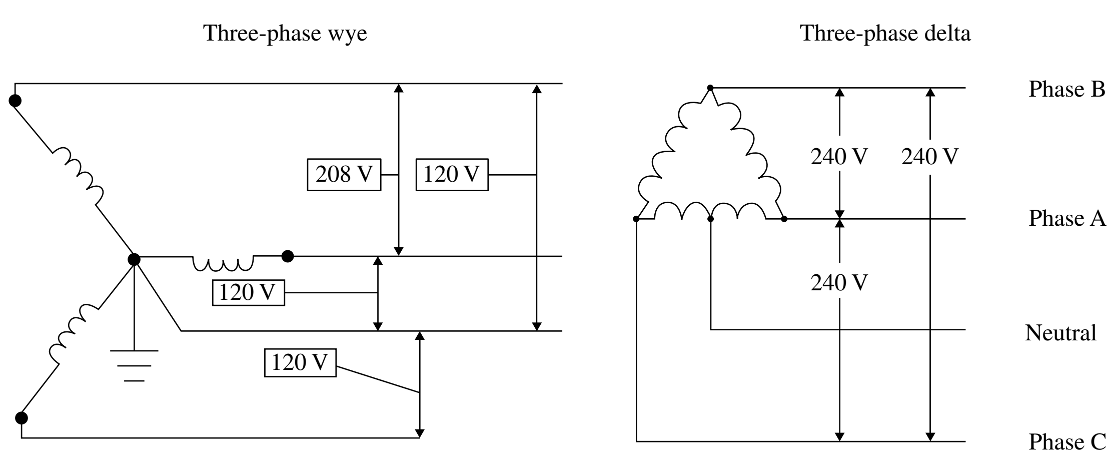


FIGURE 29.2 Three-phase wiring diagram. Courtesy of Raritan.

between two phases, for example, L1 and L2, which in the example here is 208V.

A rack PDU can also provide 400 VAC. Just as with the 208V three-phase rack PDU, if one of those lines is connected to a neutral instead of another line, this provides a single-phase output circuit that for a 400V rated PDU is 230VAC (400V/1.732=230V). This is a common deployment in Europe and is becoming more common for high-power racks in North America.

Three-phase rack PDU specifications often use the terms Wye and Delta or the Greek letters ϒ and Δ. These terms or letters were chosen because the electrical configuration diagram of a Delta transformer looks like a Δ and the electrical configuration diagram of a Wye transformer looks like a ϒ. A rack PDU that does not convert a higher input voltage, for example, 208 or 400V, to a lower output voltage, for example, 120 or 230V, but instead retains the higher voltage throughout uses a Delta transformer. A Delta transformer has three connection points, one at each corner of the triangle. Each of these points is a connection for one of the three lines. Connecting any point to any other point provides a line-to-line connection, for example, L1 to L2, and provides 208 or 400V as described in the earlier examples.

A rack PDU that does convert a higher input voltage to a lower output voltage uses a Wye transformer. A Wye transformer has three connection points for the lines, one at the end of each “arm” of the ϒ and one at the “foot” of the ϒ. The center intersection point of the ϒ is a fourth connection point and is where the neutral wire is attached. Connecting any two of the three line connections together, for example, L1 and L2, provides 208 or 400V. Connecting any one of the three line connections to the neutral, for example, L1 and neutral, provides 120 or 230V as described in the examples.

29.2.2.4 Branch Circuits: Rated Current The current flowing in a circuit is determined by the size (thickness) of its wire and terminating receptacle. Branch circuits are required to be overcurrent protected using a circuit breaker or fuse. The rating of the circuit breaker is sized to the current-carrying capacity of the branch circuit’s wiring and receptacle. For example, 10 AWG (American Wire Gauge) wire and a NEMA L21-30R receptacle are both specified at 30 A—so a circuit using these components must be protected by a 30 A circuit breaker.

In North America, the National Electric Code (NEC) for data centers (NEC Article 645) requires branch circuit wiring to be rated 125% greater than the total connected load. To insure this requirement is met without running heavier gauge wires, all electrical devices (rack PDUs, computers, etc.) used in North American data centers must be certified to Underwriters Laboratories (UL) 60950-1. UL 60950-1 limits a device to draw no more than 80% of the rating of its input plug. For example, a rack PDU containing a 30A NEMA L21-30P plug must not draw more than 24A. This 80% limitation is commonly known as “derated” current.

Table 29.2 summarizes power available for various branch circuits.

29.2.3 Plugs, Outlets, and Cords

Rack PDUs are available with several types of plugs and receptacles (or outlets), designed so that only the appropriate rack PDU plug will fit into the appropriate circuit outlet and only the appropriate device plug will fit into the appropriate rack PDU receptacle. This is done to protect equipment, for example, so a device designed for 120V only isn’t plugged into a 208V circuit, and for safety reasons, for example, a server that draws 30 A doesn’t overload a circuit designed to handle only a maximum of 15 A.

The major classifications of plugs and receptacles in data centers are the International Electrotechnical Commission (ICE) and National Electrical Manufacturers Association (NEMA).

TABLE 29.2 Branch circuit available power

Location	Rated voltage	Rated current (A)	Derated current (A)	Available power/ branch circuit (kW)
North America	120V 208V 208V 3Ø	20	16	1.9 3.3 6.7
International	230V 400V 3Ø	16	16	3.7 11.0
North America	120V 208V 208V 3Ø	30	24	2.9 5.0 8.6
International	230V 400V 3Ø	32	32	7.4 22.1

Source: Courtesy of Raritan.

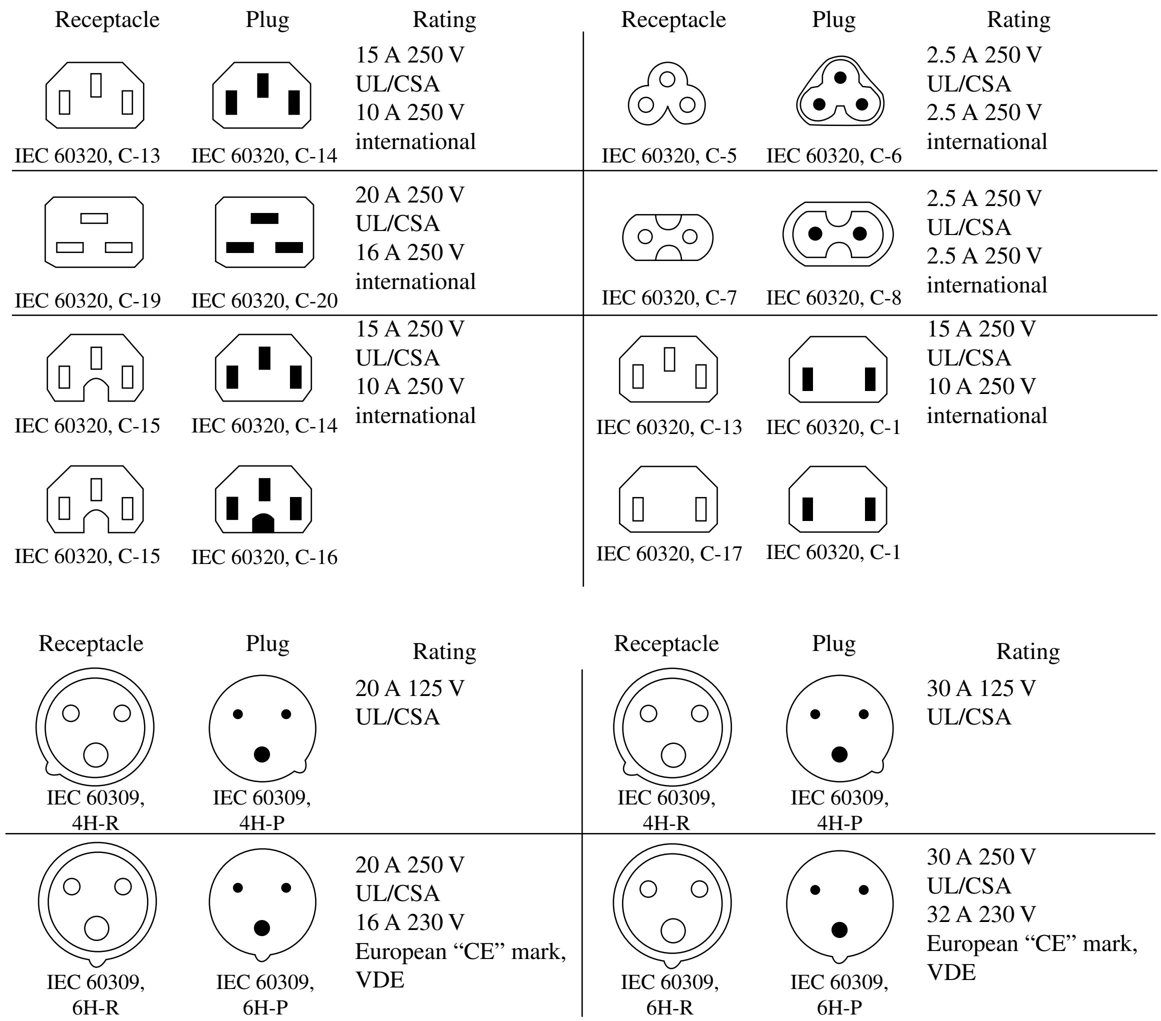


FIGURE 29.3 IEC plugs and receptacles. Courtesy of Raritan, Inc.

IEC plugs and receptacles (Fig. 29.3) are most common in Europe, and NEMA plugs and receptacles (Fig. 29.4) are most common in North America. However, many data centers in North America use IEC plugs and receptacles, and there are many families of plugs and receptacles in use in data centers around the world.

A significant concern in data center power distribution is unintentional disruption of power by accidentally disconnecting cords. Solutions exist that lock the plug into the receptacle and prevent the cord separating from the receptacle. There are three methods of securing the plug in the receptacle:

- Plug with tabs snaps into the receptacle locking them together
- Plug inserted into a receptacle with a locking mechanism that grips the plug ground blade
- Wire retention clips mounted to the PDU chassis hold the plug in the receptacle

The higher the current-carrying capability of a plug, receptacle, or cord, the greater the amount of wire conducting material, typically copper, required to prevent overheating the wire, which could lead to a fire. Note that the smaller the wire gauge number, the greater the diameter of the conductor.

The conductors are surrounded by insulating material and jacket, which may have special properties. For example, the jacketing may be designed to resist damage from exposure to oil. Typical insulating and jacket materials are PVC, rubber, and neoprene.

The number of wires in a cable can vary. Below are some typical data center configurations:

- Two: a hot wire and a neutral wire without a ground wire
- Three: a hot wire, a neutral wire, and a ground wire
- Four: three hot wires (L1, L2, and L3) and a ground wire
- Five: three hot wires (L1, L2, and L3), a neutral wire, and a ground wire

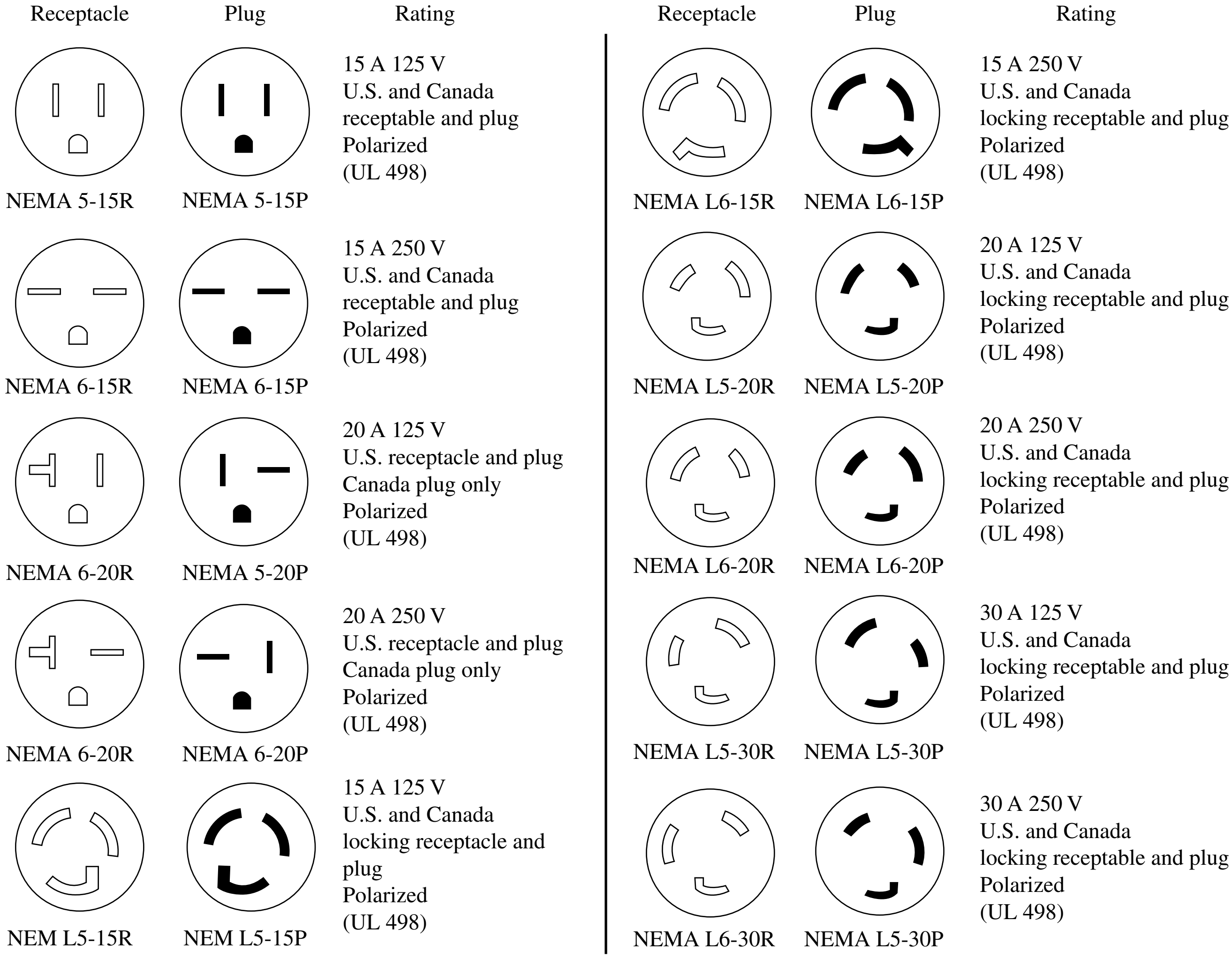


FIGURE 29.4 NEMA plugs and receptacles. Courtesy of Raritan, Inc.

29.2.4 Ratings and Safety

Rack PDUs, like all other electrical equipment, are subject to many general and specific safety standards. Furthermore, there are general industry terms and conventions that should be understood in order to ensure a reliable and safe data center. These are discussed in detail in the following.

29.2.4.1 Nameplate Data Nameplate data is the electrical power consumption information specified by the equipment manufacturer. It is typically a conservative estimate of the maximum amount of power the device could draw. This information is found on a label near the electrical power input to the device. More discussion of the use of nameplate data will follow.

29.2.4.2 Power Rating versus Load Capacity There can be confusion about power capacities and load capacity. This stems from misunderstanding approval agency regulations and from some manufacturers who may use misleading terminology. In North America, typical circuits have a maximum current-carrying capability and use circuit breakers or fuses rated at 15A, 20A, 30 A, etc. In other words, a 20 A fuse will blow or a 20 A circuit breaker will trip if a 20 A circuit experiences more than 20 A for some period of time. The period depends on the magnitude of the current and the type of fuse or circuit breaker protecting the circuit.

In North America, circuits are to be loaded to 80% of their maximum capacity. So, for example, a 15 A circuit should not carry more than 12 A, a 20 A circuit not more than 16 A, a 30 A circuit not more than 24 A, etc. The 80% value, for example, 16 A for a 20 A circuit, is often referred to as the derated value or the load capacity. In North America, a rack PDU vendor's specifications sheet may have a few current-carrying specifications. The specifications provided and the terminology used may vary by vendor, but the following are typical examples:

- Maximum line current per phase: 30 A
- Rated current: 24 A (30 A derated to 80%)
- Maximum current draw: 6 × 16 A

(Six circuits, each capable of carrying up to 16 A)

In Europe and elsewhere, circuits are simply described at their rated capacity, for example, 16 and 32 A.

Apparent power is specified in VA, which is volts × amps. Load capacity is specified in VA, where amps are the rated current, that is, the derated value. For example, for a single-phase rack PDU with 208 V and rated (not maximum) current of 24 A, the load capacity is 5.0 kVA (208 V × 24 A).

29.2.4.3 Approval Agencies In order to meet applicable local and NEC, rack PDUs must be safe and not emit electromagnetic radiation. Standards exist and recognized approval agencies are contracted by manufacturers to test products. A product that passes agency testing receives an approval listing number, and the manufacturer can then affix the agency approval listing logo on each product. The listing logo is your assurance that the product meets applicable safety and electric codes (Table 29.3). The manufacturer is required, upon request, to provide you the listing number and a copy of the testing report. You can also submit the listing number to the approval agency to verify compliance.

29.2.4.4 Proper Grounding The NEC (NEC Article 645.15) requires all exposed non-current-carrying metal parts of an IT system to be grounded. This means all equipment within a rack and the metal rack itself must be grounded.

The inlet plug of a PDU contains a ground pin. When this plug is connected to a properly wired receptacle, the PDU becomes the grounding point for the equipment plugged into it. The PDU can also be used to ground the metal rack—and most PDUs contain a special threaded hole for this purpose. Typically, a grounding wire is connected to the rack and the PDU using screws. Care should be taken to insure paint on the rack is scraped off where the grounding wire is attached to insure proper electrical conduction. There are special grounding screws with teeth under the head to ensure a good ground.

29.2.5 Overload Protection

The standard UL 60950-1 applies to the safety of ITE and requires the use of branch circuit overcurrent protection for ITE PDUs greater than 20 A. Typically, ITE PDUs greater than 20 A and certified after April 2003 must have built-in UL 489 circuit breakers or fuses (e.g., UL 248-5 fuses) suitable for branch circuit protection.

UL 60950-1 permits products at a maximum current of 15 and 20 A without circuit breakers or fuses, since the 15 or 20 A circuit breakers in the building are considered sufficient to protect the PDU; however, supplementary protection in the PDU provides additional protection. UL also "grandfathers" PDUs at more than 20 A that were certified prior to April 2003. Although such PDUs are still being sold, their use should be avoided if they are to be incorporated in larger ITE systems designed to the latest UL 60950-1 standard.

Newly certified ITE PDUs at more than 20 A are required to use overcurrent protection that meets branch circuit protection requirements in accordance with the National Electrical Code, ANSI/NFPA 70. In effect, this means these products are required to have circuit breakers listed under UL 489, "Standard for Molded-Case Circuit Breakers, Molded-Case Switches and Circuit Breaker Enclosures," or fuses, such as those listed to UL 248-5, "Low-Voltage Fuses—Part 5: Class G Fuses."

In addition to standard UL 489, UL also publishes the standard UL 1077, "Standard for Supplementary Protectors for Use in Electrical Equipment." Devices certified to this standard are called "Supplementary Protectors" and are called "Recognized" components, not "Listed" devices, as are UL 489 breakers. UL Listed Circuit Breakers meet more stringent requirements for branch circuit protection than Supplementary Protectors with UL Recognition.

TABLE 29.3 Safety and electromagnetic approval agencies

Approval	Description	Standard/revision/year	Comment
UL	Safety	UL 60950-1	Required in the United States
cUL/CSA	Safety	CAN/CSA-C22.2 No.	Required in Canada
CB	Safety	IEC 60950-1	Common replacement for UL, CSA, and CE in countries that accept CB
CE	Electromagnetic Interface (EMC)	EN 5502:2006	Europe
CE	Safety	EN 60950-1	Europe
FCC-A or FCC-B	EMC	FCC 47 CFR Part 15	United States
ICES-003	EMC	ICES-0003 issue-004	Canada

Source: Courtesy of Raritan, Inc.

Circuit breakers are used in a variety of ways. They are mounted in panel boards (also referred to as building PDUs) and rack PDUs to protect branch circuit wiring. They are also built into equipment to protect components and systems. Interrupting a short circuit—current flow limited only by the resistance of wiring—is a severe test of a circuit breaker. If the interrupting capacity of the breaker is not adequate, the device can literally explode.

UL 489 requires the breaker to be functional after being subjected to a short-circuit test. UL 1077 and the IEC standard EN 60934 allow for breakers to clear a short-circuit condition but become safely destroyed in the process. UL 489 breakers can interrupt short circuits of 5000 A or more. Typically, UL 1077 breakers can interrupt fault currents of 1000 A.

Overloads can be short term or long term. The protective device must not trip with a momentary overcurrent event that is normal for the piece of equipment being protected. Servers, for example, may create inrush currents as their internal power supply and filter circuits start. These inrush currents typically last only a fraction of a second and seldom cause a problem. If an overload lasts longer than a few minutes, the breaker should open to prevent overheating and damage. What gives a breaker the ability to discriminate between normal and damaging overcurrents is its delay curve.

29.3 ELEMENTS OF THE SYSTEM

Rack PDUs are the final endpoint of power supplied to ITE from incoming building feeds through a chain of equipment including UPS, transformers, and larger PDUs and circuit panels. IT and facilities management are increasingly viewing their rack PDUs not merely as a collection of power outlets for ITE but as a network of critical devices that significantly impact the overall efficiency and effectiveness of the data center. As such, they need to be properly managed like the ITE they power. This is driving the trend for use of more intelligent PDUs in data centers with environmental sensors and even integration with higher-level data center management systems. This section describes not only the components of the physical rack PDU and basic environmental sensors but also the rack PDU management system that leverages the intelligence in PDUs for operational improvements and energy use reduction. Further, this system can interface with and become part of a larger ecosystem of enterprise IT and facilities management systems.

29.3.1 Rack PDU

Over the past few years, average power consumption per server has rapidly increased with the adoption of high-power computing equipment like blade servers or data center containers. In addition, ongoing deployment of densely packed storage, virtualization, and cloud computing results in data centers with greater watts per square foot requirements from more densely packed racks such as a rack filled with 1U servers. To support new, power-hungry ITE data center managers have to deliver more power to the ITE rack. Over the last decade, the typical power required at a rack has increased from 2 to 12 kW and continues upward.

29.3.1.1 Single-Phase or Three-Phase Input Power for Rack PDU To accommodate the increased power demands at ITE racks, data center managers deploy rack PDUs capable of supplying multiple circuits, higher voltages, and higher currents. One way to increase the power at the rack is to increase the number of circuits and the voltage coming to the rack.

The amount of power available for use is referred to as apparent power and is calculated as volts × amps and is described as VA. A 120 V, 20 A circuit has an apparent power of 2400 VA or 2.4 kVA. A 208 V, 20 A circuit has an apparent power of 4160 VA or 4.2 kVA. Thus, one 208 V circuit provides almost twice as much power as one 120 V circuit assuming the current (amperage) remains the same. With three 208 V circuits, a substantial amount of power can be deployed in one three-phase PDU.

The cable to provide power to a three-phase PDU is thick and heavy but not as thick and heavy as the multiple, individual cables required to provide the same amount of power using either single-phase 120 V or single-phase 208 V. Running a single three-phase power cable to each three-phase rack PDU reduces both the number of cables, making installations easier, and the physical bulk of the cables, so less space is filled with cables blocking necessary cooling airflow under raised floors and within racks.

In cases where power needs to be provided at 120 V for devices such as routers, hubs, and switches, as well as at 208 V for demanding servers, three-phase PDUs can provide outlets with both 120 V (one of the three lines and a neutral) and 208 V (two of the three lines). Three-phase power at the rack is a convenient way to efficiently deploy both greater power capacity and flexibility.

29.3.1.2 Form Factor Rack PDUs are available in heights of one rack unit (1U, 1.75 in.) or two rack units (2U, 3.5 in.) for horizontal mounting in a 19 in. equipment rack.

Zero U rack PDUs mount vertically, typically to the vertical rails at the back of the rack. This can offer advantages. Zero U PDUs don't consume any rack unit spaces, and since the receptacles on the Zero U PDU line up better with the power cords for each IT device in the rack, they allow for the use of shorter power cords. This results in neater cable arrangements contributing to better airflow within the rack, which can improve cooling efficiency. Depending on the rack cabinets, Zero U rack PDUs can be mounted with screws or hung into the cabinet via buttons that are spaced 12.25 in. apart.

High-power rack PDUs are commonly equipped with circuit breakers for branch circuit protection. These circuit breakers may cause the rack PDUs to extend deeper into the racks. Consider how these PDUs are mounted in the rack, whether outlets facing center or back, to allow for cable management, airflow, and easy accessibility and serviceability of the ITE.

29.3.1.3 Outlet Density and Types Rack PDUs vary in the number of outlets supported based on the physical size (length, width, and depth) and thus the total space available for mounting outlets and internal components and the power-handling capacity of the PDU. For example, a 1U rack-mounted PDU may have enough space for eight 120 V/15 A NEMA 5-15R outlets, whereas, a 2U rack-mounted PDU may have enough space for 20 NEMA 5-15R outlets. On the other hand, a Zero U PDU may have 24 IEC C-13 230 V/10 A outlets or just four 250 V/30 A NEMA L15-30R outlets to support blade servers.

In the case of a large number of devices, each demanding a moderate amount of power, a large number of moderate power outlets are required. A typical dense "pizza box" deployment would include two rack PDUs for redundant power where each PDU is loaded to 40% so that if one power feed fails, the other feed will not exceed the NEC requirement of 80% (for North America). Typical outlets for "pizza box" servers are IEC C-13 (up to 250 V, 16 A) and NEMA 5-20R (up to 125 V, 20 A, 16 A rated).

In the case of high power consumption at a rack for a few devices, each of which consumes a lot of power, such as blade servers, storage, or network devices, the total amount of power required might be comparable to the high outlet density example given earlier, but the number and type of outlets may be different. Density for devices such as blade servers depends on their number of power supplies (often between two and six for redundancy), how the power supplies are configured (power supplies are most efficient when they operate close to their maximum level) and how many devices will be deployed.

In the case of a few devices demanding a lot of power, a large number of outlets may not be needed but outlets capable of delivering substantial power will be required. Typical outlets for high-demand devices such as blade servers at 208 V or 230 V are IEC C-13 (16 A) or C-19 (32 A) or, less commonly, NEMA L6-20R (20 A, 16 A rated) or L6-30R (30 A, 24 A rated) locking outlets.

29.3.1.4 Connectors: Ethernet, Serial, Sensor, USB, and Other Today, only the very basic rack PDUs have no external connectors—an input plug and outlets, much like a common power strip. Most rack PDUs now include a variety of connectors based on application requirements. Below, we describe four rack PDU connector configurations and general applications:

1. No connectors for external management or remote alarms and may not even have a local display. Not suitable for most data center applications today.
2. Local buttons allow navigation to see basic unit and outlet data.
3. A serial RS232 connector for local metering; local meter may be an LCD or LED. Can be plugged into a terminal or console server for Telnet or SSH remote access. Access via a menu or CLI using terminal emulation. Local buttons allow navigation to see basic unit data. No SNMP support available for alarms, unless via a specially developed serial console server. Typically nonswitched.
4. Ethernet (RJ-45) and RS232 serial (DB-9M) connectors for remote metering for the PDUs, circuit breakers, and outlets. USB-A (host) and USB-B (device) connectors to support PDU-to-PDU cascading, webcams, and wireless networking. SNMP support available for alarms, Telnet or SSH access possible for command line access. Support for sensors—like temperature, humidity, airflow, air pressure, and others—may be available on the PDU or with an add-on external device. Remote metered models typically have an LCD or LED display as well with buttons for navigation to see basic unit and outlet data.

29.3.1.5 Branch Circuit Protection Since April 2003, UL requires branch circuit protection, circuit breakers, or fuses for PDUs where the inlet current is greater than the outlet current, for example, 30 A (24 A rated) plug and 20 A (16 A rated) outlets. 15 and 20 A (12 and 16 A rated) rack PDUs can be supplied without branch circuit breakers because circuit breakers in upstream panel boards are deemed to provide the necessary protection. Rack PDUs with breakers or fuses are like mini-subpanels. For example, a 208 V, 30 A (24 A rated) three-phase PDU has three circuits, and each circuit/set of outlets has a 20 A circuit breaker.

There are four types of circuit breakers: thermal, magnetic, thermal–magnetic, and hydraulic–magnetic. Thermal circuit breakers incorporate a heat-responsive bimetal strip. This technology has a slower characteristic curve that discriminates between safe, temporary surges and prolonged overloads.

Magnetic circuit breakers operate via a solenoid and trip nearly instantly as soon as the threshold current has been reached. This type of delay curve is not ideal for servers that typically have inrush currents anywhere from 30 to 200% above their normal current draw.

Thermal–magnetic circuit breakers combine the benefits of thermal and magnetic circuit breakers. These devices have a delay to avoid nuisance tripping caused by normal inrush current and a solenoid actuator for fast response at higher currents. Both thermal and thermal–magnetic circuit breakers are sensitive to ambient temperature.

A magnetic circuit breaker can be combined with a hydraulic delay to make it tolerant of current surges. Hydraulic–magnetic breakers have a two-step response curve. They provide a delay on normal overcurrents but trip quickly on short circuits and are not affected by ambient temperature.

Circuit breakers used in rack PDUs are typically thermal–magnetic or hydraulic–magnetic with delay curves that allow for reasonable inrush currents (servers typically have inrush currents 30–200% above their normal operating load) while protecting devices from excessive fault currents.

Fuses are also acceptable for PDU circuit protection. However, replacing a fuse can be time-consuming and may require an electrician leading to longer mean time to repair (MTTR). Spare fuses must be stocked in inventory and the correct fuse must be used to ensure reliability and protection.

The following are some points to consider when selecting a rack PDU:

- Compliance with the latest fuse and circuit breaker standards
- The acceptable MTTR for fuse replacement versus circuit breaker resetting
- Impact on uptime service-level agreements (SLAs) if a fuse blows versus if a circuit breaker trips

29.3.1.6 Circuit Breakers: Single Pole versus Double and Triple Pole The reliability and flexibility of the branch circuit breaker configuration are important. Typically, circuit breakers are available as single-, double-, or triple-pole devices. Single-pole breakers are appropriate for circuits comprised of a hot wire and neutral, for example, 120 V at 20 A or 230 V at 16 A. Single-pole breakers provide a disconnect for the single hot wire used in circuits with a hot wire and neutral. Double-pole breakers provide a disconnect for circuits comprised of two hot wires, for example, 208 V at 20 A. Some PDU designs use double-pole (or triple-pole) breakers to provide protection for two different circuits, for example, two different hot wires. Since a single double-pole breaker is less expensive than two single-pole breakers, this type of design will lower the cost. Double-pole breakers will trip if either of the two circuits they protect is overloaded. It is less expensive than two (or three) single-pole breakers, but unless the poles can be operated independently, in a maintenance shutdown or trip, all two or three circuits are de-energized.

For example, assume a rack PDU with six branch circuits is protected by circuit breakers. Some rack PDUs in this configuration may protect the six circuits with three double-pole circuit breakers—one double-pole circuit breaker for the circuits with Line 1, one for the circuits with Line 2, and one for the circuits with Line 3. It is less expensive to use double-pole circuit breakers, but there are some drawbacks. Double-pole breakers will trip if either of the two circuits they protect is overloaded. This means double-pole breakers are less reliable. Double-pole breakers are also limiting because if you choose to shut off a circuit, for maintenance, for example, you have no choice but to shut off both circuits. Alternatively, some rack PDUs protect the six circuits with six single-pole circuit breakers—one breaker per circuit. This is more expensive but single-pole breakers are more reliable and less limiting. Look for rack PDUs that allow only one circuit to be de-energized for improved reliability and flexibility.

29.3.1.7 Circuit Breaker Metering Circuit breaker metering is a useful feature on any rack PDU, but it is particularly important when dealing with high power because the consequences of tripping a breaker can be disastrous if it means losing several blade servers. With circuit breaker metering, the end user sets a threshold. When that threshold is crossed, an alert is delivered so the end user knows power demand needs to be reduced or there is the risk of tripping a circuit breaker. Monitoring branch circuit breakers is important since high-power draw means a greater chance of tripping a breaker.

Line metering, intended for three-phase rack PDUs, is very useful for balancing the power drawn over each line. Overdrawing power from one line relative to another line wastes available power, and unbalanced lines can place excessive demands on the neutral in Wye-configured PDUs.

29.3.1.8 Cord Length, Feed, and Retention Rack PDU power cords vary in length depending on the whips (power cables from a building PDU) and the location of the racks. The rack PDU power cord must be long enough to reach its power source, which is typically a whip located under the raised floor or an outlet above the rack. A common power cord length is 10 ft (3 m), but other lengths can be specified, to a UL maximum of 4.5 m (15 ft).

Rack PDU power cords may exit the PDU itself from the rear, the front, the top, or the bottom. With the power cable exiting the bottom, a Zero U PDU, the data center manager will need to ensure sufficient space for the bend radius of the cable. In general, a bend radius of 5.25 in. (3U) will be sufficient, but this should be confirmed as bend radii will depend on the gauge (AWG) of the cord. A smaller bend radius may be acceptable for thin cables and a larger bend radius may be required for heavy-duty cables. The orientation of the PDU power cord may seem trivial, but it can be a potential problem depending on the physical rack and the location of the power source for the rack. Consider the orientation of the power cord and how it will be routed to connect to the whip. For example, does the power come up from the raised floor or down from cable trays above the racks and is there room inside the rack to route the cable so as not to block airflow?

Proper PDU cord retention practices, just like rack cable management, can make a big difference in operational efficiency and reliability. Taking steps to support, organize, and secure the many power cords using cord retention clips will dramatically improve your ability to access and manage the equipment connected to PDUs inside the rack. This will also minimize the chance of inadvertently unplugging power cords from rack PDUs. Finally, you should neatly arrange and secure the power cords between ITE and the rack PDU to allow for maximum airflow.

29.3.1.9 Local Display and User Interface Virtually all rack PDUs designed for data center use have built-in displays, typically LEDs, to show current draw for the entire PDU unit. Local displays have limited functionality compared to the information and control available from a remote interface, but they can be convenient and useful when working at the rack itself. The local display might allow an IT admin to toggle between current draw and voltage or, for those rack PDUs that monitor individual outlets, to sequence through the outlets to determine the current being drawn by each device. Some switched and intelligent models will have LED indicators next to each outlet to display status, whether it is on/off, booting, firmware upgrade, or fault.

In addition to a local display on the rack PDU itself, some PDUs offer a serial interface for local terminal connectivity via a laptop for configuration, diagnostics, or connectivity to a serial console server that concentrates multiple connections.

29.3.1.10 Remote User Interface For a remotely accessible rack PDU (all but the basic PDU or PDU with metering and only local display), there are typically two choices for a remote user interface to the rack PDU over an IP network. The most common is a Web-based graphical user interface (GUI) (Fig. 29.5) to an Ethernet-enabled PDU. Some PDUs support SSL-encrypted access (using https), while others support only unencrypted access (using http). Check your organization's security requirements when selecting a PDU.

The PDU can also be accessed via Ethernet over IP using SSH (encrypted) or Telnet (unencrypted) with a CLI. Security considerations should be kept in mind before enabling/disabling Telnet access. Some PDU manufacturers provide a serial console server that connects to the PDU locally via serial (RS232) and allows access to the unit remotely using SNMP or CLI.

One factor to consider is integration with central directory services for user authentication and access control. This becomes especially important when the rack PDU offers the ability to remotely turn on/off/recycle individual outlets or groups of outlets. Finally, remote access to the PDU does not eliminate the need for some local access to the PDU with an LED/LCD and associated buttons.

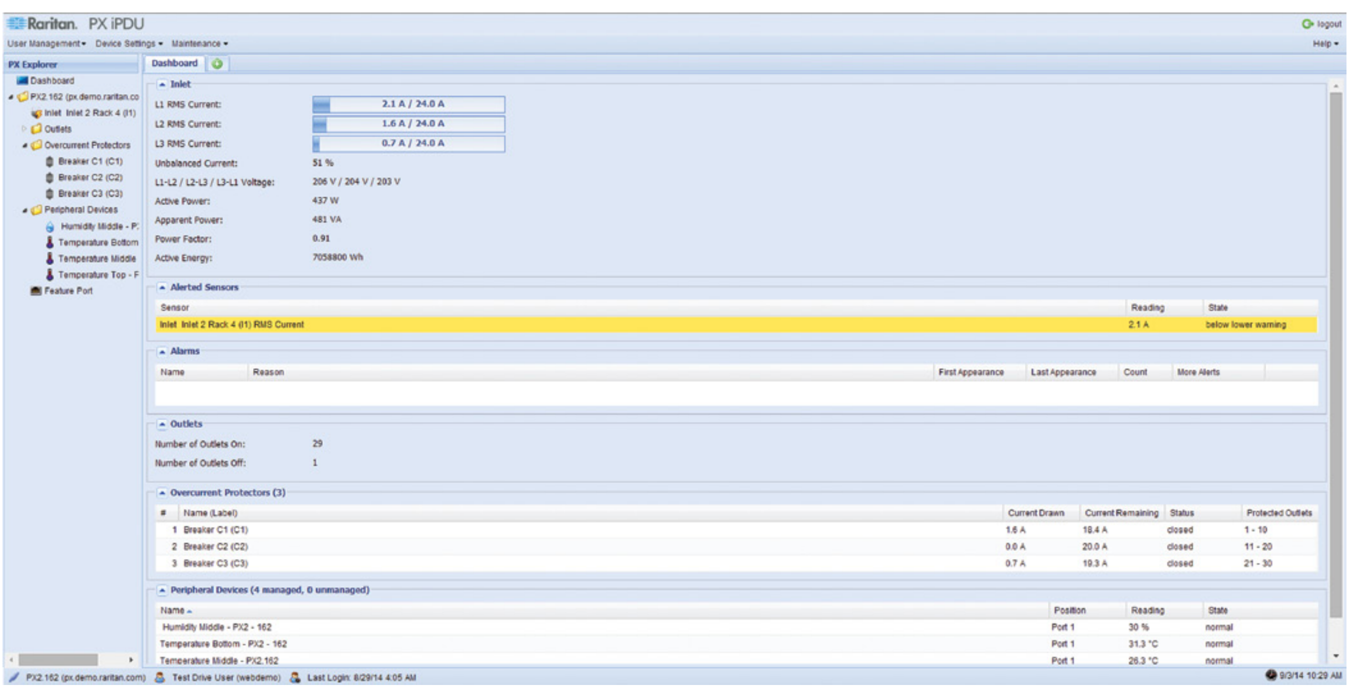


FIGURE 29.5 Rack PDU management system graphical interface. Courtesy of Raritan.

29.3.1.11 In-line Meters For data centers with existing PDUs that lack any metering capability, adding in-line meters can be useful to monitor power consumption per line or per circuit. In-line meters are usually one line or circuit in and one line or circuit out. Some vendors offer models that support up to four "In" and four "out." There are basic in-line meters that provide a simple current metering, with or without IP connectivity. Others are more sophisticated and provide richer data like active power, apparent power, and kilowatt-hour metering; and some might even include integrated environmental monitoring. By upgrading older, basic PDUs with in-line meters with IP connectivity, they can be managed along with metered and intelligent rack PDUs by a rack PDU management system so that you can have a comprehensive view of the health and usage of power for day-to-day operations and planning.

29.3.2 Environmental Management

With the IT industry's increasing focus on improving data center efficiency, more rack PDU manufacturers are offering environmental sensors. These include sensors to measure rack air temperature at the server inlets, humidity, airflow, vibration, smoke, water, and air pressure. Some PDUs will have preinstalled sensors; others provide for optional, plug-in external sensors. Another common approach is to deploy a completely independent rack management system, choosing from a wide range of environmental sensors; however, this has the disadvantage of consuming additional rack space for the rack management system as well as the cost of a separate infrastructure—for example, IP addresses, Ethernet ports, and cabling. Connectivity for sensors is typically either via RS485 or 1-Wire®.

29.3.2.1 Temperature Sensors Temperature sensors monitor the air inlet temperatures at IT devices such as servers. (See the ASHRAE sensor placement diagram in the following.) Since ITE generates considerable heat, manufacturers specify a range of acceptable temperatures for proper operation. A sensor-capable PDU should allow thresholds to be set for sending automatic alerts when the inlet temperature approaches the vendor-specified maximum to prevent servers from shutting down or failing due to overheating. In addition, it is also a good practice to set a minimum to provide alerts when the inlet temperature is colder than necessary. From a data center plant perspective, the cost of cooling and moving air is the largest infrastructure expense, so maintaining IT inlet air temperatures colder than necessary merely wastes energy and money. Temperature sensors at the rack also provide early warning about temperature extremes, hot spots, or cold spots and can help identify when an HVAC system is becoming unbalanced. To ensure that ITE is getting enough cool air, the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) has recommended that temperature probes be placed at specific locations at the inlets of equipment racks.

29.3.2.2 Humidity Sensors Understanding the basics of what humidity is and how it affects your server room can impact how long your computer equipment lasts and how much your electricity bill costs. Humidity is a measurement of moisture in the air. High humidity can cause condensation buildup on computer components, increasing risks of shorts. Likewise, if the humidity is too low, data centers can experience electrostatic discharge (ESD). Humidity can be monitored per area or zone to ensure that it is in the safe range. ASHRAE has recommended ranges for the data center that should be consulted. Appropriate thresholds and alarms should be set to indicate a potential problem.

29.3.2.3 Airflow Sensors Airflow sensors will detect a reduction of air movement that might create the potential for overheating, which can destroy ITE. There are two primary areas for monitoring airflow in the data center—above the floor (monitored at a number of points) and below the floor (monitored at select points). Differential airflow sensors are used to ensure that the pressure differential between the subfloor and the floor is sufficient to control air flowing from the subfloor to the floor above. Blockages in underfloor supply plenums can cause high pressure drops and uneven flow, resulting in cold spots in areas where cooling air is short-circuiting to the return path. Airflow sensors should have thresholds set, and alarms enabled, like other environmental sensors, to ensure that data center managers are alerted when conditions are less than optimal for efficient cooling.

29.3.2.4 Air Pressure Sensors It is important to have the appropriate air pressure in underfloor supply plenums, but sometimes, this is treated as an afterthought. Air pressure that is too high will result in both higher fan costs and greater leakage, which can short-circuit cooling air, while pressure that is too low can result in hot spots at areas distant from the cool air supply point. This can lead to poor efficiency "fixes" to correct the problem such as lowering the supply air temperature or overcooling the full space just to address a few hot spots. Differential air pressure sensors can be used to ensure that the pressure differential between the subfloor and the floor is sufficient. Maintaining appropriate room pressure prevents airborne particulates from entering the data center.

29.3.2.5 Contact Closure Sensors Contact closure sensors can be used for a variety of applications. For example, a contact closure could send an alert when a cabinet door is opened and trigger a webcam to take a picture. Contact closure sensors can be connected to any device that can open or close a contact.

29.3.2.6 Other Sensors There are a variety of other sensors that can be used in the data center. Examples include in-cabinet smoke, water, and vibration sensors. Like the other sensors mentioned earlier, these are used to send alarms when measured conditions are outside the range for proper data center operation.

29.3.3 System Connectivity

29.3.3.1 Physical Topology Like many functions of data center management, the best practices for remote management of rack PDUs are evolving. The current best practice is to connect all remotely accessible rack PDUs to the "management network" (separate from the "production network") directly in order to collect periodic meter readings, get immediate notifications of any faults or potential problems, and enable remote power cycling of ITE (depending on the intelligence of the rack PDU). When planning for a new facility, provide for a minimum of two Ethernet drops for each cabinet to support rack PDUs, since each rack will typically require two PDUs.

29.3.3.2 Communication Protocols The communication protocols used are typically TCP/IP when PDUs are Ethernet connected, and proprietary protocols for PDUs serially connected to a console server, which, in turn, connects to the TCP/IP network via Ethernet. Most often, SNMP protocol is used for management, while LDAP and Active Directory are used for authentication, authorization, and access control. SSH and Telnet may be used for command line management and HTTP/HTTPS for Web-based access.

There are rack PDUs now with USB-A (host) and USB-B (device) ports that can be used to support USB devices such as webcams and WiFi modems. Some rack PDUs support MODBUS, a common, older building management communication protocol, and some rack PDUs support the GSM modem protocol so that cell phones can receive one-way text alerts.

29.3.3.3 Managing the Rack PDU The management system for data center power is often run on a "management network" separate from the production network. This reduces the likelihood of a Denial of Service (DOS) or other attack that would affect this critical function. In mission-critical facilities, there are often two connections to each rack PDU equipped with remote communications: one for syslog, SNMP traps, access via Web browser, and kilowatt-hour logging and another for critical functions like remote power cycling, status of circuit breakers, and load monitoring. In some cases, administrative functions, like rack PDU configuration, are performed via command line scripting through a secondary interface such as a serial port, while Ethernet remains the primary interface for all other functions.

Some important management functions are listed below:

1. Audit logging to track all activity—like switching of outlets and configuration changes. Two or more syslog servers are often used for this function.
2. Fault management—via SNMP with tools like HP OpenView, IBM Tivoli, and others. SNMP V2 is still the most commonly used, but SNMP V3, with its built-in security, is recommended for applications requiring outlet control.
3. Configuration—via Web browser, SNMP, command line, or a central software tool.
4. Firmware upgrades—not an issue for older PDUs with minimal functionality, but something that may be required for Ethernet-enabled PDUs. A central tool is essential to manage potentially large numbers of PDUs to simplify management and reduce cost of ownership.
5. Alerts—via SMTP messages.

A combination of some or all of the aforementioned capabilities is required to effectively manage a data center. Check your application requirements and choose the PDU type appropriate for your application. If, however, you have multiple rack PDUs (40+), you will want to consider a comprehensive rack management system, discussed in Section 29.3.4.

29.3.4 Rack PDU Management System

A rack PDU management system is a software application (sometimes delivered as a software appliance) that consolidates communication with all your rack PDUs and in-line meters equipped for remote communication (Fig. 29.6). Its main functions are data collection, reporting, power control, element management, and fault management. The system collects and converts detailed power data into useful information and provides a central point for secure access and control across multiple rack PDUs with operation validation and an audit log. It simplifies the management of rack PDUs and alerts you to potential incidents. We include it in this section because for larger data centers (with more than 40 racks), it is a "must-have" to realize many of the benefits offered by metered, switched, or intelligent PDUs—improved energy efficiency, increased uptime, and lower operational costs.

29.3.4.1 Data Collection Data collection is the fundamental component that enables reporting and most other management functions. The management system can collect only the data elements provided by the managed PDUs. As discussed earlier, basic PDUs provide no data, metered PDUs provide total unit data, and intelligent

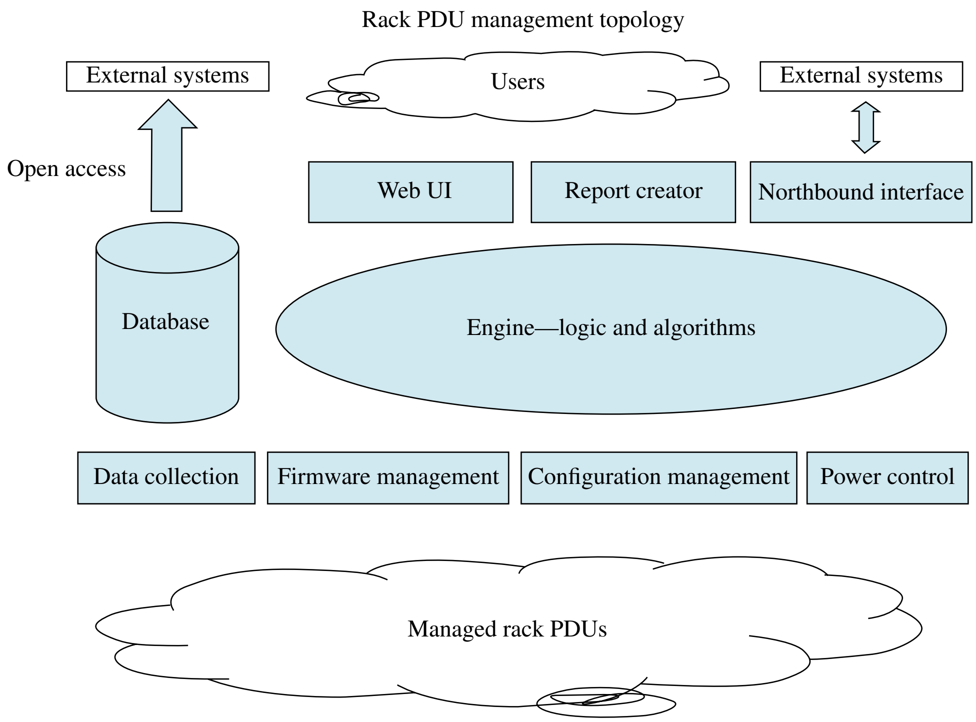


FIGURE 29.6 Rack PDU management topology. Courtesy of Raritan.

PDUs provide individual outlet data and more, so it is important to understand what data you will want to analyze when selecting rack PDUs. Typical data elements you will probably want to collect, as available from the PDUs, include total unit active and apparent power, line current and capacity, outlet-level current and active power, environmental sensor data, and real-time kilowatt-hour metering data.

Next, you will want to determine the granularity of the data. Your management system should offer a user-configurable data polling interval. For most applications, a normal polling interval is 5 min, which means the system will collect data points every 5 min, but if greater granularity is required, the rack PDU may need to store data readings so that the network is not overloaded with polling traffic. Finally, you will want to use a roll-up algorithm to collect data for long periods without causing the database to balloon and affect performance. For most energy management applications, data readings are rolled up to a maximum, average, and minimum—hourly, daily, and monthly.

Advanced polling options enable a customer to minimize network traffic while still enabling granular data collect. Advanced polling requires a rack PDU that has the memory capacity to record and store readings called samples. For example, a rack PDU might be able to store 120 samples. The rack PDU management system should offer the ability to configure optional sample rates for each rack PDU and also set optional polling intervals for the management system itself to collect the stored samples at each rack PDU not previously collected. For example, a rack PDU can be configured to record and store 1 min samples. The rack PDU management system can be configured to poll the rack PDU once an hour. In each poll, it will pull the 60 min samples since the last poll with the intelligence to know the last reading is recorded on the previous poll.

29.3.4.2 Reporting and Analytics for Power Monitoring and Measurement Reporting and graphing should include active power, current, temperature, humidity, and information derived from the basic collected data such as energy usage, cost, and carbon emitted due to the energy consumed for standard and selected time periods.

Reports on maximums and minimums for current, temperature, humidity, and active power simplify key tasks and ensure that you are not in danger of exceeding circuit breaker ratings, overcooling, or undercooling. This environmental information can give data center operators the confidence to raise temperature set points without introducing risk to the ITE. Analyzing trend line graphs and reports and "what-if" modeling can help you do capacity planning based on real-world data.

Outlet-level data and reporting granularity can help you become more energy efficient. It enables you to determine

FIGURE 29.7 Data center infrastructure management (DCIM) monitoring software. Courtesy of Raritan.

the potential savings of upgrading to more energy-efficient servers or the benefits of server virtualization. Consolidating several low-utilization physical servers as virtual servers on one high-utilization physical server can reduce overall expenses, but you will need to understand the resulting power demand of the host servers. You can also establish objectives, report on usage, and implement changes for both physical components of the data center (floor, room, row, rack, and IT device) and also logical groupings (customer, department, application, organization, and device type). This level of detail creates visibility and accountability for energy usage, and some IT organizations issue energy billback reports to users/owners of the ITE.

29.3.4.3 GUI The GUI (Fig. 29.7) is your window into all of the rack PDU management system functions. This should be clean, intuitive, and Web-based, functioning with all major Web browsers. A Web-based system provides you more remote access options and is easier to support and upgrade. The GUI will most likely include a user-configurable dashboard. The dashboard can be displayed in the data center network operations center for an easy at-a-glance view of the status of the data center power and environmental conditions. This will give your customers, either internal or external, a good indication of your data center management capabilities.

29.3.4.4 Element Management The main components of element management include centralized rack PDU access and control, firmware management, and bulk configuration. You can view all your managed rack PDUs from one Web browser window and get a summary view of the name, location, status, manufacturer model, and firmware level. You will want to be able to drill down to manage at the PDU unit level, line level, and, in the case of intelligent PDUs, outlet level. Finally, one-click, sign-on access to each managed PDU can give you control through the PDU's own GUI.

Since Intelligent PDUs run firmware with many configuration options, the rack PDU management system should allow you to centrally store rack PDU firmware/configuration versions and facilitate distribution to multiple PDUs. Configuration template storage and distribution will simplify initial PDU installation as well as future unit additions and replacements.

29.3.4.5 Fault Management Rack PDU management systems often provide a map view and a floor layout view and use a color scheme to provide an at-a-glance view of the health of all managed PDUs. Health problems are discovered in a several ways. The system can receive an SNMP trap or a syslog event so that you become aware of the problem as it happens. Also, a management system can poll the rack PDU at set intervals to collect the heath status of the communication path to the PDU or critical failures and forward events to a higher-level enterprise management system.

29.3.4.6 Local and Remote Control/Switching Switched rack PDUs allow for outlet control including on/off power cycling. However, most IT devices have more than one power supply for power redundancy purposes, and these supplies are connected to outlets on separate rack PDUs. Through the management system, you can power cycle at an IT device level, which will programmatically switch outlets from multiple PDUs. The system should allow for grouping IT devices into racks such that you can control a full rack. Finally, any switched PDU must allow for flexible sequencing and delay so an inrush current spike does not trip a circuit breaker and so that application intrasystem dependencies are taken into account during start-up and shutdown.

29.3.4.7 Security of Data and User Access Remote monitoring, metering, and management require secure remote access via Ethernet and/or serial connections. To ensure security, an intelligent rack PDU should have strong encryption and passwords and advanced authorization options including permissions, LDAP/S, and Active Directory. A Web session timeout will protect against leaving an authenticated session live while not in use.

29.3.4.8 Administration and Maintenance Most of the administration is initial setup. All systems will allow for GUI entry of this data but that can be time-consuming. Systems should also allow for the import of configuration information, for example, via CSV files. During the setup, you will add your rack PDUs and hierarchical and logical relationships. Hierarchical relationships include data center, floors, rooms, rows, racks, rack PDUs, and IT devices. Logical associations include owners/customers of the IT device and IT device type. The administrator will also set the data pruning intervals to ensure unnecessary data is pruned from the system.

29.3.4.9 Open Point of Integration Most data centers have some other management systems already in use, so it is important that the rack PDU management system can be integrated into these systems to minimize the amount of duplicate data entry and collection. Asset management and enterprise reporting systems are two typical systems that should logically interface with the rack PDU management system. The asset management system will automatically add rack PDUs, IT devices, and their associated connections to the rack PDU management system's inventory of managed devices. Integration with an enterprise reporting system enables the creation of custom reports with the additional ability to correlate data that exist in other systems. Finally, in recent years, a more comprehensive class of products called the Data Center Information Management (DCIM) have been introduced that normally include the aforementioned functions, overall capacity planning tools, and more.

29.4 CONSIDERATIONS FOR PLANNING AND SELECTING RACK PDUS

The following paragraphs in this section will address some of the basic considerations and options you will have when designing and deploying your rack PDU system.

29.4.1 Power Available and Distributed to Racks

There are several approaches to deploying power to ITE racks that affect rack PDU selection and configuration. Some approaches provide degrees of redundancy and hence higher reliability/availability than others but may not be appropriate for certain types of equipment. Redundancy and higher availability require resources, so managers of data centers that have limited power resources need to decide what ITE justifies redundant power, for example, production servers, and what equipment does not, for example, nonproduction equipment being tested or evaluated.

29.4.1.1 Single Feed to Single Rack PDU The simplest power deployment to an ITE rack is a single appropriately sized power feed to a single rack PDU (Fig. 29.8). ITE with one or more power supplies would plug into this single rack PDU. If that single feed or single rack PDU should fail, for whatever reason, power to the equipment in the rack will be lost. The failure could occur at the rack PDU itself or farther upstream, perhaps a main feed fails or a building PDU circuit breaker trips.

As noted earlier, the NEC requires that circuits be loaded to no more than 80% of their maximum capacity. For example, if a 30 A feed and rack PDU were deployed in this configuration the load allowed (the rated current) would be 24 A (30 A × 80%). The NEC would expect the feed and PDU to handle a maximum of 30 A, but the circuit should be loaded to only 24 A.

29.4.1.2 Dual Feed to Single Rack PDU with Transfer Switch The next step up in availability is still a single feed to a single rack PDU with the addition of a Transfer Switch, which typically has two feeds from the same or different building feeds (Fig. 29.9). If a feed to the transfer switch fails, it automatically switches to the other power feed and the rack PDU continues to power the ITE. However, if the single rack PDU fails, the power to the ITE is lost.

There are two types of transfer switches: static transfer switch (STS) and automatic transfer switch (ATS). An STS is based on static electronic component technology (silicon-controlled rectifier), which results in faster and better controlled transfer between sources. An ATS is less expensive and is based on electromechanical relay technology, which results in slower transfer times.

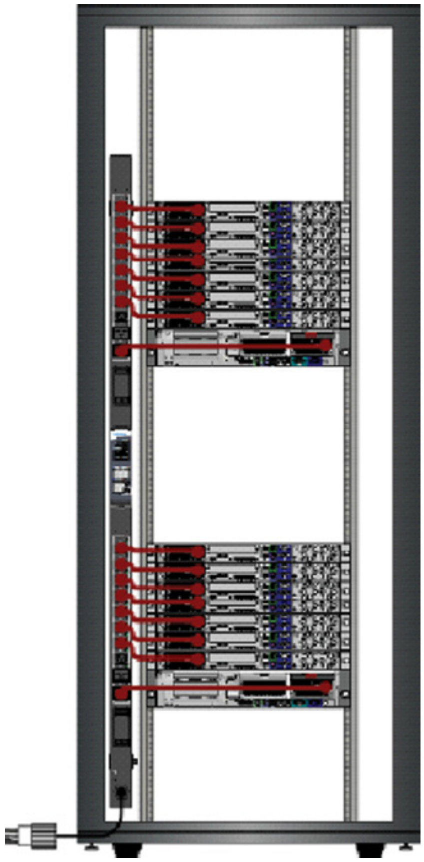


FIGURE 29.8 Single feed to single rack PDU. Courtesy of Raritan.

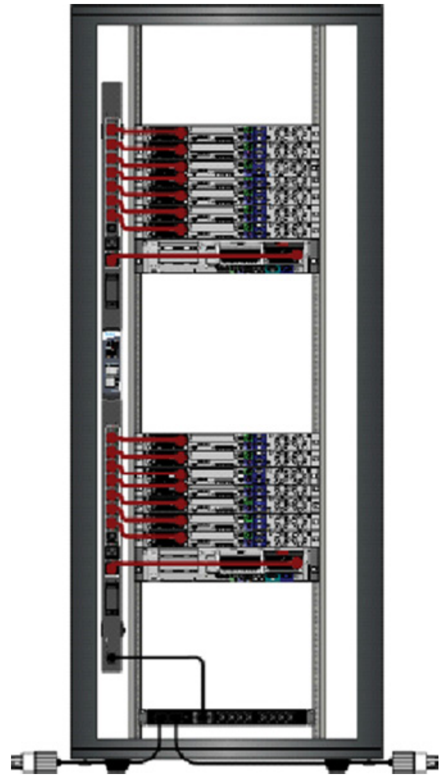


FIGURE 29.9 Dual feed to single rack PDU with transfer switch. Courtesy of Raritan.

Again, with this arrangement, the rack PDU is still loaded to 80% of the maximum, but the electrical power capacity required has doubled—one feed is operational and the second feed is a backup. It has also doubled the amount of upstream equipment necessary to supply the additional feed.

Two power feeds to an ATS and then to a single rack PDU are generally used only where reliability is a concern but the ITE itself, for example, a server, has only one power supply.

29.4.1.3 Dual Feed to Dual Rack PDUs Today, many servers, network devices, storage systems, even Keyboard, Video, Mouse (KVM) switches, and serial console servers are available with dual power supplies. Some larger servers may have as many as four or even six power supplies. The most reliable deployment here is to use two power feeds to two rack PDUs (Fig. 29.10). With this configuration, if one rack PDU or power feed fails, there is a second one available to maintain power to the ITE in the rack. A common practice when using dual feeds is to use rack PDUs with colored chassis such as red and blue. The colored chassis enables a visual control for installation of or changes to the PDU and connections. The rack will have a red chassis PDU fed by input circuit “A” and a blue chassis PDU fed by input circuit “B.” The colored chassis helps to eliminate confusion about which PDU is fed by circuit “A” or “B.”

But it is important to remember the requirement that each circuit be loaded to no more than 40%. If the two circuits feeding the rack are both loaded to 80%, the NEC requirement will be met, but what would happen if one of the circuits failed? The power demand to the second circuit would jump from 80 to 160%, and the circuit breaker for that feed would trip so the second circuit to the rack would also lose power. To prevent this, both feeds should be loaded to no more than 40% so that if one fails, the remaining circuit won’t be loaded to more than 80%. Compared to the previous case with ATS in Section 29.4.1.2, where one feed is backup, in this configuration, both feeds are powering ITE.

Note that if you intend to perform remote switching for ITE with dual power supplies, you will want to use a rack PDU that supports outlet grouping; that is, two or more outlets are controlled as though they are a single outlet.

29.4.1.4 Multiple Power Supplies ITE with two or more power supplies can vary in the way power is delivered to the equipment (Fig. 29.11). Some devices have a primary and backup power supply; some alternate between the power supplies; and some devices share power demand across all the power supplies. For example, a blade server

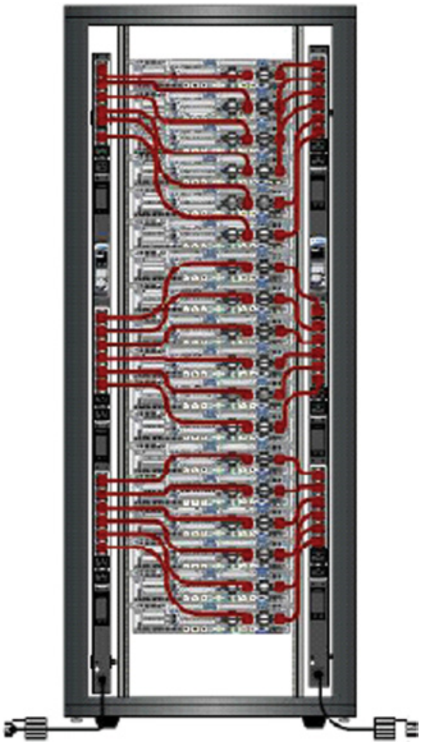


FIGURE 29.10 Dual feed to dual rack PDUs. Courtesy of Raritan.

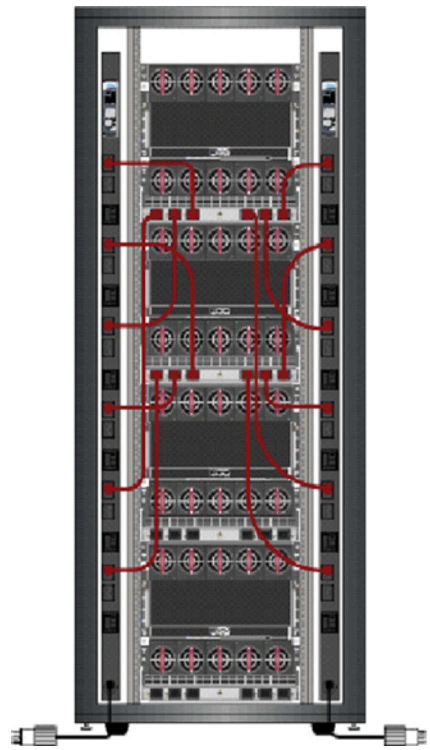


FIGURE 29.11 Multiple power supply configuration. Courtesy of Raritan.

with four power supplies in a 3 + 1 redundancy configuration would draw one-third of its power from each of its three primary power supplies, leaving one for redundancy in the event any one of the three fails. Finally, some more sophisticated devices have multiple power supplies that are designed for both redundancy and efficiency. For example, some devices might drive utilization rates higher on specific power supplies to drive higher efficiency. You will need to check with each equipment manufacturer to understand how the power supplies work so that optimal balanced load configurations can be achieved on the rack PDU, especially those with branch circuits and three-phase models.

29.4.1.5 Load Balancing Load balancing attempts to evenly distribute the rack equipment's current draw among the PDU's branch circuits so that as you come closer to perfect balance, more total current can be supplied with the greatest headroom in each branch circuit. For example, consider a PDU with two 20 A circuit breaker protected branch circuits—where each branch contains a number of outlets. The total current capacity of the PDU is 40 A with the limitation that no branch circuit of outlets can exceed 20 A. If the total load of all devices plugged into the PDU is 30 A, perfect balance is when the load is exactly divided between the two branches (15 A each branch). The headroom in each branch is then 5 A (20 A circuit breaker less 15 A load). Any other distribution of the load (16 A:14 A, 17 A:13 A) results in less headroom.

Load balancing has similar benefits for three-phase PDUs. As the load comes closer to perfect balance, the current draw is more evenly distributed among the three-phase lines (more headroom), and total current flowing in the three lines is minimized. For example, consider a 24 A three-phase Delta-wired PDU with three branch circuits. When an 18 A load is balanced across the three branch circuits (6 A load in each branch), the current flowing in each input phase line is 10.4 A, and the total current in all three lines is 31.2 A. If the entire load was carried by one branch circuit (totally unbalanced), the current in the three-phase lines are 18 A, 18 A, and 0 A, respectively, and the total current is 36 A. When the load is balanced across all three lines, the PDU has 7.6 A (18.0–10.4 A) more headroom.

Load balancing can be tricky because many IT devices draw power in varying amounts depending on the computational load. For devices with single power supplies, an estimate of the power consumption should be made for each device and then the devices plugged into the several circuits so that the circuits are loaded evenly. This is true both within a rack and across multiple racks. For devices with dual power supplies, they should be plugged into different circuits. A typical deployment would be the dual feed to dual rack PDUs mentioned earlier.

For IT devices with more than two power supplies, such as blade servers, load balancing can become even more complicated, especially if the rack PDUs are three-phase models.

As an example, assume four blade chassis are to be installed in a rack, each chassis has six power supplies, and two three-phase rack PDUs will be installed in the rack for redundant power. The first blade server will have power supplies (PS) #1, #2, and #3 plugged into circuits (C) #1, #2, and #3, respectively, on PDU A and power supplies (PS) #4, #5, and #6 plugged into circuits (C) #1, #2, and #3, respectively, on PDU B. Since we want to try to balance the load across all circuits and lines and we can't be sure that each of the four blade servers will be performing tasks that equally load the circuits, we will stagger the second blade server power supplies. So the second server will have PS #1 plugged into C #2, PS #2 plugged into C #3, and PS #3 plugged into C #1 on PDU A and PS #4 plugged into C #2, PS #5 plugged into C #3, and PS #6 plugged into C #1 on PDU B. Circuit-level metering, phase-level metering, and outlet-level metering will be very helpful for (re)balancing loads in the rack.

29.4.1.6 Inrush Current Servers draw more current when they are first turned on, known as inrush current. As discussed in the section on overload protection, rack PDUs with circuit breakers are designed not to trip during very short periods of high currents. However, it is better for upstream circuits if sudden surges during equipment powering on are minimized. For this reason, some rack PDUs provide outlet sequencing and allow users to configure both the sequence and the delay time in which the outlets are turned on. Some rack PDUs may allow programming of outlet groups and allow sequencing of groups of outlets.

29.4.2 Power Requirements of Equipment at Rack

Section 29.4.1 deals with ways to deploy electrical power to a rack. This section deals with determining how much power to deploy to a rack. Typically, the starting point is an IT device's nameplate power requirement data (see Section 29.2.4.1) that specifies a voltage and current (amps), which is typically higher than what is usually seen during actual deployment. Often, a percentage of the nameplate value, for example, 70%, is used when computing the maximum PDU load capacity required: PDU load capacity = sum (or Greek letter sigma) (device nameplate in VA × 70%). For example, 208 V × 2.4 A × 70% × 14 servers = 4.9 kVA.

For the aforementioned example, if you run 208 V, you need a 30 A (5 kVA) rack PDU since you will load it to 80% to meet North American requirements (4.9 kVA/208 V = 23.5 A; 23.5A is approximately 80% of 30 A). If you want redundancy, add a second 5 kVA rack PDU and load both PDUs up to 40%. You will need to specify the appropriate number of outlets. It is a good idea to have a few spare outlets for other devices even if the rack PDU will be at its maximum capacity. More efficient or different equipment might be installed in the rack in the future or servers may not run near full capacity, leaving additional power capacity to power more equipment. The current best practice is to standardize on IEC C-13 and/or C-19 PDU outlets and 208 V. Most servers and data center devices can run at 208 V (even up to 240 V).

Remember that the derating factor of 70% was just an estimate. Research has been done with sophisticated rack PDUs that accurately measure power at the outlet. The findings were surprising. Even at peak power consumption, 15% of the servers drew 20% or less of their nameplate rating. Equally surprising was that nearly 9% drew 81% or more of their nameplate rating. The point here is that the actual power consumed as a percentage of nameplate rating can vary widely. Ideally, data center managers should measure the actual power consumption rather than use a rule-of-thumb average such as 70%. If the actual overall average is closer to 40%, as it was in the study, deploying power at 70% of nameplate is wasteful and strands unused power.

If a cabinet populated with 30 1U servers has dual power feeds and the servers require an average of 150 W each, then the total power requirement for a cabinet is 150 W × 30 servers = 4.5 kVA. Assuming 250 VA for additional equipment, like an Ethernet switch and a KVM switch, this brings the total to 4.75 kVA. So a 208 V, 30 A PDU, which is rated at 5 kVA, would be sufficient. Such a PDU can carry the full load of 4.75 kVA in a failover situation when the power feed to one side of the cabinet fails or is taken down for maintenance. Typically, each PDU would be carrying only 40% of the 4.75 kVA.

It is also important to note that three-phase Wye 208 V rack PDUs are able to support both 120 and 208 V in the same PDU. This can be handy for situations where a variety of equipment types with different voltage requirements need to be racked together.

29.4.2.1 208 V Single-Phase versus 208 V Three-Phase Rack PDU In a rack of 42 1U servers, if each server consumes an average of 200 W, then the total power consumption is 42 × 200 W = 8.4 kW. To allow for the NEC requirement of 80%, the rack needs 10.5 kVA (8.4 kW/0.8). To allow for redundant power feeds, two rack PDUs able to provide 10.5 kVA are required. The 208 V single-phase at 60 A (48 A rated) can deliver 10.0 kVA. This could suffice, particularly if the 200 W per server estimate is on the high side. Another alternative is the 208 V three-phase at 40 A (32 A rated), which can deliver 11.5 kVA. The 208 V three-phase alternative provides headroom to add higher-power-demand servers in the future and can handle the existing servers even if their average power consumption increases from 200 to 220 W.

The use of three-phase power enables one whip or rack PDU to deliver three circuits instead of just one. The whip or input power cord on the rack PDU will be somewhat larger for three-phase power than single-phase power because instead of three wires (hot, neutral, and ground), a three-phase cable will have four or five wires.

The two three-phase alternatives are Delta and Wye. A three-phase Delta system will have four wires: Line 1 (hot), Line 2 (hot), Line 3 (hot), and a safety ground. Individual circuits are formed by combining lines. Three circuits are available—L1+L2, L2+L3, and L1+L3. The power on each of the lines is a sine wave (this is also the case for single-phase power), but each of the three sine waves is 120° out of phase with the other two.

For three-phase power, the sine waves are 120° out of phase, so calculating VA is slightly more complex because we need to include the square root of 3, which is 1.732. The apparent power formula for three-phase is V × derated A × 1.732 = VA. As an example, 208 V, 40 A (32 A derated) three-phase is 208 V × 32 A × 1.732 = 11.5 kVA. In other words, the three-phase Delta deployment provides more than 170%, or 70% more, than the comparable single-phase, single-circuit deployment.

A three-phase Wye system will have five wires: Line 1 (hot), Line 2 (hot), Line 3 (hot), a neutral, and a ground. Individual circuits are formed by combining lines and a line with the neutral. As an example, a three-phase 208 V Wye rack PDU supports three 208 V circuits (L1+L2, L2+L3, L1+L3) and three 120 V circuits (L1+N, L2+N, L3+N). Three-phase Delta and three-phase Wye have the same apparent power, but the three-phase Wye can provide two different voltages.

In North America, there may be a requirement for 120 V convenience outlets such as NEMA 5-15R (120 V, 15 A, 12 A rated) or 5-20R (120 V, 20 A, 16 A rated). These can be supported by 208 V three-phase Wye PDUs where wiring between lines (L1, L2, L3) and lines and the neutral can provide power to both 208 and 120 V outlets. Whether the three-phase wiring is Delta or Wye, the voltage is always referenced to the line-to-line voltage, not the line-to-neutral voltage. This is even true in the following 400 V example where all the outlets are wired line to neutral.

Since the Wye system adds a neutral wire, many data centers are wired for Wye and use whips terminated with Wye receptacles such as NEMA L21-30R. This means the data center can use Wye PDUs that support 120/208 V or use Delta PDUs that support only 208 V without needing to change the data center wiring. A Delta PDU would use a NEMA L21-30P (the mating Wye plug) but would not use a neutral wire inside the PDU. This is a perfectly acceptable practice. For example, a data center could deploy Delta PDUs to racks where there is only a need for 208 V and Wye PDUs to racks where there is a need for both 120 and 208 V.

Three-phase cables may be slightly larger than single-phase cables, but it is important to remember that one slightly thicker three-phase cable will be significantly smaller and weigh less than three single-phase cables for the same voltage and amperage.

29.4.2.2 Rack PDU 400 V Three-Phase As shown in the 208 V/120 V example, three-phase Wye wiring is a convenient way to step down voltage. This is particularly true for 400 V power. A generally accepted method of delivering substantial power to densely packed racks is via 400 V three-phase Wye rack PDUs. A 400 V power distribution from panels to racks is now an accepted practice. A data center designer could specify 400 V Wye whips to 400 V Wye rack PDUs. Since much data center equipment can safely operate on voltages ranging from 100 to 240 V, the 400 V Wye PDU can provide three circuits—L1+N, L2+N, L3+N—each supplying 230 V (400 V/1.732). The 400 V Wye rack PDUs do not lend themselves to supporting 120 V outlets as do 208 V Wye rack PDUs.

29.4.3 Rack PDU Selection

29.4.3.1 Rack PDU Selection and Special Application Requirements There are many factors involved in selecting a rack PDU. Data center location, application, and ITE requirements, available power, cabinet, energy management and efficiency objectives, etc. will combine to dictate what type of PDU should be used. Some of the considerations in the following will guide you to select the feature set and hence the type of PDU you will need to satisfy your requirements.

What is the type of equipment and how many devices are going into the cabinets, for example, 42 × 1U servers with a single feed per device versus three 10U high blade servers with six power supply feeds per server? The answer will help define the physical configuration, for example, number and type of outlets, and capacity of your PDU(s), for example, how much power (kW) the PDUs need to support. Average rack power requirements have risen from 6.0 kW in 2006 to 7.4 kW in 2009 and 12.0 kW in 2011, and it is not unusual to see racks wired to provide as much as 30 kVA.

Clearly, decision criteria for 24/7 manned sites will be different than remote management of lights-out facilities. If you need remote or lights-out management of a facility, then you will probably need a switched PDU, which will require more security and user access management. Remote applications may also call for SNMP management.

Integration with directory services like LDAP or Microsoft's Active Directory is increasingly a requirement for controlling access to resources, rather than requiring a separate access control system. This capability is applicable to all applications, requiring central authentication, local or remote. And for many data center applications, for example, federal government and financial institutions, encryption and strong password support are necessary for remote access.

The rack PDU must supply uninterrupted power to each device plugged into it. You will want to prevent or mitigate any events that can potentially cause the circuit breaker on the rack PDU or upstream to trip. Outlet sequencing is a valuable feature to prevent inrush current from tripping a circuit breaker by establishing a sequence and appropriate delay for powering multiple devices. Outlet sequencing not only prevents the undesired tripping of a circuit breaker but also lets the user specify the order in which services (device(s)) come on line or are shut down during power cycling. For example, you will want to power the database service before the Web servers. This capability is most useful when used in conjunction with the outlet grouping capability (see in the preceding text).

For some applications and equipment, you many need a customizable alarm threshold for each outlet, with the capability to switch off an outlet should it exceed a certain power draw. This would prevent a temperature or other sensor (see Section 29.3.2) from causing a shutdown of servers. An advanced application is HVAC control using the temperature reported by a PDU's temperature sensor.

In many mission-critical environments, managed devices often have multiple feeds, which will be fed from different feeds or circuits for failover and redundancy. The device needs to be managed as a single device regardless of the number of power supplies/plugs, and all outlets must be handled simultaneously. This capability is applicable to all applications, local or remote.

Event-driven power cycling of an outlet/device is required for some applications, particularly for remote or unmanned sites. For example, if a device in a remote location fails to respond and the WAN is not operational, there are basically two options: first, an expensive, time-wasting truck roll to restart and, second, a rack PDU with the intelligence to trigger a restart of a malfunctioning device, for example, if the device has not responded for 20 min recycle power to the device.

If there is a need to maximize power efficiency, then rack PDUs can provide valuable data to support those efforts. Look for current, voltage, and power factor measurements at the PDU, line, breaker, and outlet level. Look for accurate kilowatt-hour metering at the outlet level, especially if you intend to report or charge back individuals or groups for usage. Metering accuracy can vary significantly, and for some rack PDUs, calculations may be based on assumptions and not actual real-time measurements.

29.4.3.2 Rack PDU Functionality Rack PDUs can vary significantly, not only in operational functions they offer, but also in their monitoring and data collection. The following is an overview of the strengths and weaknesses of the four types/classes of rack PDUs previously defined in Section 29.2.1.1. Clearly, our class definition is not rigid, since features offered by vendors will vary and you will want to select PDUs based on the total fit to your requirement, but this can be a useful guide in your selection.

Basic PDUs

- **Strength**: Basic, lowest cost, proven technology, and highly reliable
- **Weakness**: Lack instrumentation and are not manageable on any level

Metered PDUs

- **Strength**: Provide real-time monitoring of PDU current draw. User-defined alarms alert IT staff of potential circuit overloads before they occur.
- **Weakness**: Limited data, for example, no outlet-level or environmental data and no outlet switching.

Switched PDUs

- **Strength**: Offer some or all the features of metered PDUs plus remote power on/off capabilities, outlet-level switching, and sequential power-up
- **Weakness**: Must be managed carefully and risk of inadvertent power cycling. May not be appropriate for some environments, such as blade servers. Usually limited data, for example, no outlet-level monitoring or critical environmental data

Intelligent PDUs

- **Strength:** State-of-the-art devices are remotely accessible via Web browser or CLI. Models include all the features of switched devices (though they may be switched or unswitched) plus outlet-level monitoring, standard-based management, integration with existing directory servers, enhanced security, and rich customization. Provide comprehensive data including current, voltage, apparent power, active power, real-time environmental data, and, often, real-time kilowatt-hour (kWh) metering.
- **Weakness:** Higher initial cost relative due to their greatly enhanced feature set.

29.4.3.3 Benefits of an Intelligent Rack PDU The IT industry has dramatically chosen to move to more sophisticated, manageable systems. This fact is no more in evidence than the dramatic trend to the use of intelligent PDUs.

A truly intelligent PDU will provide real-time outlet-level and PDU-level power monitoring, remote outlet switching, and rack temperature and humidity monitoring. For top-tier data centers, deployment of intelligent PDUs can make a significant difference in the ability of IT administrators to improve uptime and staff productivity, efficiently utilize power resources, make informed capacity planning decisions, and save money. And, in so doing, they will operate greener data centers. Clearly, if your data center has dozens of racks, then the greatest benefits will be realized by using a rack PDU management system to consolidate data acquisition, reporting, as well as PDU administration and control. Here are a few practical reasons to be selecting intelligent PDUs for your racks:

Improve uptime and staff productivity

- Monitoring power at a PDU and individual outlet level, with user-defined thresholds and alerts via e-mail or SNMP, provides awareness of potential issues before they occur.
- Remote reboot of servers and ITE from anywhere in the world via a Web browser reduces downtime and personnel costs.

Use power resources safely

- User-configurable outlet-level delays for power sequencing prevent circuits from tripping from ITE inrush currents.
- Control of outlet provisioning prevents accidently plugging ITE into circuits that are already heavily loaded and are at risk of tripping circuit breakers.

Make informed power capacity planning decisions

- Outlet-level monitoring may identify some simple rearrangements of equipment to free up power resources by balancing power demands across racks.
- Monitoring power at the outlet level can identify equipment that may need to be changed to stay within the margin of safety of defined thresholds.
- Monitoring rack temperature and other environmental conditions can prevent problems, especially when a data center is rearranged and airflow patterns change.

Save power and money

- Monitoring power at the outlet level combined with trend analysis can identify ghost or underutilized servers that are candidates for virtualization or decommissioning.
- Remote power cycling enables IT managers to quickly reboot hung or crashed ITE without incurring the cost of site visits.
- Temperature and humidity sensors help data center managers optimize air-conditioning and humidity settings and avoid the common practice of overcooling and related waste of energy.

29.4.4 Power Efficiency

29.4.4.1 PUE Levels The Green Grid defines three levels of PUE: Basic or Level 1, Intermediate or Level 2, and Advanced or Level 3. Many industry analysts recommend measuring IT power consumption at the Intermediate, Level 2, that is, at the PDU level. While it is true that PDU-level power consumption will provide the denominator needed to calculate PUE, this information alone is unlikely to be sufficient to drive the best efficiency improvement decisions. Regardless of the PUE level you choose to employ, the best practice is to gather data over a time period of “typical” power usage to ensure that the peaks and valleys have been captured in calculating your PUE to establish a baseline and to track your improvements. There are many tools for collecting the data you need, described elsewhere in this book.

29.4.4.2 Why Advanced Level 3 PUE? An improved (lower) PUE can be misleading since that can result from inefficiencies in the power consumed by ITE, which merely increases the denominator. A lower PUE is generally better than a higher one, but it is possible to implement measures that reduce data center energy consumption yet actually increase your PUE. For example, if you were to replace older, less efficient servers with more efficient ones, or eliminate ghost servers, or turn off servers that were idle during the night, or employ server virtualization, the net result is power reduction, but your PUE would actually increase. The detailed IT load data from Level 3 provides the granularity of information to reduce energy consumption, not just improve the PUE metric. Clearly, the PUE (and its inverse DCIE) becomes a more useful beacon once you have built efficiency into the ITE performance, and to do that, you will want the granular power usage data for the Advanced, Level 3 PUE metric. Then you can attack the numerator and squeeze inefficiencies out of the infrastructure.

29.4.4.3 The Advantages of High Power A single-phase 120 V at 100 A (80 A rated) circuit provides 9.6 kVA. A single-phase 208 V at 60 A (48 A rated) circuit provides 10.0 kVA. A three-phase 208 V at 40 A (32 A rated) circuit provides 11.5 kVA. A single-phase 230 V at 60 A (48 A rated) circuit provides 11.0 kVA. A three-phase 400 V at 20 A (16 A rated) circuit provides 11.1 kVA.

Running higher voltages at lower currents means smaller cables that use less copper, weigh less, take up less space, and cost less. Running three-phase power instead of single-phase power means fewer cables, which simplifies deployment as well.

Plugs and receptacles are also less expensive at higher voltages and lower current ratings. For example, a 30 A 400 V three-phase Wye (16.6 kVA) plug (Hubbell NEMA L22-30P) costs $32 and the receptacle costs $41. A 60 A 208 V three-phase Delta (17.3 kVA) plug (Mennekes IEC309 460P9W) costs $166 and the receptacle costs $216. The plug/receptacle combination is $73 versus $382.

There are other benefits to higher voltages. A 400 V power circuit will eliminate voltage transformations and can reduce energy costs by approximately 2–3% relative to 208 V distribution and approximately 4–5% relative to 120 V distribution.

Consolidating data centers will generally reduce total power consumption but may create opportunities for use of high-density racks and high-power rack PDUs. For example, a 42U rack filled with 1U servers consuming 250 W each draws 10.5 kW, which would require two three-phase 208 V, 50 A circuits providing 14.4 kVA each. Taking advantage of blade servers might lead to deploying five blade chassis in one rack, which would require two three-phase 208 V, 80 or 100 A or two three-phase Wye 400 V, 50 or 60 A rack PDUs. These examples allow sufficient headroom should one of the feeds fail. They also support the North American requirement for 80% derating.

High-density racks can be deployed in small, medium, or large data centers. Even small data centers benefit from high-power racks for multiple blade servers or densely packed 1U servers.

29.5 FUTURE TRENDS FOR RACK PDUS

Two primary forces are influencing rack PDU development and innovation trends. First is the demand for increasing power and density of ITE at the rack or compute density per U of rack space. Second is the industry-wide goal, even mission, to create energy-efficient (often called “green”) data centers, including carbon footprint reduction. Both trends challenge the PDU vendors to improve both hardware and software design; and the second requires all IT and facilities organizations to better understand how the data center power is consumed and take active measures to reduce it.

The aforementioned trends are underscored in the Frost and Sullivan December 2008 survey and report on the World Power Distribution Unit Market. The report shows a healthy 17.1% compounded annual growth rate (CAGR) for overall rack PDU revenue from 2009 to 2015. However, they project 28.6% CAGR for three-phase PDUs versus 10.4% CAGR for single-phase PDUs and predict 23.3% CAGR for PDUs with intelligence and only 8% CAGR for basic PDUs. The “high impact” factors that are driving PDU demand are increasing power consumption and higher densities along with increasing need for PDU intelligence.

29.5.1 Higher-Density, Higher-Power Rack PDUs with Sensors

The growing popularity of 1U servers, blade servers, network-attached storage, storage area networks (SANs), and multigigabit, chassis-based network communications gear places enormous demands on rack PDUs. For example, four blade server chassis in a single rack could draw in excess of 20 kW of power creating power and cooling challenges for data center managers. From a power perspective, racks will require three-phase power with 60, 80, even 100 A of service. There are some data centers bringing 400 V three-phase service to the rack, to accommodate power demand while increasing efficiency from reduced voltage step-downs. Similarly, end users are packing dozens of 1U servers into a single rack and pressing rack PDU vendors to support 40+ outlets and 20+ kW.

Server virtualization is a major trend in data centers, leading to improved efficiency and cost reduction. However, running multiple virtual machines on one server will drive up its total power consumption; and a rack containing several such servers could experience a lot more power consumption driving the need for additional power load visibility to optimally manage power capacity.

More power consumption means more cooling to remove the additional heat. PDU vendors will be expected to supply the basic environmental sensors for heat, humidity, and airflow to help understand the overall environmental conditions and to identify zones that must be fine-tuned or supplemented with dedicated or specialized cooling.

29.5.1.1 Customizing ITE for Power Efficiency One trend to watch is the design and deployment of custom servers, power supplies, rack PDUs, etc. to maximize power usage efficiency. For example, Facebook along with Open Compute has begun to deploy 480 V three-phase Wye power where each line is wired to the neutral so the outlets deliver 277 V. This Wye configuration with lines wired to the neutral is the same wiring configuration as the 400/230 V wiring described earlier. This approach is very efficient, but it is highly customized since most ITE today are not built with power supplies that support 277 V. Furthermore, common data center receptacles are IEC C-13 and C-19, which do not support 277 V.

The savings and efficiencies (1–2% over 400/230 V three-phase systems) are sufficient that Facebook/Open Compute can justify building custom triplet racks, custom servers with custom power supplies, custom battery/UPS, and 480/277 V rack PDUs with custom Tyco 3-pin Mate-N-Lock outlets.

29.5.2 Increased Intelligence at the Rack to Support Efficiency Initiatives: “Smart Rack”

Many data centers have grown larger and more complex in recent years as the consolidation trend continues. With increasing size and complexity, there is a greater need to drive intelligence to the ITE at the rack to create what industry people are beginning to think of as the “Smart Rack.”

Every data center, regardless of size, is designed to support the servers at the rack where the actual computing is taking place. It is also where the vast majority of the power is being consumed. Proper monitoring and metering of the

ITE along with environmental sensors at the rack will collect the data necessary to produce the most significant overall efficiency, savings, and operational improvement. Collection and analysis of actual energy data will enable you to maximize the use of current resource capacity and take advantage of the capacity of planning tools to "right size" the data center for future requirements. This will allow you to eliminate or defer capital expenses of data center expansions while improving day-to-day energy efficiency and overall IT productivity.

Capacity planning based on nameplate data is no longer sufficient. Efficiency improvement is an information-driven activity. In order to formulate and drive the most effective decisions, you will need to collect IT device CPU utilization and their corresponding actual power usage. More energy efficiency will be gained if such planning is based upon the trends observed from the actual data over time. Furthermore, the actual data collected at the rack level can be integrated with the overall data center infrastructure management (DCIM) systems and data center energy management systems for complete data center and power chain visualization, modeling, and planning, which can lead to further improvements in the data center ecosystem, for example, computing carbon emissions generated by IT devices to report on and take steps to lower your carbon footprint.

Efficiency can also be gained from software that offers policy-based power control to automatically turn servers on and off based on granular power consumption data and a set of preestablished static or even dynamic rules. These power saving applications can be found in development labs, Web server farms, and cloud computing environments. They enter into mainstream data centers where the deployment of intelligent PDUs will enable their functions.

Creating energy-efficient behavior throughout your organization is a key factor in reducing waste and costs; and the essential ingredient to affect behavior is individual awareness/accountability for energy usage. Of course, to be effective, any such energy reporting or charge-back system must be based on credible, comprehensive, and coherent usage data, so PDU vendors will be expected to deliver the highest accuracy for energy usage at every level of the organization.

29.5.3 Integration with Higher-Level Data Center Management Systems

In recent years, a variety of software products have been introduced to help both IT and Facilities people manage the data center. While the category name may differ—Physical Infrastructure Resource Management (PRIM), DCIM, Data Center Service Management (DCSM)—these applications provide most of the following major functions: database of all physical data center assets with detailed data for IT, power and HVAC equipment, physical data center layout, and cable connections; change management; 2D or 3D visualization of the data center building with drill down to lowest-level data element; and capacity planning based on availability of floor and rack space, power, cooling, etc.

The data required to manage data center infrastructure and energy effectively are collected from power devices along the entire power chain up to the IT devices, from the IT devices themselves, from the environmental sensors, and from data center layout maps, cable plans, and cooling system design documents. The more data collected, and the more accurate the data will be, the better the data center personnel are enabled to manage the data center to support critical IT operations reliably, efficiently, and cost-effectively.

The following is a simplified view of data measurement, collection, compilation, analysis and correlation, and decision support:

1. Intelligent rack PDUs measure essential power data at a predefined frequency and store such data in memory.
2. Data collection from rack PDU management (or power management) system polls the intelligent PDUs through industry standard management protocols such as SNMP.
3. The data collection service can be part of the intelligent PDU vendor's rack PDU management system (Raritan's Power IQ is an example), or it can be part of a DCIM (Emerson/Aperture's Vista is an example) or energy management system (Schneider's ION Enterprises can be an example). For scalability reasons, data collection is typically delegated to specific PDU vendor's PDU management system, which is deployed along with intelligent PDUs to administer, maintain, and troubleshoot the PDUs, as well as to collect power statistics from these intelligent PDUs.
4. The rack PDU management system can use the collected data to perform first level of analysis. This will help to visualize the power trends and pinpoint some potential issues. Then the collected data as well as the compiled information can be used by DCIM or energy management system for further analysis.
5. The energy management or DCIM system has visibility beyond the PDU management system. They can, for example, poll information from upstream smart power devices; and they typically also feature the static information such as data center physical layout, cable plan, and HVAC deployment information, making them more suitable for analysis that must take into consideration many more factors beyond the intelligent PDUs.

With the advanced analysis conducted by a DCIM or energy management system, data center management staff can make their day-to-day operational decisions as well as longer-term strategic planning, to provide reliable and high-quality power for business applications while reducing waste in data center energy consumption.

FURTHER READING

Alger D. Build the Best Data Center Facility for Your Business: A Comprehensive Guide to Designing and Operating Reliable Server Environments. Indianapolis: Cisco Press; 2005.

ASHRAE. *Thermal Guidelines for Data Processing Environments*. 2nd ed. Atlanta: ASHRAE; 2004.

ASHRAE. ASHRAE Workshops on Improving Data Center Energy Efficiency and Best Practices. NYSERDA Sponsored Workshop; New York: ASHRAE; November 6, 2008a.

ASHRAE. 2008 ASHRAE Environmental Guidelines for Datacom Equipment. Atlanta: ASHRAE; 2008b.

ASHRAE. *High Density Data Centers: Case Studies and Best Practices*. Atlanta: ASHRAE; 2008c.

ASHRAE. *Best Practices for Datacom Facility Energy Efficiency*. 2nd ed. Atlanta: ASHRAE; 2009.

Cuthbertson D. Practical data centre management training workshop, part 1—managing the facility. Workshop; Somerset: Square Mile; July 24, 2008.

Data Center Users Group, Emerson Network Power. Data Center Users' Group Special Report: Inside the data center 2008 and beyond. Columbus: Data Center Users Group, Emerson Network Power; 2008. White Paper WP165-118, SL-24634.

Digital Realty Trust. *kW of IT Load. The New Chargeback Mechanism*. San Francisco: Digital Realty Trust.

Frost and Sullivan. Worldwide Power Distribution Unit Market, N2FE-27. Rockville: Frost and Sullivan; December 2008.

Haas J, Monroe M, Pflueger J, Pouchet J, Snelling P, Rawson A, Rawson F. Proxy proposals for measuring data center productivity. Beaverton: The Green Grid; 2009. White Paper #14.

Information technology equipment—safety—part 1: general requirements. Northbrook: Underwriters Laboratories, Inc.; 2007. UL 60950-1.

NFPA 70: National Electric Code. Quincy: National Fire Protection Association; 2008.

Raritan Inc. Data center power overload protection: circuit breakers and branch circuit protection for data centers. Somerset: Raritan Inc. 2009a. White Paper. Available at http://www.raritan.com/resources/white-papers/power-management/. Accessed on May 22, 2014.

Raritan Inc. Data center power distribution and capacity planning: understanding what you know—and don't know—about power usage in your data center. Somerset: Raritan Inc.; 2009b. White Paper. Available at http://www.raritan.com/resources/white-papers/power-management/. Accessed on May 22, 2014.

Raritan Inc. Power distribution units (PDUs): power monitoring and environmental monitoring to improve uptime and capacity planning. Somerset: Raritan Inc.; 2009c. White Paper. Available at http://www.raritan.com/resources/white-papers/power-management/. Accessed on May 22, 2014.

Raritan Inc. Deploying high power to IT equipment racks. Somerset: Raritan Inc.; 2012. White Paper V1156.

U.S. Environmental Protection Agency. Report to congress on server and data center energy efficiency public law 109–431. Washington: U.S. Environmental Protection Agency, ENERGY STAR Program; 2007.

Verdun G, editor. The Green Grid metrics: Data Center infrastructure Efficiency (DCiE) detailed analysis. Beaverton: The Green Grid; 2008. White Paper #14.

Wikipedia. "1-Wire," of Dallas Semiconductor Corp. Available at http://en.wikipedia.org/wiki/1-Wire. Accessed on May 22, 2014.

30

RENEWABLE AND CLEAN ENERGY FOR DATA CENTERS

William Kao
University of California Santa Cruz, Silicon Valley Extension, Santa Clara, CA, USA

30.1 INTRODUCTION

With the tremendous growth of the Internet, social media, and cloud-based computing, all the information generated—videos, audios, emails, status updates, news, and tweets—ends up in giant data processing facilities called data centers. These facilities consume huge amounts of electricity, amounting to 1.5–2% of global energy demand and it's growing at a rate of 12% a year. Data centers are thus becoming the fastest growing users of energy.

Growth of corporate sustainability and social responsibility awareness, potential future legislation on carbon emissions such as "cap and trade," and pressure from environmental activist groups like Greenpeace International are driving industry leading companies like Google, Apple, Facebook, and Yahoo to make heavy investments in renewable and clean energy to power their data centers. At the same time, clean tech fuel cell companies are now targeting their fuel cell line products at data center operators as a new market for distributed cleaner power.

In this chapter, we plan to cover the fundamentals of renewable energy and fuel cells. The definition of a green data center is presented along with the latest green data center trends. How most companies (especially cloud computing ones) are going "green" by adopting heavy use of the clean technologies presented in the previous sections in their new data centers are described.

All new corporate data centers are going green by:

1. Implementing renewable energy and clean technology solutions (solar, wind, hydro, geothermal, and fuel cells). Google and Microsoft selected states like Iowa, Oklahoma, and Oregon as sites for their new data centers to have access to that state's wind power. They built data centers in the northwest for hydropower. Google also supports geothermal energy by investing more than 10 million on Enhanced Geothermal Systems (EGS), which involve artificial geology enhancements such as drilling a hole, cracking rock, and conveying water to make steam.
2. Measuring key metrics and developing "energy efficiency" initiatives:

   Measuring the data center's two key metrics: the power usage effectiveness (PUE),[1] which is total facility power divided by IT equipment power, and the carbon usage effectiveness (CUE),[2] a metric by which data center operators can gauge the intensity of their CO_2 emissions per kilowatt-hour of energy used.

   PUE represents the measure of the power going into the facility at the utility meter divided by the power going to the IT load, measured either at the power distribution unit or uninterruptible power supply (UPS):

$$\text{PUE} = \frac{\text{total energy}}{\text{IT energy}}$$

   When the PUE is closer to 1.0 it means that most or all of energy is going toward IT usage, while a larger number means more is being lost or getting diverted to other uses. According to industry sources, a typical PUE number for most data centers is 1.6.

[1] http://www.thegreengrid.org/~/media/WhitePapers/WP49-PUE%20A%20Comprehensive%20Examination%20of%20the%20Metric_v6.pdf?lang=en

[2] http://www.thegreengrid.org/~/media/WhitePapers/Carbon%20Usage%20Effectiveness%20White%20Paper_v3.pdf?lang=en

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

Together, the CUE and PUE metrics help describe a data center's relative energy efficiency and emissions intensity.

The ideal "green data center" is extremely energy efficient, with a PUE close to 1.0, high asset utilization, and a low CUE through use of green power. Thus, environmental impacts are mitigated from two essential angles—through a high degree of energy efficiency and the use of clean energy.

3. Some measures being adopted by new green data centers are as follows:
   - Separating hot and cold aisles to keep the hot and cool air separated
   - Raising the thermostat temperature and using more outside air for cooling, thereby reducing the use of energy-hungry chillers
   - Using wireless monitoring and management systems to monitor and control the temperature, humidity, and energy consumption of the data center
   - Using new direct current (dc) network that can significantly cut down electricity conversion losses

The remainder of this chapter will cover renewable energy basics, fuel cell basics, green data centers, and what various large corporations are embarking to make their data centers more green and efficient through the use of clean technology.

30.2 RENEWABLE ENERGY BASICS

Renewable energy is called "renewable" because the sources harnessed to create the energy renew and replenish themselves constantly and within a reasonably short period of time (i.e., months or years, not centuries as in the case of fossil fuels). These sources of energy include sun, wind, water, biomass, and heat from the Earth's interior and atmosphere.

The term "renewable energy" excludes energy created by nuclear fuels, such as uranium, and fossil fuels (oil, gas, and coal). Fossil fuels take millions of years to form and, once removed and used, require as many years to form again. The world's supplies of uranium and fossil fuels are limited.

An important point about renewable energy is that although quickly replenished, some of these forms of energy are intermittent on either a daily or a seasonal basis. There are days when the sun does not shine or the wind does not blow; and certainly, it's rare that sunshine and wind are consistent throughout the day. In some instances, the technology requires a way of storing the power that is created. In most cases, the electricity generated from these intermittent renewable sources is supplemented by electricity generated by energy storage systems such as pumped hydro, compressed air, molten salt, flywheel, batteries, etc.

30.2.1 Why Renewable Energy

Energy derived from fossil fuels takes a heavy toll on the environment. When fossil fuels are burned, they generate carbon dioxide and other greenhouse gases. These gases trap heat within our atmosphere, warming the planet and altering climate in a much shorter time than would occur naturally. This is already having devastating extreme weather effects such as increased and more frequent droughts, floods, and more severe storms.

Additionally, fossil fuels are finite. Oil production from easily accessible reserves is in decline and while huge coal reserves do exist, they are also limited. It takes many thousands of years for natural processes to create crude oil and coal, far longer than the rate at which we are using them.

Because of their uneven distribution around the world, they are a major driver of price hikes, as well as much political and economical unrest.

Electricity from renewable energy produces fewer greenhouse gas emissions, which are associated with climate change, than electricity produced from burning coal and fossil fuels. Similarly, renewable energy generally adds fewer other pollutants to the air, which include the following:

- Sulfur dioxide and nitrogen oxides that form acid rain `
- Particulate matter, which along with ground-level ozone forms smog
- Mercury, which can be transformed in the environment to become highly toxic that causes brain damage and heart problems

30.2.1.1 Greenpeace Although all hi-tech companies have made extraordinary gains in overall data center efficiency, Greenpeace has maintained a singular focus on the use of renewable energy, either through on-site generation or by choosing locations where electricity is sourced from renewable energy sources.[3]

Greenpeace recently released a report that profiles the energy usage of the facilities powering major Internet services. The Greenpeace's report, "How clean is your cloud?[4]" looks at the energy choices that power data centers and cloud computing.

30.3 RENEWABLE ENERGY TYPES

There are seven types of renewable electricity generation: solar, wind, bioenergy, hydropower, geothermal, wave, and tidal. A brief description of each will be given in the following sections.

[3]http://www.greenpeace.org/usa/en/

[4]http://www.greenpeace.org/international/Global/international/publications/climate/2012/iCoal/HowCleanisYourCloud.pdf

30.3.1 Solar Power

The sun is a renewable source of energy that is plentiful and environment-friendly. It is the source (directly or indirectly) for most of the forms of renewable energy. Of the seven forms of renewables, only tidal and geothermal energy are not attributed to the sun.

There are two major forms of solar energy: solar thermal, also known as concentrated solar power (CSP), and photovoltaic (PV).

30.3.1.1 Solar Thermal or CSP Systems CSP systems are of three kinds—power tower, parabolic trough, and parabolic dish.

- Power tower systems: This type of system uses a number of large, sun-tracking, flat-plane mirrors to focus the sun's light onto a central receiver at the top of a tower. The system pumps a fluid, either a high-temperature synthetic oil or molten salt, through the receiver where it is heated to 550°C and then used to generate electricity (Fig. 30.1).
- Parabolic troughs systems: This type of system (Fig. 30.2) uses a series of long troughs in the shape of a parabola. The parabola concentrates the light onto a receiver tube that is positioned along the focal line of the parabolic trough.

  Temperatures at the receivers can reach 400°C and produce steam for generating electricity. Usually, the troughs track the sun in one dimension (1D) as it moves from east to west during the day.
- **Parabolic dish systems:** A parabolic dish system (Fig. 30.3) uses parabolic dish-shaped mirrors to focus the sun's radiation onto a receiver positioned at the focal point of the disk. There is fluid in the receiver, which, when the sun's rays hit it, heats up to 750–1000°C. The very hot fluid is then used to generate

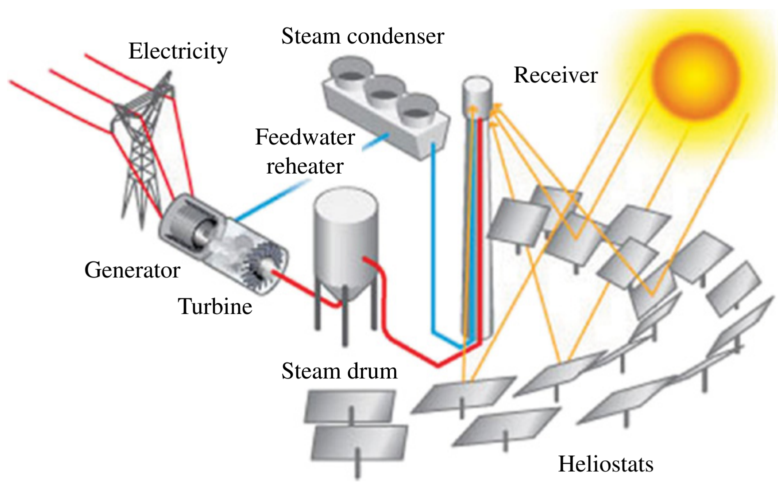


FIGURE 30.1 Solar power tower energy plant. Plant illustration from Energy Efficiency and Renewable Energy (EERE), U.S. Department of Energy.

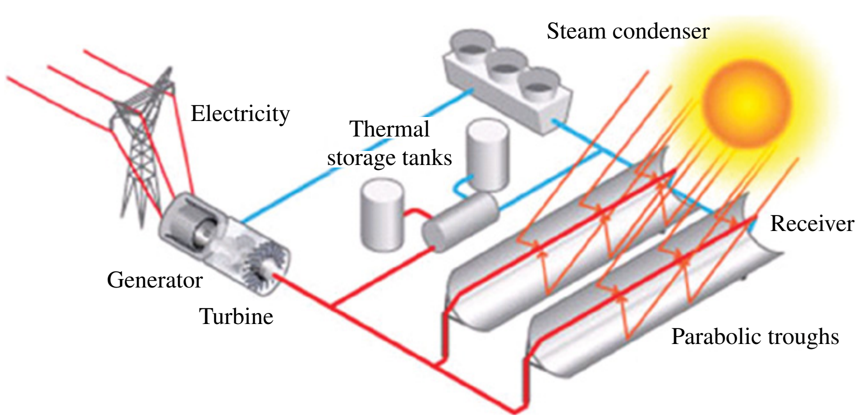


FIGURE 30.2 Solar parabolic troughs power plant. Plant illustration from Energy Efficiency and Renewable Energy (EERE), U.S. Department of Energy.

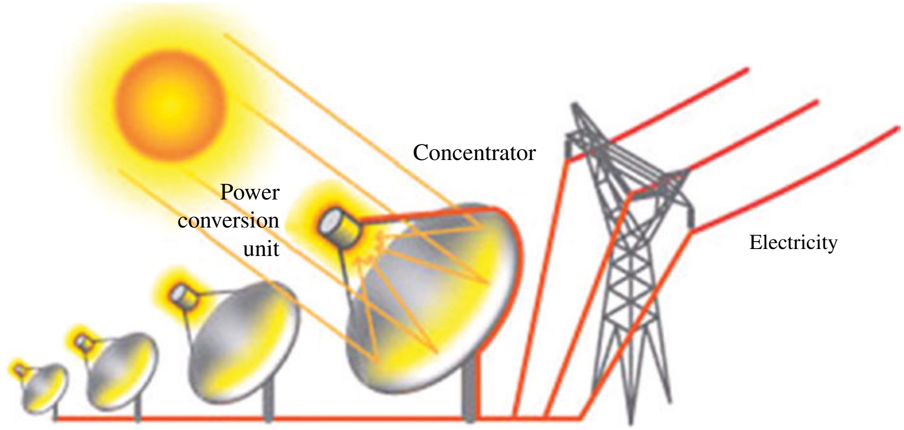


FIGURE 30.3 Parabolic dish engine power plant. Plant illustration from Energy Efficiency and Renewable Energy (EERE), U.S. Department of Energy.

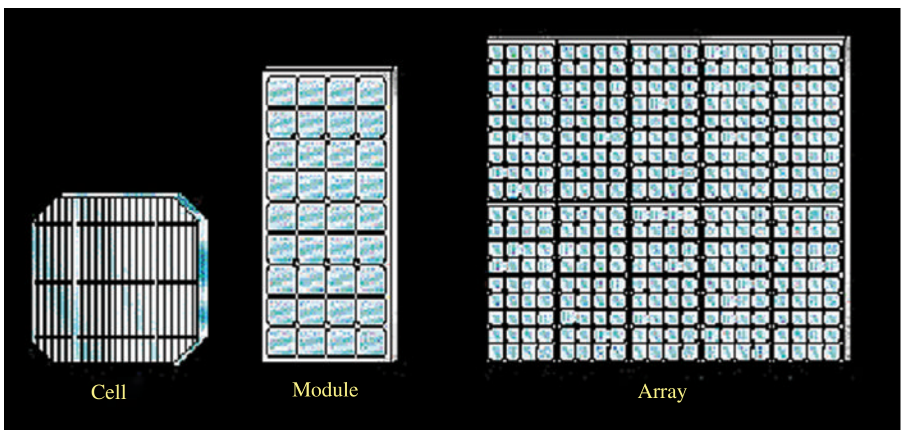


FIGURE 30.4 PV cells, modules, and arrays. Photo illustration from the U.S. Department of Energy, Energy Efficiency, and Renewable Energy Program.

electricity in a small engine attached to the receiver. Like the parabolic trough, a parabolic dish also tracks the sun's movements, but it does it in 2D, that is, both east–west and north–south.

30.3.1.2 Photovoltaic Energy The photovoltaic (PV) process turns the radiant energy of the sun into direct current electrical energy.

Photovoltaic Cells or solar cells are small semiconductor devices. First-generation solar cells are made of silicon, while second-generation uses thin film technology. As long as the sun shines on a PV cell, it produces a small flow of electricity— about 0.5V. To produce electricity in useful amounts, the cells are usually grouped together in panels or modules (Fig. 30.4). A full solar system is made up of an array of panels linked together. PV cells only work when the sun shines, so some PV systems include batteries that store power so it can be used at night or on cloudy days. PV cells produce direct current electricity. Most electric appliances and lights run on alternating current (AC) electricity, so PV systems often include a device called an inverter to convert direct current to alternating current electricity.

PV panels and arrays do not produce emissions when they create electricity and need only the energy from the sun to power them.

PV panels are a cost-effective source of power, which are specially effective and convenient for remote rural areas where there is no existent electric grid.

There is a growing interest in integrating PV arrays in the windows, roofs, and walls of houses and office buildings. This use of PV energy is called building-integrated PV (BIPV).

30.3.2 Wind Power

People have been using wind power for hundreds of years. Windmills were used to pump water and grind grain in the eighteenth and nineteenth centuries. The modern versions of windmills are called wind turbines. There are two primary

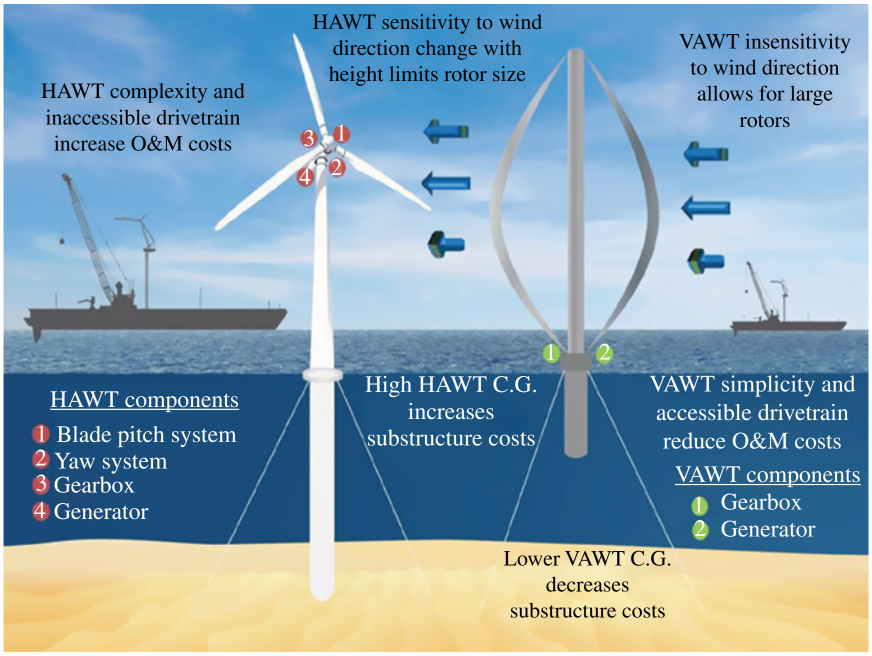


FIGURE 30.5 Two types of wind turbines. From Sandia National Laboratories, https://share.sandia.gov/news/resources/news_releases/vawts/#.U8dMNWdOVMy.

designs for wind turbines—horizontal (left) and vertical (right) axis wind turbines (Fig. 30.5).

The horizontal axis wind turbine (HAWT) looks like a windmill with two, but more often three, rotor blades affixed like a propeller to the front of the tower at its top. The gearbox, brake, and generator are housed in a casing or nacelle behind the rotor blades at the top of the tower.

The vertical axis wind turbine (VAWT) looks like an egg-beater. The rotor blades are attached at the top and close to the bottom of the tower and bulge out in the middle. The gearbox and generator are housed in a protective structure at the tower's base.

30.3.2.1 How Wind Turbines Work? The wind passes over the rotor blades, causing them to turn. The harder the wind blows, the more the energy that can be captured and the more the electricity that can be generated. If the wind is too strong, then the turbine will shut down by turning out of the wind and applying a braking mechanism that prevents the blades from turning too quickly and being damaged.

Wind turbines generally produce electricity when winds blow at more than 8 miles an hour. Production increases until it hits a maximum power at about 34 miles an hour. When winds blow at 55 miles an hour or more, most large wind turbines shut down for safety reasons.

Some wind turbines stand on their own. Others are grouped together at wind farms (Fig. 30.6). At wind farms, wind turbines need to be spaced at least five to six times the diameter of the rotor blades to prevent the turbulence of one turbine from affecting the flow of wind at another.

FIGURE 30.6 Wind farm. From EPA, http://blog.epa.gov/blog/2008/09/science-wednesday-better-together-wind-and-solar-power-in-california/.

Wind is an intermittent source of energy because it does not always blow at the speed required to generate electricity. Wind turbines generally capture an average of 15–40% of the total rated electricity generation capacity of the wind turbine. There are no significant air pollution and greenhouse gas emissions associated with this form of renewable energy. There is some noise, however, created by the rotor blades as they cut through the air. But slower rotor speeds (15–25 rpm) and new designs and materials have significantly reduced the noise level in the past several years. Today, the noise level at 250 m can be as low as 42–43 dB, which is less than the average background level of noise in city residential areas. Similarly, results of studies show that wind turbines have little effect on the population of birds, in part because utilities and private companies go to great lengths to make sure they do not site wind farms in the middle of migratory flight paths, and in part because in most areas the migratory flight paths of birds are higher than the turbines or the reach of the blades. The slower, constant blade speeds and solid tower designs that typify today's wind turbines also serve to lessen the potential for bird impingement impacts.

30.3.2.2 Wind Power Calculation The amount of power transferred to a wind turbine is directly proportional to the area swept out by the rotor, to the density of the air, and the cube of the wind speed. Thus, the usable power potentially available in the wind is given by

$$P = \frac{1}{2}\alpha\rho\,\pi r^2 V^3$$

where P = power in watts, α = an efficiency factor determined by the design of the turbine, ρ = mass density of air in kilograms per cubic meter, approximately $\rho = 1.225\,\text{kg/m}^3$ at sea level temperature, r = radius of the wind turbine in meters, and V = velocity of the air in meters per second.

In order for a wind turbine to work efficiently, wind speeds usually must be above 12–14 miles per hour. Wind has to be at this speed to turn the turbines fast enough to generate electricity. The turbines usually produce about 50–300 kW of electricity each. Wind speeds are a function of altitude H. Knowing the velocity V_0 at height H_0, one can compute the velocity V at different height H:

$$V = \left(\frac{H}{H_0}\right)^{\alpha} V_0$$

where $\alpha = 0.143$ (the 1/7 power law).

30.3.3 Bioenergy

People have been using bioenergy for thousands of years, burning wood and peat to warm their homes, cook their food, and forge their utensils.

30.3.3.1 Biomass Biomass refers to plant and animal material that can be burned to generate energy. It can take the form of crops, crop waste, trees, and animal waste. While these are renewable resources, biomass has similar challenges to fossil fuels in that carbon dioxide and other greenhouse/toxic gases are generated when combusted. Biomass is uniquely suited among renewable energy sources for conversion to transportation fuel. In the process of plant growth, light from the sun provides energy for the conversion of carbon dioxide and water into the carbohydrates that make up plant matter. The process of photosynthesis produces oxygen while removing carbon dioxide from the atmosphere:

$$\text{Sunlight} + 6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2 \quad (\text{carbohydrates, plant matter})$$

When the reaction is reversed during gasification (see Gasification), the carbon dioxide given off is the carbon dioxide that was previously removed from the atmosphere as the plant grew. While a minor amount of fossil fuel will be needed to produce and transport biomass, its net carbon balance is close to zero.

There are several ways of turning biomass into heat and electricity, including direct combustion, anaerobic digestion, cofiring, pyrolysis, and gasification:

- Direct combustion. Any organic material that is dry enough can be burned. The heat is used to boil water to produce steam, which turns a turbine attached to a generator to create electricity.
- Anaerobic digestion is a process that breaks down organic matter, such as the organic portion of municipal waste, in a tank, container, or lagoon that doesn't have any oxygen in it. The waste contains microorganisms that, when they digest biomass such as manure, organic waste, or waste in a landfill site, produce a combustible gas. The gas primarily comprises methane and carbon dioxide and is called biogas. This biogas, which is a reasonably clean fuel, can be used in an electrical generating plant. The digestion process also produces a "digestate" that can be separated into a liquid component, which can be used as a fertilizer, and a solid component, which can be used as a soil conditioner.
- Cofiring refers to the practice of introducing biomass into the boilers of coal-fired electricity plants. Adding biomass as a source of fuel helps to reduce the use of coal.
- Pyrolysis refers to the thermochemical process used to convert solid biomass to liquid fuel. During the process, biomass is heated in an oxygen-free tank to produce a gas that is rich in hydrocarbons, which is then quickly cooled to an oil-like liquid and a solid residue, or char, which is usually called charcoal and used for burning. Pyrolysis offers the advantage of producing renewable liquid fuels that can be more easily stored, transported, and burned than solid wood wastes.
- Gasification is a form of pyrolysis. It uses more air than pyrolysis when the biomass is heated. The resulting gas, called producer gas, is a mixture of carbon monoxide, hydrogen, and methane, as well as carbon dioxide and nitrogen. This gas is burned to produce steam, or used in gas turbines to produce electricity.

If biomass resources are managed wisely and the resulting combustion emissions are properly controlled, biomass has the potential to provide significant amounts of energy more cleanly and with much lower greenhouse gas emissions than nonrenewable fossil fuels, such as coal and oil. The direct combustion of biomass, however, can result in air emissions of concern.

30.3.3.2 Biofuel Biofuel can be broadly defined as solid, liquid, or gas fuel consisting of, or derived from, recently dead biological material, most commonly plants. Biofuel can be produced from any carbon source; the most common are photosynthetic plants that capture solar energy. Examples of liquid fuel such as ethanol or biodiesel are created from plant material and more recently, algae. The carbon cycle of a biofuel such as ethanol is shown in Figure 30.7. The use of food crops as fuel has caused prices of some grains such as corn to skyrocket in recent years. There are various current issues with biofuel production and use being discussed in the popular media and scientific journals. These include the effect of moderating oil prices, the food versus fuel debate, sustainable biofuel production, deforestation and soil corrosion, impact on water resources, human rights issues, poverty reduction potential, biofuel prices, and centralized versus decentralized production models.

Currently, research into refining the cellulosic ethanol process is continuing—this is where more woody stock and crop waste can be converted into ethanol rather than the crop itself, but a more promising technology is the use of algae to create biofuels.

Ethanol: Ethanol is used as an additive, usually mixed with gasoline in a blend of 10% ethanol and 90% gasoline. This is called E10. Drivers can use it in recent model cars without modifying the engines. Most of the ethanol made today is the result of a fermentation process using corn, grains, potatoes, sugar beets, or sugarcane. Using ethanol to fuel vehicles reduces drivers' dependence on gasoline, which is not a renewable fuel, and cuts emissions of carbon dioxide and some pollutants associated with smog.

Concerns have been expressed that to ensure a constant supply of raw material to produce ethanol, a company might buy up vast tracts of land to grow crops needed as feedstock. This may jeopardize an area's biodiversity as these tracts of land are devoted to one crop. There are also concerns that crops once used to feed people may be diverted to industry and that soil quality may deteriorate because parts of plants or trees once left behind to nourish the soil will now be used as raw materials in bioproducts.

Biodiesel: Biodiesel is made from renewable sources, such as vegetable oils from canola seeds, corn seeds, sunflower seeds, or flax seeds. These can be treated to create a clean-burning fuel known as biodiesel. The most direct way to extract the oil from the seeds is to use mechanical or mechanical/solvent extraction.

30.3.3.3 Biogas Biogas is methane gas, a by-product of decomposition, livestock production, cultivation of certain plants, and landfills that can be captured and burned. The

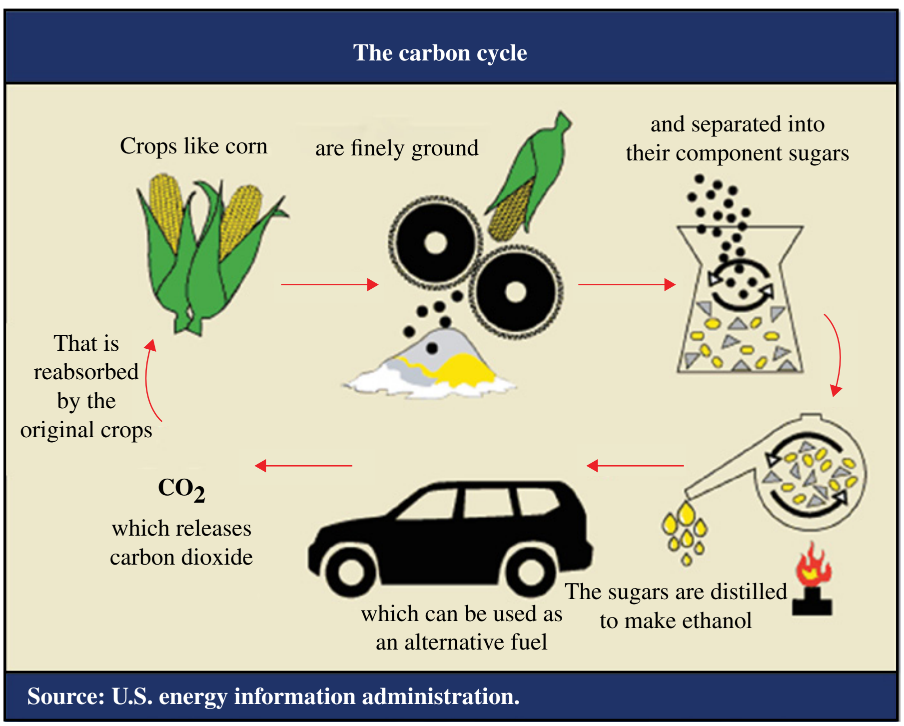


FIGURE 30.7 Carbon cycle for a biofuel. Image from http://www.window.state.tx.us/specialrpt/energy/renewable/ethanol.php.

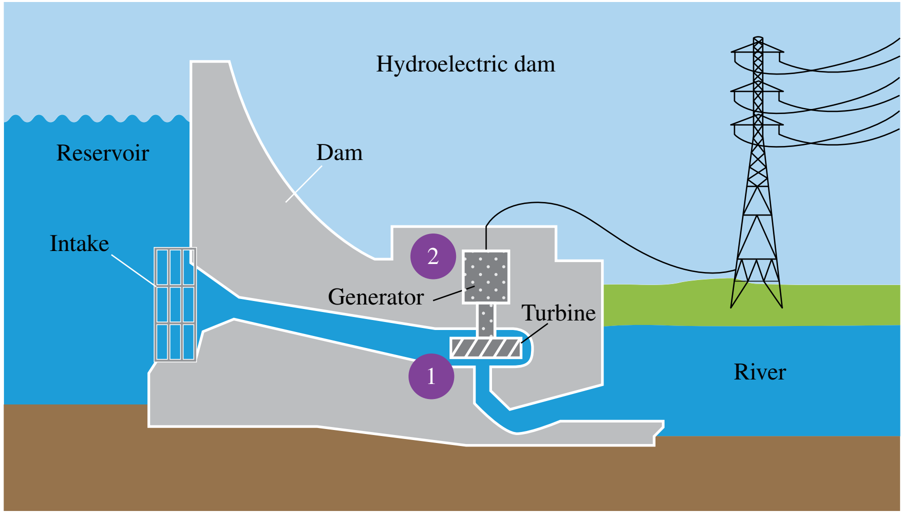


FIGURE 30.8 Diagram of a hydroelectric dam. From Environmental Protection Agency.

burning of methane gas is more desirable than allowing it to escape into the atmosphere as methane has a Global Warming Potential (GWP) 62 times that of carbon dioxide.

30.3.4 Hydropower (Hydroelectric Power)

Hydropower has been in use for thousands of years. It simply uses the inertia of the flow of a body of water to spin turbines, or power a mill for grinding grain.

Hydroelectric plants (Fig. 30.8) convert the potential energy of water to electrical energy by creating a drop in the elevation of the water. The amount of electricity generated depends on the vertical distance that the waterfalls and the water's flow rate. Flow is a measure of the volume of water moving past a point during a certain amount of time, usually a second. Many hydroelectric generating stations use dams to raise water levels upstream of the station and increase the drop in height to produce more electricity and/or to store water and release it to produce electricity to match changes in demand. Here is how it works:

- The water in the river or reservoir behind the dam flows through an opening, usually called an intake, and from there through a pipe called a penstock.
- The water flows through the penstock under pressure to its end, where there is a turbine.
- The force of the water turns the blades of the turbine, which turn the shaft inside the turbine. The turbine shaft is connected to a generator, which generates electricity.
- Once past the turbine, the water flows through a pipe, called a draft tube, out of the generating station into a channel, called the tailrace, and back to the river.

Most hydroelectric power comes from the potential energy of dammed water driving a water turbine and generator. The energy extracted from the water depends on the volume and on the difference in height between the source and the water's outflow. This height difference is called the head. The amount of potential energy in water is proportional to the head. To obtain very high head, water for a hydraulic turbine may be run through a large pipe called a penstock.

A simple formula for approximating the electric power production at a hydroelectric plant is

$$P = hrk$$

where P is power in watts, h is height/head in meters, r is flow rate in cubic meters per second, and k is a conversion factor.

Hydroelectric plants produce electricity relatively efficiently. In fact, they convert about 90% of the available energy from water into electricity; this is more efficient than any other method of generating electricity.

Some large-capacity hydroelectric projects require huge dams and reservoirs, which flood thousands of hectares of wilderness and disrupt the migration patterns of fish and wildlife. Many unique wilderness areas have been lost as a result, and many people have been forced to evacuate and relocate.

30.3.5 Wave Power

Winds blowing along the surface of the oceans create waves. Wave energy converters (Fig. 30.9) rely on the up and down motion of waves to generate electricity. Special equipment such as floating structures move with the waves and are attached to a generator that converts this movement into electricity.

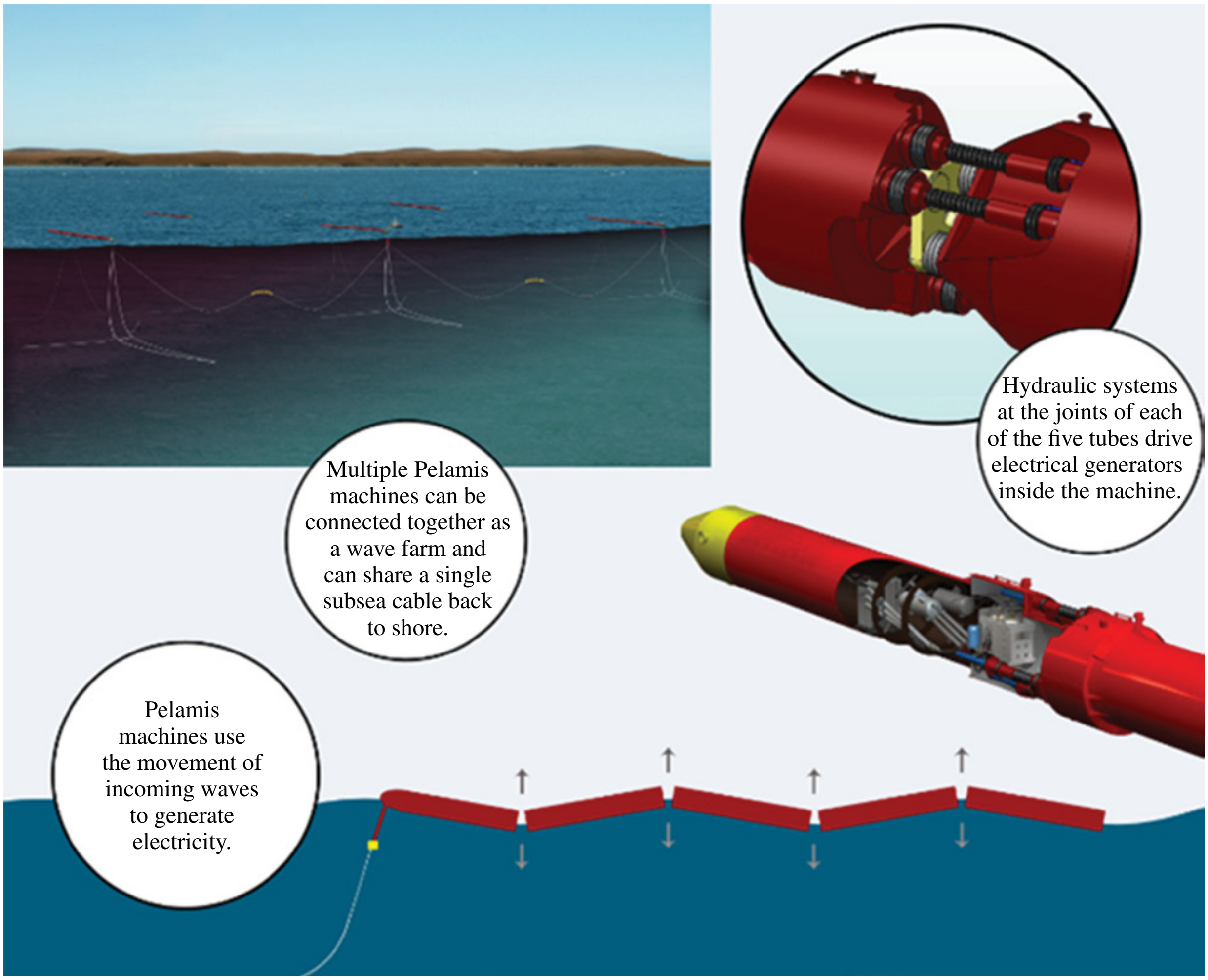


FIGURE 30.9 Wave power generator. Courtesy of Pelamis Wave Power, www.pelamiswave.com.

Wave energy systems do not need fuel to operate and do not produce polluting emissions. But they do have to be durable enough to withstand the beatings they take during severe storms. Some systems, such as the ones that are offshore, use visual and radar devices as navigational aids to boats and ships to avoid potential collisions.

Wave power is still relatively new; but aside from the further refinement and development required and posing navigational hazards, it has few drawbacks.

30.3.6 Tidal Power

Every day, twice a day, the tides rise and fall, in some places by only a few feet and in others by as much as 20 ft. As tides move in and out, the water embodies a huge amount of kinetic energy. This energy can be tapped with special turbines that will work regardless of whether the tide is ebbing or flooding.

The French built the first, and the world's largest, commercial-scale tidal generating plant in La Rance in the 1960s with a capacity of 240 MW. Although the tides rise and fall twice a day in all coastal areas, there must be a difference of at least 5 m between high and low tides for a tidal generating station to create cost-effective electricity. Today, about 40 areas in the world are considered suitable for tidal generating stations.

How Tidal Energy Works?

- Barrage: The simplest and oldest technology involves building a dam, known as a barrage, across a bay or estuary that has large differences in elevation between high and low tides. When the tide comes in, the water fills the area behind the barrage. When the tide starts to ebb, the gates of the barrage shut to hold back the water at its maximum height. Once the tide is out, the water is allowed to flow through holes near the bottom of the barrage where the turbine is located. The water, now running with great energy, turns the blades of the turbine that, in turn, generate electricity.
- Tidal turbine: Tidal turbines resemble wind turbines (Fig. 30.10), except that the blades or rotors are about

one-third of the way up the structure and are completely submerged in water. These turbines use the currents of tides that have velocities of between 2 and 3 m/s to turn the rotors or blades. Currents of more than 3 m/s put too much stress on the blades in the same way that gale force winds damage wind turbines. The major disadvantages of tidal power are the costs involved, corrosion, and high maintenance.

30.3.7 Geothermal Power

Geothermal energy harnesses the heat from the Earth's molten core. This heat can be drawn from several sources: hot water or steam reservoirs deep in the Earth that are accessed by drilling, geothermal reservoirs located near the Earth's surface, and the shallow ground near the Earth's surface that maintains a relatively constant temperature of 50–60°F. Water can be injected into holes leading to these rocks to create steam that can drive turbines to generate electricity.

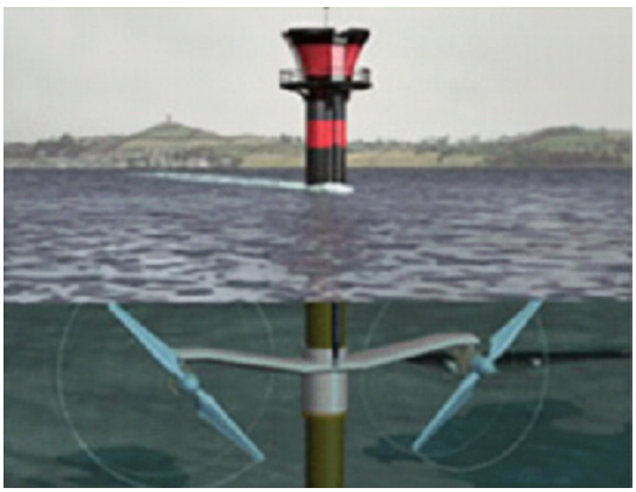


FIGURE 30.10 Underwater turbine. Photo courtesy of Marine Current Turbines.

Geothermal power plants (Fig. 30.11) use steam produced from reservoirs of hot water found a few miles or more below the Earth's surface to produce electricity. The steam rotates a turbine that activates a generator, which produces electricity.

There are three types of geothermal power plants—dry steam, flash steam, and binary cycle:

- Dry steam power plants draw from underground resources of steam. The steam is piped directly from underground wells to the power plant where it is directed into a turbine/generator unit. The Geysers in northern California is the only dry steam plant in the United States.
- Flash steam power plants are the most common and use geothermal reservoirs of water with temperatures greater than 360°F (182°C). This very hot water flows up through wells in the ground under its own pressure. As it flows upward, the pressure decreases and some of the hot water boils into steam. The steam is then separated from the water and used to power a turbine/generator. Any leftover water and condensed steam are injected back into the reservoir, making this a sustainable resource.
- Binary cycle power plants operate on water at lower temperatures of about 225–360°F (107–182°C). Binary cycle plants use the heat from the hot water to boil a working fluid, usually an organic compound with a low boiling point. The working fluid is vaporized in a heat exchanger and used to turn a turbine. The water is then injected back into the ground to be reheated. The water and the working fluid are kept separated during the whole process, so there are little or no air emissions.

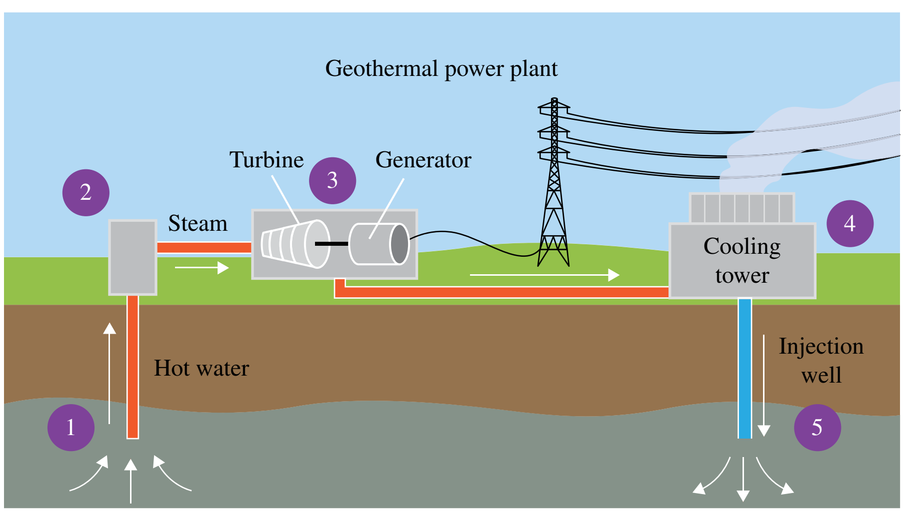


FIGURE 30.11 Diagram of a geothermal power plant. From www.epa.gov.

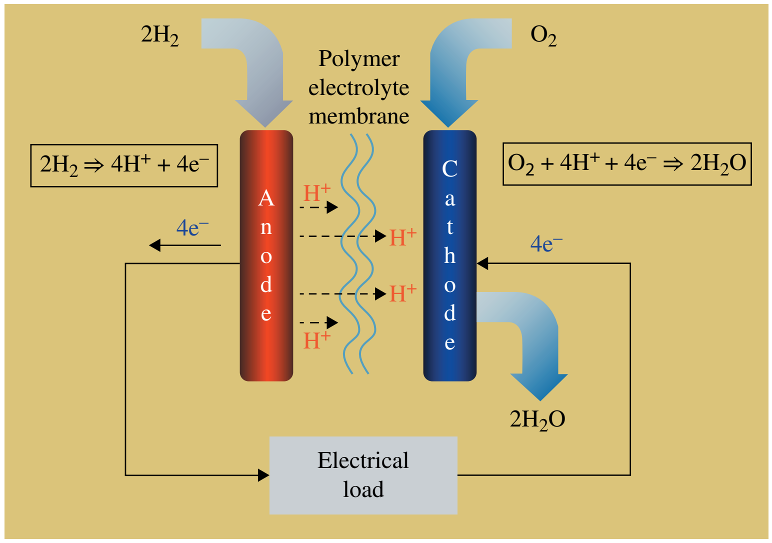


FIGURE 30.12 Block diagram of a fuel cell. Diagram from EPA, http://www.epa.gov/fuelcell/basicinfo.htm#background.

Currently, two types of geothermal resources can be used in binary cycle power plants to generate electricity—EGS and low-temperature or coproduced resources:

- EGS provide geothermal power by tapping into the Earth's deep geothermal resources that are otherwise not economical due to lack of water, location, or rock type.
- Low-temperature and coproduced geothermal resources are typically found at temperatures of 300°F (150°C) or less. Some low-temperature resources can be harnessed to generate electricity using binary cycle technology. Coproduced hot water is a by-product of oil and gas wells in the United States. This hot water is being examined for its potential to produce electricity, helping to lower greenhouse gas emissions and extend the life of oil and gas fields.

While the construction of geothermal plants is very costly, creating electricity from that point onward is quite cheap. The main disadvantage of geothermal power is that it's only viable in limited geographical areas due to the depth of the hot rocks and the availability of water.

30.4 ALTERNATIVE ENERGY: FUEL CELL

The definition of alternative energy varies. In this chapter, we consider: "Alternative energy is any form of energy that does not come from fossil fuels. Alternative energy sources are often renewable, such as solar power and wind power. Alternative energy supplies are clean.[5]"

[5]http://www.universetoday.com/74599/what-is-alternative-energy/

The intermittency of sun and wind power and the geographical limitations of hydroelectric and geothermal power require another 24/7 baseload power alternative, such as fuel cell, to current combustion-based plants.

30.4.1 Fuel Cell Design

A fuel cell is a device that converts the chemical energy from a fuel into electricity through a chemical reaction with oxygen or another oxidizing agent. Hydrogen is the most common fuel, but hydrocarbons such as natural gas and alcohols like methanol are sometimes also used. Fuel cells are different from batteries in that they require a constant source of fuel and oxygen to run, but they can produce electricity continually for as long as fuel and air are supplied.

Fuel cells are made up of three adjacent segments: the anode (negative side), the electrolyte, and the cathode (positive side) (see a fuel cell block diagram in Fig. 30.12). Two chemical reactions occur at the interfaces of the three different segments. The net result of the two reactions is that fuel is consumed, water or carbon dioxide is created, and an electric current is created, which can be used to power electrical devices, normally referred to as the load. Individual fuel cells produce very small amounts of electricity, about 0.7 V, so cells are "stacked," or placed in series, to increase the voltage and placed in parallel circuits to increase the current output to meet an application's power generation requirements. In addition to electricity, fuel cells produce water, heat, and, depending on the fuel source, very small amounts of nitrogen dioxide and other emissions. The energy efficiency of a fuel cell is generally between 40 and 60% or up to 85% efficient if waste heat is captured and used in a combined heat and power (CHP) system.

30.4.2 Fuel Cell Technology Benefits

Some of the most salient benefits of fuel cells are as follows:

- Fuel cell technology has lower CO_2 emissions.
- Fuel cell technology uses less water in the creation of electricity.
- Distributed generation (DG)—fuel cells can generate power "on-site," eliminating the need for transmission over long distances. They avoid 7–15% losses from transmission across the grid.

30.4.3 Fuel Cell Types

The main difference between fuel cell types is the electrolyte; therefore, fuel cells are classified by the type of electrolyte they use. At the anode, a catalyst oxidizes the fuel, usually hydrogen, turning the fuel into a positively charged ion and a negatively charged electron. The electrolyte is a substance specifically designed so ions can pass through it, but the electrons can't. The freed electrons travel through a wire creating the electric current. The ions travel through the electrolyte to the cathode. Once reaching the cathode, the ions are reunited with the electrons and the two react with a third chemical, usually oxygen, to create water or carbon dioxide.

The most important design features in a fuel cell[6] are as follows:

- The electrolyte substance usually defines the *type* of fuel cell.
- The fuel that is used: the most common fuel is hydrogen.
- The anode catalyst, which breaks down the fuel into electrons and ions. The anode catalyst is usually made up of very fine platinum powder.
- The cathode catalyst, which turns the ions into waste chemicals, like water or carbon dioxide. The cathode catalyst is often made up of nickel but it can also be a nanomaterial-based catalyst.

We will cover five major fuel cell types. The first to be fired into space was the polymer electrolyte membrane (PEM) fuel cell, which was developed by GE and performed successfully on the Gemini orbital missions of the mid-1960s.

30.4.3.1 PEM or Proton Exchange Membrane Fuel Cells These fuel cell types incorporate a solid polymer membrane as its electrolyte (Fig. 30.13). The solid, flexible electrolyte will not leak or crack, and these cells operate at a low enough temperature to make them suitable for homes and cars. But their fuels must be purified, and a platinum catalyst is used on both sides of the membrane, raising costs.

[6]http://www1.eere.energy.gov/hydrogenandfuelcells/fuelcells/fc_types.html#phosphoric

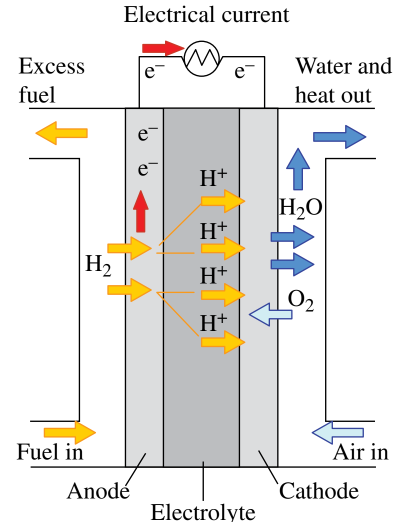


FIGURE 30.13 Diagram of PEM fuel cell. Photo courtesy of EERE, Department of Energy.

Protons (H+) are transported from the anode to the cathode. The operating temperature range is generally 60–100°C.

Today, PEM is the main type being commercialized to power automobiles. An advantage for PEM is that it begins generating power at room temperature and attains its peak power at about 80°C (176°F), allowing the relatively fast start-up needed for cars. And it responds almost instantaneously to changing power demands, which is crucial for transportation.

30.4.3.2 Alkali Fuel Cells Alkali fuel cells (AFCs) (Fig. 30.14) consume hydrogen and pure oxygen, producing potable water, heat, and electricity. They are among the most efficient fuel cells, having the potential to reach 70%.

AFCs operate on compressed hydrogen and oxygen. They generally use a solution of potassium hydroxide (chemically, KOH) in water as their electrolyte. Efficiency is about 70%, and operating temperature is 150–200°C (about 300–400°F). Cell output ranges from 300 W to 5 kW. Alkali cells were used in Apollo spacecraft to provide both electricity and drinking water. They require pure hydrogen fuel, however, and their platinum electrode catalysts are expensive, and like any container filled with liquid, they can leak.

30.4.3.3 Phosphoric Acid Fuel Cell The electrolyte in this type of fuel cell consists of concentrated phosphoric acid (H_3PO_4) (Fig. 30.15). Protons (H+) are transported from the anode to the cathode. The operating temperature range is generally 160–220°C. Efficiency ranges from 40 to 80%, and operating temperature is between 150 and 200°C (about 300–400°F). Existing phosphoric acid cells have outputs up to

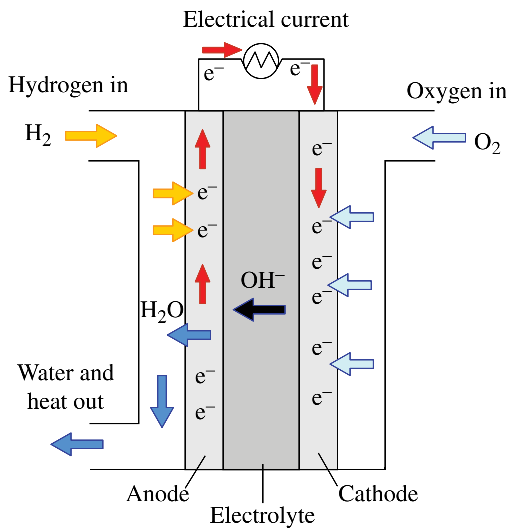


FIGURE 30.14 Alkaline fuel cell. Photo courtesy of EERE, Department of Energy.

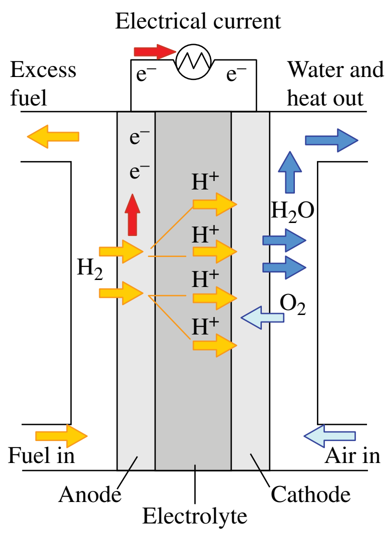


FIGURE 30.15 Diagram of phosphoric acid fuel cell. Photo courtesy of EERE, Department of Energy.

200 kW, and 11 MW units have been tested. Phosphoric acid fuel cells (PAFCs) tolerate a carbon monoxide concentration of about 1.5%, which broadens the choice of fuels they can use. If gasoline is used, the sulfur must be removed. Platinum electrode catalysts are needed, and internal parts must be able to withstand the corrosive acid.

30.4.3.4 Molten Carbonate Fuel Cells Molten carbonate fuel cells (MCFCs) use high-temperature compounds of salt (like sodium or magnesium) carbonates (chemically, CO_3) as the electrolyte. Carbonate ions (CO_3^{-2}) are transported from the cathode to the anode (Fig. 30.16). Operating temperatures

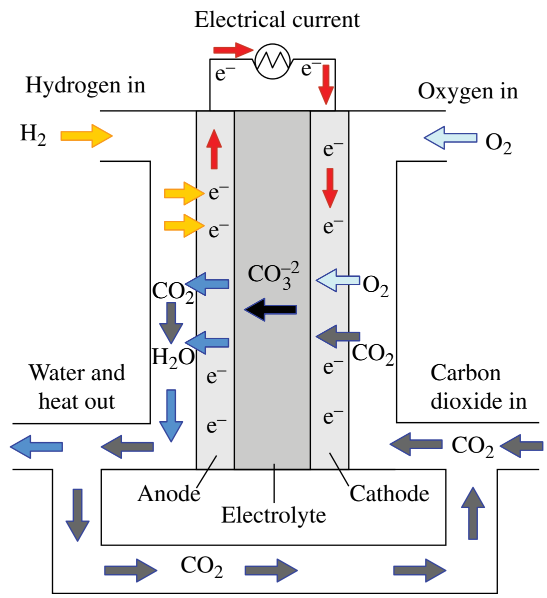


FIGURE 30.16 Molten carbonate fuel cell. Photo courtesy of EERE, Department of Energy.

are typically near 650°C. Efficiency ranges from 60 to 80%, and operating temperature is about 650°C (1200°F). Units with output up to 2 MW have been constructed, and designs exist for units up to 100 MW. The high temperature limits damage from carbon monoxide "poisoning" of the cell and waste heat can be recycled to make additional electricity. Their nickel electrode catalysts are inexpensive compared to the platinum used in other cells. But the high temperature also limits the materials and safe uses of MCFCs—they would probably be too hot for home use. Also, carbonate ions from the electrolyte are used up in the reactions, making it necessary to inject carbon dioxide to compensate.

The MCFC uses an inexpensive catalyst, has high efficiency, and produces excess heat that can be captured and utilized. It can run not only on natural gas and propane but even on diesel fuel, which makes it suitable for ships and stationary power in remote places, such as islands, where delivering a supply of natural gas is difficult or impossible.

The chemical reactions for an MCFC system can be expressed as follows:

$$\text{Anode reaction}: CO_3^{-2} + H_2 \rightarrow H_2O + CO_2 + 2e^-$$
$$\text{Cathode reaction}: CO_2 + ½O_2 + 2e^- \rightarrow CO_3^{-2}$$
$$\text{Overall cell reaction}: H_2 + ½O_2 \rightarrow H_2O$$

30.4.3.5 Solid Oxide Fuel Cell A solid oxide fuel cell (SOFC) is so named because of the solid ceramic material at the center of the device. While solid electrolytes cannot leak, they can crack. SOFCs use a hard, ceramic compound of metal (like calcium or zirconium) oxides (chemically, O_2) as

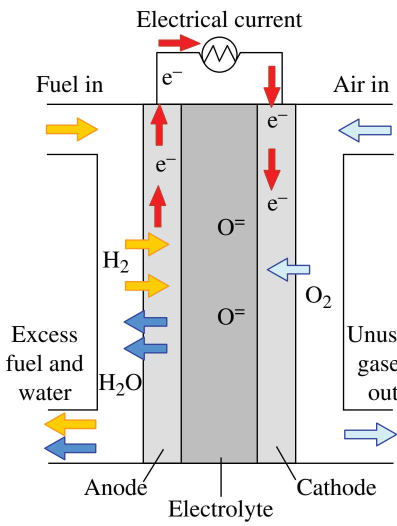


FIGURE 30.17 Solid oxide fuel cell. Photo courtesy of EERE, Department of Energy.

electrolyte. Efficiency is about 60%, and operating temperature is about 1000°C (about 1800°F). Cells output is up to 100 kW. At such high temperatures, a reformer is not required to extract hydrogen from the fuel, and waste heat can be recycled to make additional electricity. However, the high temperature limits applications of SOFC units and they tend to be rather large.

Air enters the cathode side of the cell (Fig. 30.17). At the cathode, oxygen in the air is converted (reduced) to oxide ions, which cross the ceramic interior to the anode. On the anode side, the fuel is electrochemically oxidized producing both heat and electrical energy.

If the fuel comprised only hydrogen, water would be the only emission. If a fossil fuel is used, containing carbon as well as hydrogen, carbon dioxide is formed at, and expelled from, the anode in addition to water. The electrical energy (electrons) produced during the oxidation of the fuel flows through an external circuit, doing some useful work along the way, to the cathode where it is used to convert oxygen to oxide ions, thus completing the circuit. The chemical reactions for the SOFC system can be expressed as follows:

$$\text{Anode reaction}: 2H_2 + 2O^{-2} \rightarrow 2H_2O + 4e^-$$
$$\text{Cathode reaction}: O_2 + 4e^- \rightarrow 2O^{-2}$$
$$\text{Overall cell reaction}: 2H_2 + O_2 \rightarrow 2H_2O$$

SOFCs operate at high temperatures (600–800°C) and can thus tolerate many of the fuel components that poison proton exchange membrane (hydrogen) fuel cells.

SOFCs cogenerate electricity and useful high-temperature heat at efficiencies of >85%.

SOFCs do not produce the sulfur/nitrogen oxides and particulates formed by burning fossil fuels. They can operate on natural gas, propane, and diesel. The fuel flexibility of SOFCS also allows operation on emerging fuels such as biofuels, coal syngas, and pure and impure hydrogen.

The waste stream from SOFCs operating on hydrocarbon fuels contains primarily humidified carbon dioxide (CO_2), and thus SOFCs can serve as an excellent CO_2 capture technology, further reducing greenhouse gas emissions.

30.4.4 Comparison of Fuel Cell Technologies

The table in Figure 30.18 compares the five types of fuel cells described earlier in terms of the electrolyte used, operating temperature, typical stack size, efficiency, advantages, disadvantages, and applications.

30.4.5 Fuel Cell Technology Challenges

Cost and durability are the major challenges to fuel cell commercialization. Size, weight, and thermal and water management are barriers to the commercialization of fuel cell technology. The key challenges include the following:

Cost: The cost of fuel cell power systems must be reduced before they can be competitive with conventional technologies. For stationary systems, the acceptable price point is considerably higher ($400–$750/kW for widespread commercialization and as much as $1000/kW for initial applications).

Durability and reliability: The durability of fuel cell systems has not been established. For stationary applications, more than 40,000 h of reliable operation in a temperature at −35 to 40°C will be required for market acceptance.

System size: The size and weight of current fuel cell systems must be further reduced to meet the packaging requirements for automobiles. This applies not only to the fuel cell stack but also to the ancillary components and major subsystems (i.e., fuel processor, compressor/expander, and sensors) making up the balance of power system.

Air, thermal, and water management: Air management for fuel cell systems is a challenge because today's compressor technologies are not suitable for automotive fuel cell applications. In addition, thermal and water management for fuel cells are issues because the small difference between the operating and ambient temperatures necessitates large heat exchangers.

Improved heat recovery systems: The low operating temperature of PEM fuel cells limits the amount of heat that can be effectively utilized in CHP applications. Technologies need to be developed that will allow higher operating temperatures and/or more-effective heat recovery systems and improved system designs that will enable CHP efficiencies

Comparison of fuel cell technologies

Fuel cell type	Common electrolyte	Operating temperature	Typical stack size	Efficiency	Applications	Advantages	Disadvantages
Polymer electrolyte membrane (PEM)	Perfluoro sulfonic acid	50–100°C 122–212° typically 80°C	<1–100kW	60% transportation 35% stationary	• Backup power • Portable power • Distributed generation • Transporation • Specialty vehicles	• Solid electrolyte reduces corrosion & electrolyte management problems • Low temperature • Quick start-up	• Expensive catalysts • Sensitive to fuel impurities • Low temperature waste heat
Alkaline (AFC)	Aqueous solution of potassium hydroxide soaked in a matrix	90–100°C 194–212°F	10–100 kW	60%	• Military • Space	• Cathode reaction faster in alkaline electrolyte, leads to high performance • Low-cost components	• Sensitive to CO_2 in fuel and air • Electrolyte management
Phosphoric acid (PAFC)	Phosphoric acid soaked in a matrix	150–200°C 302–392°F	400 kW 100 kW module	40%	• Distributed generation	• Higher temperature enables CHP • Increased tolerance to fuel impurities	• Pt catalyst • Long start up time • Low current and power
Molten carbonate (MCFC)	Solution of lithium, sodium, and/ or potassium carbonates, soaked in a matrix	600–700°C 1112–1292°F	300 kW to 3 MW 300 kW module	45–50%	• Electric utility • Distributed generation	• High efficiency • Fuel flexibility • Can use a variety of catalysts • Suitable for CHP	• High temperature corrosion and breakdown of cell components • Long start up time • Low power density
Solid oxide (SOFC)	Yttria stabilized zirconia	700–1000°C 1202–1832°F	1 kW to 2 MW	60%	• Auxiliary power • Electric utility • Distributed generation	• High efficiency • Fuel flexibility • Can use a variety of catalysts • Solid electrolyte • Suitable for CHP & CHHP • Hybrid/GT cycle	• High-temperature corrosion and breakdown of cell components • High-temperature operation requires long start up time and limits

FIGURE 30.18 Fuel cells comparison table. From U.S. Department of Energy, Fuel Cell Technologies Program.

exceeding 80%. Technologies that allow cooling to be provided from the low heat rejected from stationary fuel cell systems also need to be evaluated.

30.4.6 Fuel Cell Manufacturers and Users

Given the large power consumption needs from data centers, it is not at all surprising that many fuel cell companies are now targeting their fuel cell line products at data center operators as a new market for distributed cleaner power. Their new plans are to sell the fuel cells as primary power to data center operators, and the grid would be the backup power to their fuel cells. That's a contrast to some previous cases, where data center operators were using fuel cells for backup and auxiliary power.

Fuel cell products look like industrial refrigerators. They offer a solution for businesses seeking to reduce energy costs, increase energy security and lower carbon emissions. Their claim is that they can cut utility bills by up to 50%, provide continuous power that stays up even when the grid goes down, reduce emission by 41%, and can produce 11 times more energy than an equivalent solar installation while taking up only 1/20th of the surface area.

Many companies have been using fuel cells in their data centers or facilities, including Apple, AT&T, eBay, Facebook, Google, NTT America, Samsung, Sprint, Verizon, etc. Readers may check "Fuel Cells 2000" (http://www.fuelcells.org/) or other sources for lists of fuel cell manufacturers.

30.5 CASE STUDIES

Greenpeace[7] is a nongovernmental environmental organization. Greenpeace states its goal is to "ensure the ability of the Earth to nurture life in all its diversity" and focuses its campaigning on worldwide issues such as global warming, deforestation, overfishing, commercial whaling, and antinuclear issues.

Although all hi-tech companies have made extraordinary gains in overall data center efficiency, Greenpeace has maintained a singular focus on the use of renewable energy, either through on-site generation or by choosing locations where electricity is sourced from renewable energy sources.

Greenpeace recently released a report that profiles the energy usage of the facilities powering major Internet services. The Greenpeace's report, "How clean is your cloud?[8]" looks at the energy choices that power data centers and cloud computing.

[7]http://www.greenpeace.org/usa/en/
[8]http://www.greenpeace.org/international/Global/international/publications/climate/2012/iCoal/HowCleanisYourCloud.pdf

In this section, it presents hi-tech companies that are going green by adopting clean technologies in their new data centers at the time of writing.

30.5.1 Apple Computer

Apple powered its three current data centers with coal-free energy by the end of 2013.[9] Apple powered its 500,000 ft^2 data center in Maiden, North Carolina, entirely with renewable energy by the end of 2012. Apple will produce about 60% of renewable power for the site from solar and fuel cells. The facility has earned LEED Platinum certification from the U.S. Green Building Council, after being touted as exceptionally energy efficient.

Apple built two solar array installations in Maiden. These sites use high-efficiency solar cells and an advanced solar tracking system. A 100 acre, 20 MW on-site installation will produce 42 million kWh of energy annually. Apple calls the 20 MW solar project "the nation's largest end user-owned, onsite solar array." A second 100-acre site located a few miles away will produce another 42 million kWh. Together, that's 83 million kWh of clean, renewable energy supplied annually. A biogas-powered (gas captured from decomposing biomass) 5 MW fuel cell installation came online later in 2012, which provided more than 40 million kWh of 24×7 baseload renewable energy annually. This means Apple will be producing enough on-site renewable energy at 124 million kWh, which is equivalent to power 10,874 residential homes.

Some energy-efficient design elements of Apple's Maiden facility include the following:

- 20 MW of solar panels from San Jose, California-based SunPower Corporation.
- 4.8 MW of fuel cells from Sunnyvale-based Bloom Energy Corporation. The fuel cells will be powered with biogas from landfills. The use of biogas, which displaces conventional natural gas to generate electricity, will reduce greenhouse gas emissions and smog-forming pollutants.
- 200 MW of wind power from local utility grids to lower the carbon footprint from its operations.
- A chilled water storage system to improve chiller efficiency by transferring 10,400 kWh of electricity consumption from peak to off-peak hours each day.
- Use of "free" outside air cooling through a waterside economizer operation during night and cool-weather hours, which, along with water storage, allows the chillers to be turned off more than 75% of time.
- White cool-roof design to provide maximum solar reflectivity.
- High-efficiency LED lighting combined with motion sensors.
- Real-time power monitoring and analytics during operations.

Apple's newest data center, located in Prineville, Oregon, will be every bit as environmentally responsible as their Maiden data center. At Prineville, they have access to enough local renewable energy sources to completely meet the needs of the facility. To achieve that goal, they are working with two local utilities as well as a number of renewable energy generation providers to purchase wind, hydro, and geothermal power—all from local sources.

30.5.2 eBay

eBay's new "data center in Utah will rely on a 6 MW fuel cell array supplied by Bloom Energy, based in Sunnyvale, California, which makes an innovative solid oxide system.[10] It will be the largest stationary fuel cell bank (30 Bloom cells) ever installed in a nonutility setting, and the first time a data center has been designed to rely on fuel cells as its primary energy source, with the grid serving as backup. The normal procedure is for data centers to get electricity from the grid, with some kind of backup system to kick in when the grid goes down—an expensive procedure."

"In principle, whether the plant is running exclusively on biogas or biogas production is being subsidized to compensate for natural gas consumed at the plant, the facility would appear to be doubly green: It runs on a renewable fuel and produces no solid waste, carbon dioxide being its only undesirable byproduct. So it's easy to see why the Bloom Energy Server is attractive to high-tech companies that depend on big energy-guzzling data centers and fervently wish to build green credentials."

30.5.3 Google

Google has invested $915 million to date in clean energy development and sited its data centers in Iowa[11] and Oklahoma[12] with long-term wind energy contracts. Google will buy power from a planned 100 MW wind farm in Mayes County, Oklahoma, located near a data center now being built, another step in the company's goal to be carbon neutral.

The power purchase agreement to buy power from Minco II wind farm in Mayes County for 20 years is similar to the one Google signed in 2012 with the project developer, NextEra Energy Resources, a wind farm in Iowa.

[9]https://www.apple.com/environment/renewable-energy/

[10]http://spectrum.ieee.org/energywise/green-tech/fuel-cells/ebay-will-rely-on-fuel-cells-to-power-major-data-center-

[11]http://www.google.com/about/datacenters/inside/locations/council-bluffs/

[12]http://www.google.com/about/datacenters/inside/locations/mayes-county/

The deal follows Google's investment in two large renewable energy projects, including a huge 825 MW wind farm in Oregon[13] and the Ivanpah solar power plant[14] in Southern California, both of which are under construction.

Google touts the energy efficiency numbers of its facilities, Google's "TTM energy-weighted average PUE" could be found at "Efficiency: How we do it."[15]

30.5.4 IBM

IBM's India Software Lab[16] in Bangalore has set up a 50 kW rooftop array to power about 20% of its data center.

When IBM considered to run servers on high-voltage direct current, they determined to use solar panels, which produce direct current, as a source. This direct current mini-grid solution can cut energy consumption of data centers by about 10% due to alternating current–direct current conversion loss.

The system is designed so that power will be pulled from the grid at night or when there isn't sufficient voltage to run servers directly. Power conditioning units dedicated to supplying the data center can automatically switch between power sources.

30.5.5 Microsoft

Microsoft Corporation is relying on green technologies in its "the most power-efficient" data center. Their San Antonio, Texas, data center will cost \$550 million and 477,000 ft^2 and contain tens of thousands of servers.[17] Microsoft plans to use 602,000 gallons of recycled water (from San Antonio's waste water system) a day in its cooling systems. The recyclable water makes use of water that is not fresh or drinkable but is not contaminated by any toxic substances from the local utility. It is considered environment-friendly because it reduces demands for freshwater and doesn't consume the energy required to purify it at wastewater treatment sites. In addition, a significant portion of electricity in Texas is generated by wind and abundant sunlight with solar panels, and that clean source of energy was attractive to site selection by Microsoft.

30.5.6 Yahoo

Yahoo unveiled its new data center in Lockport, New York, boasting one of the most energy-efficient data centers in the world. The company says the new data center will use significantly less energy and water than conventional data centers. Yahoo says "with a low Power Usage Effectiveness (PUE) for the facility is 1.08, compared to the industry average of 1.92."[18] The data center will be powered in part by hydropower and reduce energy costs to "less one cent for cooling for every dollar spent on electricity." Yahoo designed its own data center, choosing a location with cool weather and winds. Its "chicken coop" design is inspired by the long, narrow architecture of chicken coops. The result: air moves through the building naturally and cools the servers without the usual need to crank up the air conditioning and in turn, high electricity bills.

Environment Control: Greenpeace International focuses on environmental impact, and use of renewable energy.

30.6 SUMMARY AND FUTURE TRENDS

This chapter provided a summary of renewable energy technologies, focusing primarily on those used to power data centers (solar, wind, hydropower, and geothermal). At the same time, to remedy some of renewable energy limitations (solar and wind energy intermittency), fuel cell technology is discussed. Fuel cells provide compact, quiet, and reliable baseload power in DG (see the following text) and on-site power generation.

The following trends are being observed:

Today, most of the electricity produced in the United States is provided by regional utilities and supplied to customers via the grid. DG refers to power generation at the point of consumption rather than centrally. DG eliminates the cost and inefficiencies of transmission and distribution, reduces grid congestion, and provides flexibility. These distinct advantages of DG are changing and users are choosing DG to become self-reliant in terms of their energy needs.

Combined Heat and Power (CHP): Fuel cells allow the use of waste heat generated by the fuel cell to heat the building, thus reducing energy costs. Readers may wish to read "Opportunities for Combined Heat and Power in Data Centers[19]" prepared by ICF International for U.S. Department of Energy.

A number of IT companies are signing long-term power purchase agreements (PPAs) to procure energy from renewable energy systems. Such PPAs help renewable energy developers to obtain preferential financing, and allow customers to purchase energy at set rates, typically below utility electric prices. Companies can sign PPAs for on-site or off-site renewable solutions. PPAs require that the customer organization has excellent credit, and is willing to sign a long-term contract. Google is an excellent example of a company using PPAs to procure clean energy for new data centers. By signing a

[13] http://www.google.com/about/datacenters/inside/locations/the-dalles/
[14] http://ivanpahsolar.com/about
[15] http://www.google.com/about/datacenters/efficiency/internal/
[16] http://www-03.ibm.com/press/us/en/pressrelease/35891.wss
[17] http://blogs.msdn.com/b/microsoft-green/archive/2008/09/22/microsoft-opens-san-antonio-data-center.aspx
[18] http://yodel.yahoo.com/blogs/yahoo-corporate/yahoo-unveils-world-class-green-data-center-4735.html
[19] https://www1.eere.energy.gov/manufacturing/datacenters/pdfs/chp_data_centers.pdf

long-term PPA, Google[20] has provided NextEra Energy with a secure revenue source, which allows them to obtain financing and helps to stimulate demand for more renewable energy.

Hi-tech companies are electing to purchase renewable energy directly through competitive retail markets, power purchase agreements, and/or renewable energy certificates (RECs). RECs represent the environmental attributes of the generation and delivery of 1 MW/h of green power to the U.S. Grid. RECs have become a popular option for easily and inexpensively offsetting emissions from data center electricity use. Leading hi-tech companies including Intel, Microsoft, Cisco, and Dell are among the top purchasers of RECs. RECs can be sourced locally or nationally, meaning that there may or may not be local environmental benefits from the purchase of RECs.

Finally, the goal of using renewable energy is to minimize environment impacts due to the energy requirement to remove heat from data centers. Efficiently running servers—the heat source—will result in efficient utilization of power. Applying a "green algorithm" theorem such as described in Ref. [5], which determines the optimal speed for all tasks assigned to a computer, will efficiently operate cloud servers with adjustable speeds and parameters to effectively reduce energy consumption and complete all tasks.

REFERENCES

[1] Boyle G. *Renewable Energy*. Oxford: Oxford University Press; 2004.

[2] Twidell J, Weir T. *Renewable Energy Resources*. 2nd ed. London: Taylor & Francis; 2006.

[3] National Renewable Energy Laboratory. Available at www.nrel.gov. Accessed on May 22, 2014.

[4] Renewable Energy World. Available at www.renewableenergyworld.com. Accessed on May 22, 2014.

[5] Zhang LM, Li K, Dan Chia-Tien Lo, Zhang Y. Energy-efficient task scheduling algorithms on heterogeneous computers with continuous and discrete speeds. Sustainable Computing: Informatics and Systems. 3(2), 109–118; 2010.

FURTHER READING

Bloom Energy. Available at www.bloomenergy.com. Accessed on May 22, 2014.

Energy Department. *Annual Energy Outlook 2013 with Projections to 2040*. U.S. Energy Information Administration; 2013. Available at http://www.eia.gov/forecasts/aeo/pdf/0383(2013).pdf. Accessed on May 22, 2014.

Fuel cells articles. Available at www.eoearth.org/article/Fuel_cells. Accessed on May 22, 2014.

Green Data Center. Available at http://www.42u.com/green-data-center.htm. Accessed on May 22, 2014.

Kishore VV. *Renewable Energy Engineering and Technology*. Sterling: Earthscan; 2009.

Levelized Cost of New Generation Resources in the Annual Energy Outlook 2013. U.S. Energy Information Administration; 2013. Available at http://www.eia.gov/forecasts/aeo/pdf/electricity_generation.pdf. Accessed on May 22, 2014.

[20] https://static.googleusercontent.com/media/www.google.com/en/us/green/pdfs/renewable-energy.pdf

31

SMART GRID-RESPONSIVE DATA CENTERS

Girish Ghatikar, Mary Ann Piette, and Venkata Vish Ganti

Lawrence Berkeley National Laboratory, Berkeley, CA, USA

31.1 INTRODUCTION AND CONTEXT FOR GRID-RESPONSIVE DATA CENTERS

Since 2008, the Industrial Demand Response (DR) Team of the Demand Response Research Center (DRRC) at Lawrence Berkeley National Laboratory (LBNL) has been evaluating DR opportunities in industrial facilities [1] and their ability to be grid responsive. This initial research included collecting and analyzing data on recommended DR strategies included in utility integrated audits and evaluating the applicability of these strategies for use in automated demand response programs (known as AutoDR). These programs use OpenADR, a national Smart Grid standard for DR and distributed energy resources. OpenADR refers to the use of an open standard for communicating DR prices and signals, which allows utilities or energy service providers to send common signals to facilities [2]. The facility controls are preprogrammed to respond to these signals. The team supported a number of California electric utilities and their contractors in identifying potential automated industrial DR participants and provided technical assistance in evaluating DR sites. The team also conducted in-depth analyses of industrial sectors that appeared to have good AutoDR potential and analyzed their DR technical capacity. In 2008, the DRRC selected data center facilities as a focus for new research because of their high and increasing energy use. Data center energy use is expanding rapidly in California and nationally. In the Pacific Gas and Electric (PG&E) service territory alone, data centers are estimated to consume 500 MW of peak electricity annually [3].

According to a 2007 U.S. Environmental Protection Agency (EPA) report, the national energy consumption by servers and data centers doubled from 2000 to 2006 to 61 billion kWh; and if such current trends continue, we will nearly double again to more than 100 billion kWh by 2011. With an estimated annual electricity cost of $7.4 billion, an estimated 20% of the energy use is in the Pacific region alone [3]. A recent study has shown that there was an increase in energy consumption at data centers by 85.6 billion kWh; however, the recent studies indicate that it was lower than the EPA forecast [4]. The 2008 financial crisis, the resulting global economic slowdown, and further improvements in virtualization technologies leading to a reduced server-installed base were attributed for lowered energy use than the EPA forecast. The EPA's identification of the San Francisco Bay and Los Angeles areas in California as having the largest concentration of data centers in the United States and as "areas of concern" and "critical" for electricity transmission congestion formed the impetus for the LBNL studies for energy efficiency and DR.

31.1.1 What Are Grid-Responsive Data Centers?

With rapid acceleration and investment in Smart Grid deployment in the United States and other parts of the world, one key question remains unanswered—How can customers benefit from the Smart Grid? While the value to the customer is often not well defined, few initial studies have looked at the valuation framework through DR [5]. Such metrics are a good starting point for data center customers' integration with the Smart Grid. DR is a set of actions taken to reduce electrical loads when contingencies, such as power grid emergencies or congestion, threaten the electricity supply–demand balance and/or market conditions that cause the cost of electricity to increase. DR programs and tariffs are designed to improve grid reliability and decrease electricity

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

use during peak demand periods, which reduces total system costs [6–8]. The term "Smart Grid-Responsive" (or Grid-Responsive in short) data centers indicates that facilities are not only "self-aware" to meet local needs but also "grid aware" to respond to changing grid conditions (e.g., price or reliability) and gain additional benefits resulting from incentives and credits and/or lowered electricity prices.

31.1.2 Smart Grid and DR Role

For the U.S. Smart Grid framework developed by the National Institute of Standards and Technology (NIST), the Smart Grid is defined as a "complex system of systems for which a common understanding of its major building blocks and how they interrelate must be broadly shared" [9, 10].

The NIST framework provides a high-level conceptual reference model for Smart Grid domains with secure communication and electrical interfaces, including the integration with the customer's facilities. The framework also provides the relevance of hardware and software technologies used in the Smart Grid. Figure 31.1 shows these domains and communication models, including the DR and OpenADR interfaces. This framework formed the basis to identify interoperability standards to facilitate communication between different Smart Grid domains and their relevant security measures.

In many instances, analogy is made between smart meters and the Smart Grid. The Smart Meter acts as one of the Energy Services Interfaces (ESIs) that are used as a demarcation point between the Smart Grid and the customer domain. As per NIST framework, the Smart Grid scope is outside of the meter or ESI. Other ESIs can exist within the customer domain such as the energy management and control systems (EMCSs) and IT equipment management tools. The DR scope is within the service provider, operations, and the customer domains, which is shown as OpenADR DR signals in Figure 31.1. However, other Smart Grid domains influence the need for DR.

31.1.3 Study Objectives

The Smart Grid relevant LBNL studies evaluated the technical and institutional capabilities and opportunities as well as challenges and unique issues related to DR in data

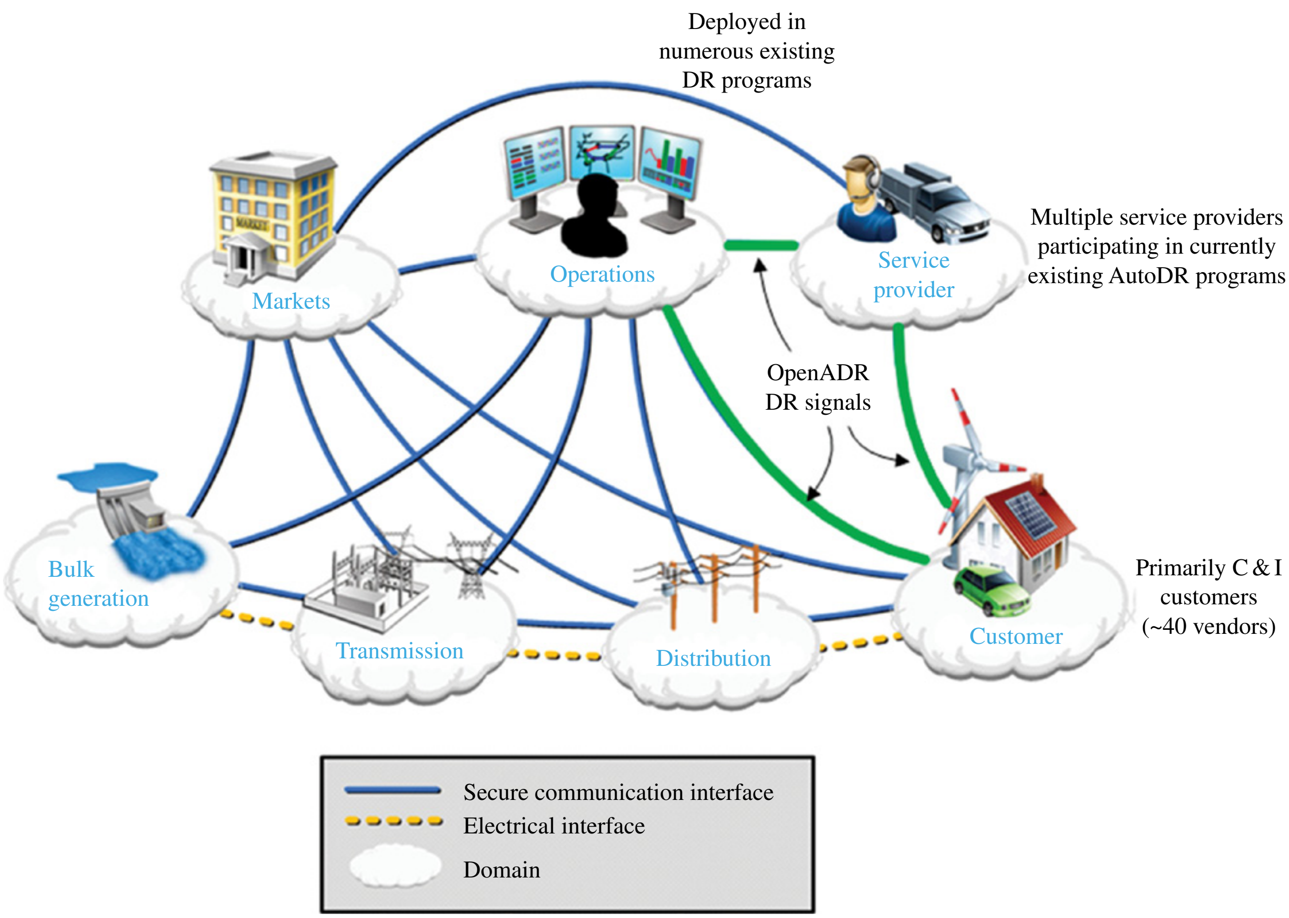


FIGURE 31.1 Smart grid domains and actors' interaction through secure electrical and communication flows, including the OpenADR consumer interfaces. NIST Publication 1108 Ref. [9].

centers. The findings from these studies form majority of the content in this chapter. LBNL is conducting studies to evaluate the performance of DR control strategies through a series of data center field tests.[1] Specific project objectives of earlier studies were to:

- Identify different types of existing data centers and data center technologies.
- Determine technologies and strategies that could be used for DR and/or AutoDR using open standards such as OpenADR.
- Identify emerging technologies (e.g., virtualization, load migration, cloud computing, and storage) that could be used for DR and/or OpenADR.
- Verify load patterns and the potential magnitude of load shed or shift in data centers that could be achieved with little or no impact on data center business or operations.
- Assess the readiness of technologies that could be used with the existing OpenADR infrastructure in California utilities.
- Identify concepts and opportunities for providing OpenADR-enabled products to facilitate full automation of data center DR strategies.
- Identify next steps and field study requirements as well as barriers, if any, for data center participation in DR or OpenADR.

The study draws on more than 6 years of previous research and ongoing data center and high-tech building-related energy efficiency projects at LBNL [11]. Previous related work includes benchmarking of data centers; development of best practices and assessment tools for the U.S. Department of Energy (DOE); case studies and demonstrations of energy efficiency; development of a certified practitioner program and a joint American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE)-DOE awareness training curriculum; and studies of uninterruptible power supply (UPS) and DR power efficiency.

31.2 SMART GRID AND DR APPLICATIONS IN THE UNITED STATES

Significant investments by both the government and industry are under way in the United States to create a more efficient electric grid, termed “Smart Grid.” A fully functional Smart Grid provides dynamic optimization of electric grid operations and resources. Such electric grid systems will enable incorporation of DR and consumer participation [12]. There are many definitions of Smart Grid (e.g., modernization of the electricity delivery system so it monitors, protects, and automatically optimizes the operation of its interconnected elements [13]), and the choice of which systems to develop and deploy depends on the contextual and regional needs.

In 2009, U.S. President Obama announced $3.4 billion investment to spur Smart Grid transition, which was followed by over $11 billion investments through American Recovery and Reinvestment Act [14]. Such investments through federal and ARRA funds for Smart Grid have the potential to lead innovation through measured data, visualization, and automation of both facility and grid resources.

Many other countries have followed such developments and have their own Smart Grid programs to meet local needs. For example, in India, the additional focus is on improving electric reliability and better accounting of electric losses. In South Korea, there is a need for Smart Grid to provide DR programs to address price volatility during high-demand periods and large-scale smart meter deployment. In Japan and China, there is a need of Smart Grid to offer DR to address increasing demand and integration of renewable energy systems.

31.2.1 Mission

The goal of the NIST Framework 1.0 and its preceding Framework 2.0 was to develop a roadmap to identify and develop a pathway for interoperability standards that will facilitate easy communication and operation of the Smart Grid domains across different markets. This activity was intended to advance the federal goals and mission through analysis and coordination activities. The key stakeholders provided inputs and advancements of Smart Grid interoperability standards with work in different domains. For the purposes of this chapter, we focus on the consumer, and the service provider and operations interfaces, which is the domain for DR and OpenADR.

31.2.2 Stakeholders

This content of this chapter is of potential interest to the following DR stakeholders:

- Utilities, energy, or DR service providers wishing to identify new DR potential in the data center industry and to create targeted industrial DR programs
- Data center operators wishing to reduce energy costs, explore DR value and strategies, and incorporate energy

[1]The Pacific Gas and Electric Company’s (PG&E) Emerging Technologies Program funded the earlier studies with cofunding from the Public Interest Energy Research (PIER) Program at the California Energy Commission (CEC). The PG&E, CEC, and San Diego Gas and Electric Company (SDG&E) have funded the studies to conduct field tests and analysis.

efficiency or demand-side management measures beyond those already planned or implemented

- Federal and state policy makers and regulators wishing to identify new DR opportunities and review technology availability and maturity as a basis for implementing recommendations for building codes and new construction
- Members of the public wishing to know about utility and data center industry efforts to provide for energy and grid security and reliability
- Product vendors and companies wishing to identify new business opportunities in the energy value chain

Additionally, the NIST framework also defines the Smart Grid stakeholders to include industry utilities, vendors, academia, regulators, system integrators and developers, and others for the decision-making process.

31.2.3 Benefits

Grid-responsive data centers have the potential to provide many benefits through well-coordinated activities with the electric grid. The data center energy use in the United States is growing locally and globally. For example, the data center energy use is significant in local areas such as those of California to around 10% than the national average of 1.5–2%. EPA study findings suggest that in the PG&E service territory alone, data centers represent an estimated 500 MW of peak load (~2.5% of total) and are growing fast. This energy use is increasing rapidly both within and outside of California [3].[2] Concentration of data centers in certain areas of the state will strain electricity distribution and supply systems if current trends continue. The LBNL study was the first comprehensive exploration of data center DR opportunities. Although the emphasis was on impacts of data center DR in California, the findings and recommendations apply to other regions as well. Data center energy use is not only a domestic challenge but also a growing global concern.

The general benefits of DR extend across different Smart Grid domains and to customer facilities such as data centers. For example, reduction of peak power demand by data center facilities results in reduced new generation capacity or peak load power plants. To provide electricity for limited hours in a year, the peak plants often result in increased carbon emissions if the generation source is coal, oil, etc. The resulting cost savings can be passed to the consumers in the form of lowered electric prices and/or incentives, credits for participating in DR programs. The NIST framework defines additional benefits of a modernized electric grid as follows:

- Improves power reliability and quality
- Optimizes facility utilization and averts construction of backup (peak load) power plants
- Enhances capacity and efficiency of existing electric power networks
- Improves resilience to disruption
- Enables predictive maintenance and “self-healing” responses to system disturbances
- Facilitates expanded deployment of renewable energy sources
- Accommodates distributed power sources
- Automates maintenance and operation
- Reduces greenhouse gas emissions by enabling electric vehicles and new power sources
- Reduces oil consumption by reducing the need for inefficient generation during peak periods
- Presents opportunities to improve grid security
- Enables transition to plug-in electric vehicles and new energy storage options
- Increases consumer choice
- Enables new products, services, and markets and consumer access to them

31.2.4 Current Smart Grid and DR Status

As the U.S. Smart Grid interoperability standards, government and industry demonstrations take shape; there are many ARRA projects to demonstrate the technical feasibility and the value to the stakeholders. One such study is LBNL’s DR opportunities for data centers, which formed the significant basis for the contents of this chapter [15].

One of the key success metric of the Smart Grid interoperability standards is the testing and certification framework to enable applications. Several organizations have started to successfully implement guidelines for the implementation of testing and certification programs. The NIST-initiated Smart Grid Interoperability Panel (SGIP) and its subcommittee, the Smart Grid Testing and Certification Committee (SGTCC), created the Interoperability Process Reference Manual (IPRM). The IPRM provides a best practice approach for certification schemas from actual testing to the act of certifying a product itself [16]. In the DR space, the OpenADR standards development is under way to create a testing and certification, including deployment roadmap with the purpose of [17]:

- Creating interoperable standards.
- Testing of conformance and interoperability.
- Certifying the products.

[2]The EPA estimates that in 2006 the energy use of the nation’s servers and data centers was more than double that was used in 2000. Nationally, data center electricity use was estimated to be 61 billion kWh in 2006 (1.5% of total U.S. electricity consumption) for a total electricity cost of about $4.5 billion. PG&E represents about one-third of California’s electricity sales.

Through the NIST framework, the next steps for Smart Grid include getting the consumers and the stakeholders engaged at federal, state, and local levels. The roadmap and activities should eventually lead to benefits identified and a fully functional Smart Gird. The NIST framework is also intended to provide inputs to regulators and policymakers to evaluate the investments proposed by the utilities and other entities.

31.2.5 Data Center Power Distribution and Technologies

This section reviews power distribution and efficiency technologies that are applicable to data centers and could be used for DR. The key technologies for site infrastructure are control and other strategies to reduce cooling energy use; a key efficiency technology for IT infrastructure is virtualization. This study also includes synergistic technologies that integrate IT and site infrastructure efficiency efforts. This review includes both mature and emerging technologies, emphasizing those that can be used for DR and integrated with OpenADR. The primary uses of almost all of the technologies addressed in this section are energy efficiency and operations optimization. For DR, the technologies and control systems would need to allow for open integration with multiple vendors for interoperability and scalability to different data center types.

Figure 31.2 shows a typical data center's power distribution architecture and end uses. Typically, EMCSs regulate site infrastructure loads. Without EMCSs, energy is distributed directly by the switchgear, commonly known as the electricity grid. IT infrastructure comprises electronic components to transform and smooth power so that equipment can safely consume it. In most cases, the IT equipment consumes, on average, nearly half (40–50%), and the site infrastructure consumes the remaining 50–60% of total data center energy. The recent trends, resulting through energy efficiency measures, seem to show significant reduction in the site infrastructure loads, primarily the cooling loads.

31.3 SITE INFRASTRUCTURE CONTROL SYSTEM TECHNOLOGIES

EMCSs primarily regulate data center site infrastructure systems: cooling, power delivery, and lighting. Most current data center cooling systems use fans to push cool air to equipment. Some data centers use efficient direct water refrigerant cooling systems [18].

Several distributed EMCSs are currently on the market and are primarily used for monitoring and implementing energy efficiency measures. Along with supervisory control and data acquisition (SCADA) systems, these automation and control systems regulate operation of the heating ventilation and air-conditioning (HVAC), lighting, and related facility electrical systems in an integrated fashion. Communication building control protocols such as BACnet®, Modbus®, and LonTalk® allow EMCS systems to communicate with site infrastructure equipment. These protocols are important to understand and could be programmed to communicate any efficiency or potential DR strategy and oversee technology interoperability within data centers. In many cases, such EMCS or SCADA systems can be preprogrammed to manage data center support loads in response to a DR event notification.

31.3.1 Cooling, Power Delivery Systems, and Lighting Technologies

As described earlier, IT equipment is the primary data center end use, consuming approximately half of total data center energy. Site infrastructure systems such as cooling, power delivery, and lighting also consume a significant amount of energy, from 35 to 50% of total energy use. Generally, for every watt of consumption by IT equipment, another watt is required for the entire infrastructure. There is a potential for as much as a 15% reduction in cooling system energy use based on the operations of best practice "green" data centers. In a small-sized best practice data center, this amounts to a saving of more than 1 million kWh [19]. Data centers using technologies such as air economizers for cooling systems,

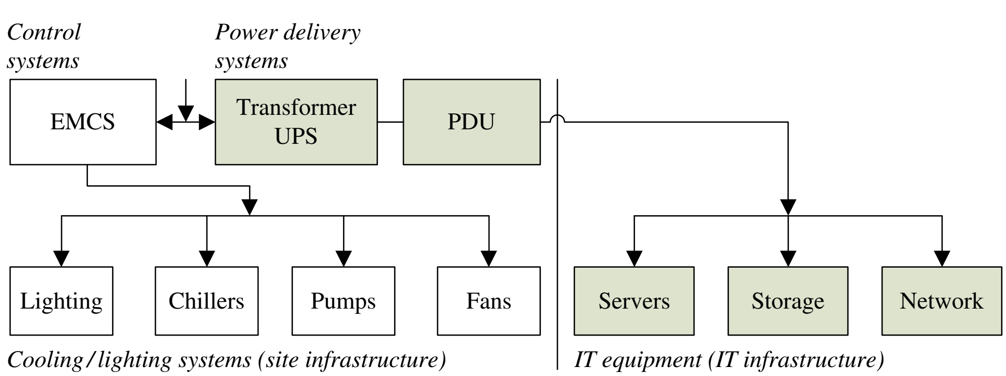


FIGURE 31.2 Typical data center power consumption and distribution architecture. NIST Special Publication 1108R2 Ref. [10].

power loss reduction for power delivery systems, and lighting controls could also use these technologies for DR.

31.3.2 Cooling System Technologies

Air or water economizers could save significant energy and costs for data centers, and it is a likely DR strategy. Air economizers use outdoor air directly to meet indoor cooling needs whenever the outside air temperature is lower than the return air temperature set point. In one LBNL study of air economizers, mechanical cooling power dropped by approximately 30% when economizers were active, which saved significant energy costs [20]. Other energy efficiency cooling system technologies that could be used for DR are:

- Raising data center temperature set points and improving airflow management.
- Regulating humidification controls or eliminating them completely in areas where humidification is not necessary (e.g., temperate climates such as in California).

31.3.3 Power Delivery System Technologies

Some recent technological advancements to cut losses resulting from power distribution systems such as UPSs, transformers, and using UPS bypass can increase the overall efficiency of data centers. Almost all data centers use backup storage system technologies and standby generators for power interruptions, emergencies, and disaster recovery. Such power delivery system technologies could be useful for DR.

31.3.4 Lighting Control Technologies

Bilevel and dimmable lighting controls use sensors to automatically regulate lighting use as needed. Because lighting accounts for a small portion of data center energy use, except in mixed-use data centers, the magnitude of savings from lighting controls and more efficient lighting is smaller than for cooling and power delivery system efficiency measures.

31.4 IT INFRASTRUCTURE VIRTUALIZATION TECHNOLOGIES

Future data center cost management will rely on reducing IT equipment energy consumption, which, in turn, reduces site infrastructure energy use [21]. Virtualization technologies consolidate and optimize servers, storage, and network devices in real time, reducing energy use by enabling optimal use of existing data center equipment. The business and operational needs, the Service-Level Agreements (SLAs), and the energy management goals guide the use of virtualization technologies. Virtualization technologies are increasingly being used not only to improve energy efficiency but also to reduce the expensive floor space required for IT equipment and to manage and optimize legacy systems in real time. Virtualization allows data centers to:

- Optimize use of existing servers, storage, and network devices based on business needs.
- Reduce electricity and new hardware/software commissioning costs.
- Consolidate for improved energy efficiency of IT equipment.
- Manage bandwidth requirements, power constraints, and time-differentiated rates.

Server power supply efficiencies vary dramatically by load, with peak efficiency at 50–60% loads, high efficiency at high loads, and significantly lower efficiency at loads of less than 30%. Most server power supplies operate at 20–50% of load, and power supplies are often oversized for equipment requirements, leading to inefficient power use and excess heat [22]. Virtualization technologies increase server power efficiency and reduce cooling loads by eliminating redundant IT equipment, and are mature enough to meet performance and reliability needs of data centers without compromising quality of service.

An example is the Simple Network Management Protocol (SNMP) in data centers that allows IT equipment to communicate and to be used with virtualization technologies. Other communication protocols and languages such as Transmission Control Protocol over Internet Protocol (TCP/IP) and eXtensible Markup Language (XML) would enable open, standard-based information exchange within a data center's virtualization network and interoperability as well as integration with Smart Grid. Technologies that integrate site and IT infrastructure would be useful to provide a single source of information for integrated implementation of DR strategies.

31.5 DR OPPORTUNITIES, CHALLENGES, AND AUTOMATION CONSIDERATIONS

Sizeable potential DR opportunities exist in data centers, and numerous strategies are available for data centers. The DR strategies that pertain to a given data center will depend on operational and functional (or type) characteristics of the data center, and these characteristics provide decision support for DR program participation. Data center managers may perceive that some strategies are applicable for energy efficiency; however, raising the bar and temporarily reducing service levels without impact to operations can achieve further incremental benefits through DR. These DR strategies generally fall into the categories of load shedding (dropping load completely) and load shifting (moving load from peak to off-peak periods). We separated the data center DR opportunities according to the area of the data center facility to which they apply:

- *Site infrastructure*, where the opportunities in HVAC, power delivery, and lighting have been well studied and, for example, include raising temperature set points
- *IT infrastructure*, where the main opportunities are in virtualization and other emerging technologies for servers, storage, and networking equipment, enabling, for example, consolidation of redundant server

The following subsections summarize the main DR data center opportunities and strategies and the advantages and challenges of each strategy. The opportunities include strategies that are currently research concepts—emerging technologies that are still under development. Each of the strategies listed in Table 31.1 is described in further detail in the subsections following the table.

In addition to opportunities, we look at key challenges to implement DR in data centers, which include traditional conservative operating strategies, the need for performance evaluation metrics that support DR performance assessment, and lack of information about DR in data centers and resulting perception of the risks of DR (Table 31.1). For data centers that are not already employing energy efficiency strategies, DR may be a manageable first step toward energy efficiency practices, enabling the data centers to save energy and gain financial benefits.

This characterization of data center DR opportunities and challenges is based on the LBNL scoping research study. Through field tests, further research is under way in LBNL to assess the opportunities for wide-scale adoption of some of these DR strategies and their feasibility for automation using open standards such as OpenADR, which is also described further.

31.6 DATA CENTERS WITH DR PROVISIONS

Site infrastructure DR opportunities include strategies for load sheds or shifts through changes in cooling and lighting energy use. IT infrastructure opportunities include the use of virtualization strategies to consolidate redundant servers and storage and improve the efficiency of networks and tasks such as routine backups.

31.6.1 Demand-Response Strategies for Site Infrastructure

Data center site infrastructure loads (cooling, lighting, power delivery) support data center IT infrastructure. The site infrastructure end uses and control systems are similar to those in commercial buildings although in data centers these systems serve IT equipment needs rather than human comfort needs. DR strategies for cooling and lighting control systems could apply only to site infrastructure or could be designed to respond to IT equipment energy use. For mixed-use data centers that contain large office areas, extensive studies have looked at optimal DR strategies for HVAC and lighting [23]. These studies found that HVAC and lighting are excellent candidates for DR, which achieves significant peak load reduction with no impact to occupants or facility operations and may apply to data centers as well.

31.6.2 Demand-Response Strategies for IT Infrastructure

Data center IT infrastructure end uses include servers, storage, and network devices, which typically account for half of total data center energy use. Cooling systems protect these devices from failure by eliminating the heat they generate. Any DR strategy for IT infrastructure load will, by definition, reduce cooling load. Virtualization technologies can be used to consolidate redundant servers. Section 31.6.3 details a few virtualization technologies available today. As little or no information exists on applicability to DR, no empirical DR load reduction estimates are provided.

31.6.3 IT and Site Infrastructure Synergy

In 2007, the LBNL team determined that synergistic DR using IT and building control technologies to manage IT and site infrastructure loads together could have significantly greater impact than stand-alone DR in either IT or site infrastructure. The synergy here would enable faster response of site infrastructure loads to the changing IT loads. This determination was consistent with the results of other studies that show greater potential energy savings from integrated building controls. For example, an integrated lighting, HVAC controls, and automated blind system that monitored light levels and temperature can control the building systems to achieve least energy cost [24]. In data centers, intelligent coordination of site infrastructure controls to respond automatically to IT infrastructure load reductions could enable fast and efficient whole-building load reduction. Improved interaction between site and IT management (and technologies) would not only allow for general efficiency improvements but also facilitate the synergy between virtualization or server consolidation for significant IT energy reduction and corresponding reductions in the need for site infrastructure (e.g., cooling) [25]. Current technologies and systems do not provide a platform for integrating IT and site infrastructure. Current market partial solutions provide middleware to bridge this gap.

One key to integrate the energy management of data centers IT and site infrastructure is the use of different communication protocols. Many such solutions from vendors currently available on the market provide middleware to integrate IT and site

TABLE 31.1 Challenges to implement DR in data centers

Data center infrastructure	DR strategy[a]	Advantages	Future considerations and cautions[b]
Site infrastructure and mixed-use data centers	1. Adjust supply air temperature and/or humidity set points to industry and ASHRAE ranges (recommended or allowable): a. Adjust data center zone supply air temperature and humidity set points b. Adjust HVAC temperature set point for mixed-use data center zones	• Sequence of operation for this strategy is well studied and implemented in offices and commercial buildings • Strategy could be part of control system sequence of operation	• Not applicable to data centers already operating at higher temperatures • Airflow management issues • Perceived risk of IT equipment failure if strict environmental conditions are not maintained
	2. Use innovative cooling system management: a. Shut down redundant chillers, pumps, and CRAC units in response to IT equipment needs b. Expand outside air temperature range for economization (water or air)	• Significant savings when used with IT infrastructure strategies	• Higher outside air wet-bulb temperature may raise cooling water temperature • Weather dependence of air or water-side economizer • Research concept for DR
	3. Use lighting controls: a. Use bilevel switching or dimmable lighting controls to reduce lighting levels	• Sequence of operation for this strategy is well studied and implemented in office spaces and commercial buildings • Lights could be shut down completely	• Minimal impact as stand-alone strategy in non-mixed-use data centers
	4. Reconfigure redundant power delivery and backup electric storage systems: a. Use UPS bypass technology b. Shut down redundant transformers c. Use backup storage	• Strategy for shorter duration • Backup storage in use outside California; system testing can coincide with DR event	• Perceived impact on equipment or risk of error or malfunction (a) • Perceived need for additional backup storage during DR (c) • Air-quality regulatory issues if diesel generators are used (c) • Research concept for DR (a and b)
IT infrastructure	1. Use virtualization technologies: a. Increase server processor utilization rate and consolidate b. Increase storage density and consolidate c. Improve networked device efficiency	• Enabling technology available (a and b) • Enabling technology maturing (c)	• Increased utilization rates for servers may increase cooling needs with overall efficiency (a) • Research concept for DR (b and c)
	2. Shift or queue IT or backup job processing	• Enabling technology in use • Could be used as load shift	• Suited for laboratory or research and development data centers • Research concept for DR
	3. Use built-in equipment power management	• Built-in power management present in most equipment already • Energy savings higher in newer systems	• Minimal energy savings for most current equipment • Needs to be combined with virtualization and load shifting of IT or backup job strategies for DR impact • Research concept for DR
	4. Use load migration technologies for shed or shift	• Enabling technology available for some • Perennial strategy ("anytime DR")	• Infrastructure available in only a few data centers and used primarily for disaster recovery • May need local utility and coordination • Research concept for DR

TABLE 31.1 (*Cont'd*)

Data center infrastructure	DR strategy[a]	Advantages	Future considerations and cautions[b]
IT and site infrastructure synergy	1. Integrate virtualization, HVAC, lighting controls, etc. for faster load-shed response	• These intelligent strategies have higher potential energy savings than stand-alone strategies	• No enabling technologies available currently • IT and site infrastructure technology and performance measurement currently separate • Research concept for DR

[a]Except where indicated, all strategies need DR demonstration and assessment.
[b]"Research concept" in this column indicates that this DR strategy is still under development, and the impact on energy savings and scalability needs to be quantified.

infrastructure systems, and provide analytical capabilities. For example, a newly developed power supply technology that can generate reports of server power supply consumption and efficiency data could be used to coordinate with EMCS or HVAC equipment (protocols such as BACnet, Modbus, etc.) so that the cooling system communicates and responds to IT equipment (protocols such as SNMP) heat output.

31.7 AUTODR USING OPEN STANDARDS

AutoDR using open standards such as OpenADR is feasible in data centers and offers opportunities to participate in commercial DR programs. OpenADR is a set of specifications for continuous, open, secure two-way signals over a communication channel such as the Internet that allows facilities to automate their DR programs with "no human in the loop" [2]. In the NIST draft roadmap report to DOE, OpenADR was recommended as a national Smart Grid [12] standard for DR [26]. Considering the sophistication of data center technology, fully automated DR (AutoDR) should be a feasible option. OpenADR offers the following benefits:

- Reliable AutoDR for existing and new IT and site infrastructure technologies
- Integration with existing control and software systems for interoperability
- Application to commercial and industrial end uses such as lighting, HVAC, and IT equipment using existing technology and controls infrastructure

AutoDR using OpenADR has demonstrated increased reliability in comparison with the performance of facilities with manual or semiautomated DR [27].[3,4] The feasibility of OpenADR in a data center depends on the specific OpenADR-based AutoDR program offering and its suitability for integration with the data center's IT and site infrastructure systems, networks, and communications security. California utilities have invested significantly in OpenADR technology and communication infrastructure, which data center technologies can utilize. As described earlier, the site infrastructure control strategies, such as changing supply air and zone temperature set points and adjusting lighting, have already been demonstrated as OpenADR strategies within data centers that include commercial office spaces. IT virtualization technologies have also been used to consolidate servers in proof of concept studies. These virtualization technologies could be integrated with utility OpenADR infrastructure using a software client [28].

Further study of data center equipment and technologies is needed for vendors to offer built-in OpenADR features. "OpenADR ready" systems would allow data centers to participate in utility DR communications infrastructure and integrate with preprogrammed DR strategies. Most data centers using OpenADR for external communications will be concerned about network and security. Knowledgeable network administrators and software programmers can ensure secure communication of control systems and integration of automation.

[3]Manual DR: Manual turning off or changing of comfort set points or processes or of individual equipment, switches, or controllers.

[4]Semi-automated DR: Automation of one or more processes or systems within a facility using an EMCS or centralized control system, with the remainder of the facility on manual operation.

31.7.1 OpenADR Architecture

Figure 31.3 shows a typical AutoDR architecture using OpenADR standards, which is commercially implemented by California's three investor-owned utilities (IOUs): PG&E, SCE, and SDG&E. The Demand Response Automation Server (DRAS) is a middleware or broker between the utility or ISO and the participating facility's systems. DRAS can enable a standard Internet-based interface. Participating facilities or DR aggregators use a hardware or software "client" to communicate with the DRAS and retrieve DR event information; the facility then programs the DR strategies into its EMCS or other end use technology. The client can be a hardware client, such as Client and Logic with

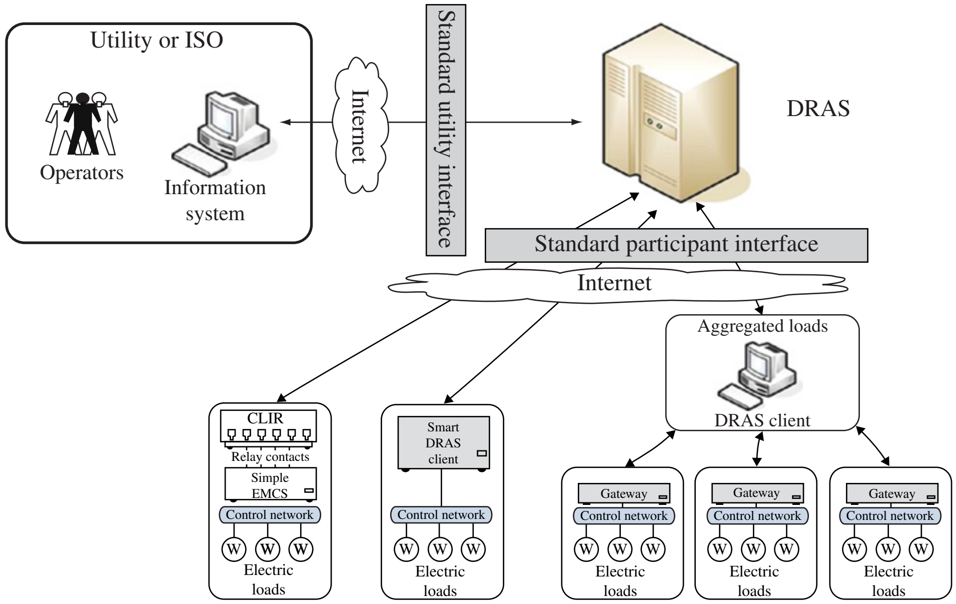


FIGURE 31.3 AutoDR architecture concept and OpenADR standards. From Ref. [15].

Integrated Relay (CLIR), any third-party device, or a software-based client integrated within facility controls and technology subsystems. Data center technologies could use a software-based client for OpenADR.

31.7.1.1 OpenADR Integration with Control Systems All sizeable data centers (>1 MW IT load) have control systems to monitor and allow regulation of cooling, power delivery, and lighting. These systems could be used for DR though they might require custom programming to reduce load automatically for DR events.

31.7.1.2 OpenADR Integration with Virtualization Technologies Virtualization technologies designed to improve IT infrastructure energy efficiency could also be used for DR strategies via a software client. For example, more aggressive virtualization strategies than normal could be activated in response to a DR event notification to increase energy savings for the duration of the event. The California ISO started a demonstration project in 2008 testing three servers in a laboratory setting; this project has shown that virtualization technologies could be integrated with existing utility or ISO OpenADR infrastructure.

31.8 GRID-DISTRIBUTED DATA CENTERS AND NETWORKS

Some data centers maintain fully networked and distributed locations on different electrical grids or geographic locations as backup for disaster recovery. In 2007, LBNL discussed with data center experts the emerging technologies currently available or in development that could allow temporary load migration of data center IT equipment loads outside a region that is experiencing a DR event. Because of this shift, IT equipment could be shut down or enabled for intelligent power management. Although this is primarily an IT infrastructure strategy, the shift in IT loads would reduce supporting site infrastructure (cooling) loads as well. Data centers that participate in DR using this strategy would likely need advance notice of the need for load migration for planning and coordination purposes. With such notice, transferring a partial or total data center workload to another data center outside the utility service territory or electric grid is possible during a DR event. Even data centers running at 100% efficiency could use this strategy.

Given the unique characteristics of data centers and their established disaster recovery scenarios to allow transfer of computing based on computer network congestion and other reasons, LBNL is conducting research to identify ways to enable DR capabilities within data centers through strategies such as load migration (computing and resulting electrical load). Such strategies could be applied for both electricity grid reliability and price-response programs through a distributed grid network, nationally and internationally. Such data center-based DR shed/shift strategies to be tested include both IT equipment and site infrastructure systems by migrating the load to a backup data center in response to grid conditions, renewable generation, and/or prices.

31.9 SUMMARY OF DR STRATEGIES

Data center DR opportunities depend on several factors, including the institutional and technical capabilities identified in previous sections. The main opportunities, from those listed in order of ease of implementation earlier, include:

- Those with the largest potential using virtualization and other emerging technologies for the servers, storage, and networking equipment that make up IT infrastructure in a data center along with corresponding reductions in cooling energy use.
- Those with the most immediate opportunities, such as raising temperature and humidity set points and lighting strategies, which have been well studied in prior research.

Challenges to implementing DR in data centers include traditional conservative operational strategies with regard to temperature and other conditions, current energy performance metrics that do not give the information needed to assess the success of DR, and lack of information about DR—its risks, benefits, and possibilities—in data centers. Additional field studies are needed to validate the performance of these strategies in data centers.

31.10 CHALLENGES TO GRID-RESPONSIVE DATA CENTERS

In addition to the cautions for implementing specific DR strategies listed earlier, some key organization and decision-making challenges need to be addressed before data centers consider DR participation. These key challenges are as follows:

1. *Perception of risk to business and operations:* Operators of "mission-critical" data centers are concerned if load reduction strategies might adversely impact reliability of their operations. Some of the strategies described in this report are research concepts whose performance and impact need to be quantified before they are adopted for DR.
2. *Performance measurement strategies:* Data centers currently measure energy performance separately for site and for IT infrastructure. However, for DR purposes, performance must be evaluated at the Whole Building Power (WBP) level. Measuring WBP indicates how much total energy the facility has saved, which the utility needs to quantify for settlement purposes.
3. *Lack of information:* Data centers may not perceive that DR is feasible for them, and those currently employing energy efficiency measures may not recognize that additional savings are possible from DR.

31.10.1 Perceived Risk to Business and Operations

Many facilities maintain data center zone temperatures at the low end of the recommended ranges because of the risk of equipment damage from overheating. Most external data centers have service-level agreements (SLAs) that specify the environmental conditions they must maintain although minor variations to temperature set point within recommended ranges could be allowed during a DR event. Data centers that can readily test and allow temperature variations are likely to be internal, R&D, or laboratory data centers offering non-mission-critical services.

In most mixed-use data centers, office HVAC and lighting account for a good portion of energy use. Office spaces use HVAC systems that are separate from the rest of the plant except when air handling units for the office space utilize chilled water from the plant that is also used for data center cooling. For integrated HVAC and CRAC systems, any DR strategy to raise temperature set points within data center zones could also affect office occupants. With well-designed strategies for mixed-use data centers, HVAC and lighting load could be reduced temporarily with no immediate impact on office occupant comfort or data center operation. The perceived risks are applicable to different data center types.

31.10.2 Performance Measurement Strategies

Data center energy use is unique in that IT and site infrastructure operations are separate and subject to different performance measurement practices. DR strategies to reduce IT infrastructure energy use can also reduce site infrastructure energy use because site end use loads respond to IT equipment energy use. Therefore, a metric is needed that captures these savings. Current data center efficiency measurement practices would not capture WBP changes. To evaluate DR performance at the WBP level, combined IT and site infrastructure performance measurements are needed. To measure DR performance, the following measurements should be made at the WBP peak demand level [29]:

- *Total watts reduced*: Used by utilities for DR load reduction estimates and payments
- *Total percent reduction*: Shows change from normal operations against a baseline
- *Watts/ft^2 reduced*: Normalizes performance to a benchmark for similar sites

31.10.3 Lack of Information

Lack of information encompasses both data center operators who do not perceive that participating in DR is feasible at all and those already practicing energy efficiency and not aware that additional savings are possible from DR. Even efficient data centers that have relatively high base load characteristics could temporarily reduce service levels during a DR event and realize additional incremental energy (and financial) savings. In addition, the data center energy use is unique in that IT and site infrastructure operations are separate and subject to different performance measurement practices. Thus, the measurement and verification metric for DR must consider the load reduction at the whole building level, which is used for calculation of load reduction and eventual payments of incentives and credits.

31.11 U.S. POLICIES GOVERNING SMART GRID EMERGING TECHNOLOGIES

Many policies exist in the United States at federal and state level. The most important of these is the Energy Independence and Security Act of 2007 (EISA-2007). The EISA-2007 Act set the stage for modernization of the aging U.S. Power grid to "Smart Grid." This federal legislation encompasses upgrading the electric utility transmission, distribution system and develops new smart appliance standards. The DOE is required to conduct Smart Grid research, development, and demonstration. As result, a new role was assigned to NIST to establish interoperability standards for Smart Grid. At the state level, apart from conducting research and development, utilities have been mandated to conduct smart grid demonstration pilots to investigate economic feasibility and interoperability with the current and emerging technologies.

31.12 THE ENERGY INDEPENDENCE AND SECURITY ACT OF 2007

The Smart Grid technology needs result from the EISA-2007, Title XIII under the following sections: 1301 (Statement of Policy on Modernization of Electricity Grid), 1302 (Smart Grid System Report), 1303 (Smart Grid Advisory Committee and Smart Grid Task Force), 1304 (Smart Grid Technology Research, Development and Demonstration), 1305 (Smart Grid Interoperability Framework), and 1306 (Federal Matching Fund for Smart Grid Investment Costs) [13].

The DOE, Federal Energy Regulatory Commission (FERC), and NIST are the primary government agencies developing smart grid policy. The underlying conceptual model defined in the NIST framework 1.0 is a legal and regulatory framework that includes policies and requirements that apply to various actors and applications and to their interactions [9]. Regulations adopted by the FERC and by public utility commissions at the state and local levels govern many aspects of the Smart Grid.

Such regulations are intended to ensure that electric rates are fair and reasonable and that security, reliability, safety, privacy, and other public policy requirements are met; see, for example, the mission statements of NARUC [30]. The transition to the Smart Grid introduces new regulatory considerations, which may transcend jurisdictional boundaries and require increased coordination among federal, state, and local lawmakers and regulators. The conceptual model must be consistent with the legal and regulatory framework and support its evolution over time. The standards and protocols identified in the framework also must align with existing and emerging regulatory objectives and responsibilities.

Standards play an important role in enabling technological innovation by defining and establishing ground rules upon which product differentiation, innovative technology development, and other value-added services may be developed and offered to utility customers. Standards are also essential for enabling seamless interoperability between and across products and systems. In the United States, private sector-led standards development that is informed by market needs has played a foundational role in facilitating competition, innovation, and global trade. A proven example is a set of standards established as part of the U.S. EPAct 2005 for energy-efficient appliances for residential and commercial use. As per the act, private sector companies carried out aggressive research and development to make appliances energy efficient and to meet the standards.

Federal government engaged in a leadership or coordinating role in private sector standardization activities to address national priorities established in statute or the administration policy. In the case of emerging technologies in smart grid area, government leadership brought together stakeholders from the various domains constituting to develop a consortium for smart grid innovation. Ordinarily, it might have taken much longer for these different stakeholders to coalesce and rapidly identify critical gaps and needs for development and adoption of an interoperable Smart Grid. This open process for standardization provides the following benefits:

- *Transparency:* Essential information regarding standardization activities is accessible to all interested parties.
- *Open participation:* All interested or affected parties have an opportunity to participate in the development of a standard, with no undue financial barriers to participation.
- *Flexibility:* Different product and services sectors rely on different methodologies for standards development that meets their needs.
- *Effectiveness and relevance:* Standards are developed in response to regulatory, procurement, and policy needs and take account of market needs and practices as well as scientific and technological developments.
- *Coherence:* The process avoids overlapping and conflicting standards.
- *International acceptance:* To benefit from international markets, the public and private sectors are best served by standards that are international in scope and applicability.
- *Net benefit:* Standards used to meet regulatory and procurement needs should maximize net benefits of the use of such standards.

31.13 STATE POLICIES FOR SMART GRID ADVANCEMENT

At the state level, several states have created legislation to facilitate the development of smart grids. In California, Senate Bill 17 requires the California Public Utility Commission (CPUC) to create a smart grid deployment plan. The bill required that standards be adopted for California that complied with standards from NIST, the Gridwise Architecture Council, the International Electrical and Electronics Engineers, the North America Electric Reliability Cooperation, and FERC. In recognition of the importance of such needs, the CPUC is taking measures to meet these requirements. For example, CPUC adopted privacy and security rules for customer data generated by smart meters that are deployed by California IOUs. At a local level, in California, The Renewables Portfolio Standard (RPS) program requires IOUs, electric service providers, and community choice aggregators to increase procurement from eligible renewable energy resources to 33% of total procurement by 2020 [31]. Recent scoping studies have evaluated the data centers and AutoDR technologies to estimate the technical potential for demand-side resources to enable better management of the Smart Grid due to the intermittent behavior of renewable generation [32].

31.14 CONCLUSIONS AND NEXT STEPS

The previous and ongoing studies suggest that there is significant potential for cost and energy savings from implementation of DR in data center facilities. Specific characteristics of data center loads that make them promising candidates for DR include minimal load variability and weather sensitivity, increasing energy costs and peak demand, and lack of existing DR programs for data center IT infrastructure.[5]

Data centers' unique operational characteristics and the use of highly advanced technology and control systems for both IT and site infrastructure make them good candidates for OpenADR, using the technology already deployed by California's IOUs. Implementation of DR in data centers faces a number of challenges, both practical and perceived. These include:

- Lack of studies and demonstrations of DR in data centers (both IT and site infrastructure) lead to perceived uncertainties about the capabilities of data centers to participate in DR programs and any resulting impacts, particularly on sensitive IT equipment and data center performance.
- Concerns about interrupting data center processes and adversely affecting quality of service and IT equipment lifespan drive data center energy use, particularly cooling loads, and hesitation to participate in DR.
- No comprehensive strategy guide exists for implementing DR strategies in data centers.
- Energy efficiency and DR technologies and practices are underutilized in data centers.
- Wide variation in different data center sizes, types, energy use, processes, and business strategies means there is no "one-size-fits-all" data center DR solution.
- Resource-dependent load patterns that are driven by outside factors such as customer needs, mission-critical applications, and resource availability lead to continuous data center availability and redundant IT equipment.
- No comprehensive assessment has quantified the value to data centers of participating in DR programs.
- Most data centers currently measure energy performance separately for site and IT infrastructure. However, for DR purposes, performance must be evaluated at the WBP level because utilities need WBP savings for DR settlement purposes.

Production data centers, which run business mission-critical applications and are fully operational at all times, have the greatest perceived equipment reliability and continuity-of-service risks, making them less motivated to participate in DR. Laboratory, research, and other "non-mission-critical" data centers are the likeliest candidates for early adoption of DR. DR strategies applicable to data center site infrastructure have been well studied for other types of commercial buildings and could be readily deployed in data centers. These include:

- Energy efficiency measures for HVAC (e.g., raising cooling set points).
- Optimization and use of lighting control systems.

DR strategies applicable to data center IT infrastructure include:

- Raising environmental set points to conform to industry standards and recommendations (e.g., raise zone supply air temperature and humidity limits).
- Using virtualization technologies to temporarily improve IT equipment efficiency, such as server consolidation and workload migration as is typically used for disaster recovery and situations where availability requirements are high.
- Using lighting control systems to reduce or turn off unnecessary lighting.
- Using synergistic strategies to integrate IT and site infrastructure energy use and load reductions.
- Using other emerging technologies, such as built-in equipment power management, UPS, transformers, bypass, and backup storage strategies.

[5]Mixed-use data centers, which contain large office spaces, have more weather-dependent loads than other types of data centers. Even within mixed-use data centers, however, the base load is very high, and load variability is only about 20%.

The results of LBNL study suggest that using virtualization technologies as a DR strategy to turn off underutilized IT infrastructure equipment will likely result in the greatest reduction of data center energy use because this strategy reduces both direct IT loads and supporting loads, particularly cooling. IT equipment technologies using OpenADR can integrate with utility communications infrastructure and data center strategies to respond to AutoDR events.

The LBNL study also found that most data centers maintain temperature and humidity levels well below industry-recommended standards. A moderate increase in temperature set points would result in immediate load reduction. If no negative impacts are experienced from reductions in cooling loads during DR events, permanent adjustments could be made to temperature set points, reducing overall data center energy use. Data centers that already employ efficient lighting and HVAC practices could benefit from additional DR strategies. Previous commercial DR program studies have shown that building control systems can be integrated with OpenADR [26].

The initial studies recommended that further research is needed on the deployment of DR in data centers to determine the specifics of DR strategies—which loads to shed, for what duration, using which technologies, and how best to interact with utilities. Studies and demonstrations of both specific strategies for DR in data centers, particularly strategies applicable to IT infrastructure, are needed to quantify the value to data centers of participating in DR programs and to create confidence that participation in DR will not undermine data center performance or equipment reliability and lifespan. The 2007 PG&E AutoDR program assessment recommends further studies of data centers for OpenADR in particular.

31.14.1 Commercialization Potential

This report looks at the potential use of existing and emerging energy efficiency technologies for DR and AutoDR. Significant commercial potential may exist in integration of OpenADR clients with data center virtualization technologies. The "integrated systems" approach of IT and a site infrastructure technology is another area that has significant energy efficiency and DR potential and could be integrated with OpenADR. The adoption of DR strategies will depend on data centers' willingness to participate in DR, which in turn will depend on technology maturity and quantification of the energy savings of these strategies and their scalability.

31.14.2 Next Steps

The earlier studies have revealed the potential of DR within various types of data center end uses. These study findings helped develop a focused scope for next phase of studies of data center DR. Cosponsored by the IOUs, PG&E, SDG&E, and the California Energy Commission, the next phase of studies are under way at LBNL to:

- Evaluate existing enabling technologies (monitoring infrastructure, server virtualization software, network capabilities for distributed computing).
- Identify growth areas in monitoring infrastructure, data center virtualization technologies and network capabilities for distributed computing clusters.
- Investigate and demonstrate DR opportunities within various end use loads in a data center IT infrastructure, cooling systems, lighting, and other additional loads and how such end uses can support federal and state policies (e.g., California Renewables Portfolio Standard (RPS)).
- Recruit production, backup, scientific computing, and geographically distributed data centers who are interested in becoming the early adopters of DR to conduct field tests.
- Analyze DR measures resulting from load shed, load shift, and load migration strategies.
- Potentially integrate externally with utility or grid operator-driven OpenADR standards and automation of control strategies for AutoDR.

Early study results, which are yet to be formally published, show that IT loads can manually be turned off in a DR event in less than 8 min, and there is a significant load reduction potential. This has the potential for data centers to be excellent candidates to participate in AutoDR programs and integrate with OpenADR for both retail and wholesale DR markets.

Future areas of research will involve developing cloud-based distributed data center management automation software, which is capable of seamlessly migrating storage and computing loads across geographically distributed data centers. Petascale and Exascale computing systems will be capable of responding to price and DR signals from electric utilities or CAISO via OpenADR to dynamically shift processing and data storage loads across various utility territories within the United States and globally.

ACKNOWLEDGMENTS

The work described in this report was coordinated by the DRRC and funded by the California Energy Commission (Energy Commission), Public Interest Energy Research (PIER) Program, under Work for Others Contract No. 500-03-026, by the Pacific Gas and Electric Company (PG&E) under Work for Others Contract No. PGZ0803, and by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. SDG&E under Work for Others Contract No. POSD01-L01 also funded the ongoing study references.

The authors would like to thank the study contributors, Sydny Fujita, Aimee McKane, June Dudley, Anthony Radspieler, and Nance Matson (LBNL), KC Mares (Megawatt Consulting), Dave Shroyer (SCG), Bill Tschudi, and Dale Sartor for their data center energy efficiency expertise. The authors also want to acknowledge all others who assisted in review of this document and for their ongoing support, including Energy Commission PIER program's Ivin Rhyne, Chris Scruton, Anish Gautam, Paul Roggensack, and Mike Gravely, PG&E's Albert Chiu and Jonathan Burrows, and SDG&E's Eric Martinez.

REFERENCES

[1] McKane AT, Piette MA, Faulkner D, Ghatikar G, Radspieler Jr, A, Adesola B, Murtishaw S, Kiliccote S. 2008. Opportunities, barriers, and actions for industrial demand response in California. LBNL-1335E. Lawrence Berkeley National Laboratory. Available at http://drrc.lbl.gov/sites/all/files/lbnl-1335e.pdf. Accessed on June 12, 2014.

[2] Piette MA, Ghatikar G, Kiliccote S, Koch E, Hennage D, Palensky P, McParland C. Open automated demand response communications specification (version 1.0). California Energy Commission, PIER Program; 2009. CEC-500-2009-063. LBNL-1779E.

[3] EPA. EPA report to congress on server and data center energy efficiency public law 109–431; 2007. Available at http://www.energystar.gov/index.cfm?c=prod_development.server_efficiency, http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf. Accessed on April 26, 2012.

[4] Koomey JG. Growth in data center electricity use 2005–2010. Report by Analytics Press; 2011. Available at: http://www.mediafire.com/file/zzqna34282frr2f/koomeydatacenterelectuse2011finalversion.pdf. Accessed on May 24, 2014.

[5] Heffner G. 2009. Demand response valuation framework. LBNL-2489E. Lawrence Berkeley National Laboratory. Available at http://drrc.lbl.gov/sites/all/files/lbnl-2489e.pdf. Accessed on May 24, 2014.

[6] Lekov A, Thompson L, McKane A, Rockoff A, Piette MA. 2009. Opportunities for energy efficiency and automated demand response in industrial refrigerated warehouses in California. LBNL-1991E. Lawrence Berkeley National Laboratory. Available at http://drrc.lbl.gov/sites/all/files/lbnl-1991e.pdf. Accessed on June 12, 2014.

[7] Pacific Gas and Electric Company. What is demand response?; 2008. Available at http://www.pge.com/mybusiness/energysavingsrebates/demandresponse/whatisdemandresponse. Accessed on April 26, 2012.

[8] Flex Your Power. Demand response programs. 2008. Available at http://www.fypower.org/flexalert/demand_resp_faq.html. Accessed on April 26, 2012.

[9] NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 1.0, NIST special publication 1108. January 2010. Available at http://www.nist.gov/public_affairs/releases/upload/smartgrid_interoperability_final.pdf. Accessed on May 24, 2014.

[10] NIST Framework 2.0: NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 2.0, NIST special publication 1108R2; February 2012. Available at http://www.nist.gov/smartgrid/upload/NIST_Framework_Release_2-0_corr.pdf. Accessed on August 16, 2014.

[11] Lawrence Berkeley National Laboratory, High Performance Buildings for High-Tech Industries. 2009. Data center server power supplies. Available at http://hightech.lbl.gov/. Accessed on July 13, 2009.

[12] U.S. DOE Secretary Dr. Steve Chu's presentation; GridWeek. 2009. Available at http://www.pointview.com/data/2009/09/31/pdf/Steve-Chu-4774.pdf. Accessed on May 24, 2014.

[13] Energy Independence and Security Act (EISA). 2007. Available at http://www.whitehouse.gov/assets/documents/recovery_plan_metrics_report_508.pdf. Accessed on May 24, 2014.

[14] American Recovery and Reinvestment Act of 2009 (ARRA). Available at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_public_laws&docid=f:publ005.pdf. Accessed on May 24, 2014.

[15] Ghatikar G, Piette MA, Fujita S, McKane AT, Han JQ, Radspieler A, Mares KC, Shroyer D. 2010. Demand response and open automated demand response opportunities for data centers. Lawrence Berkeley National Laboratory. Available at http://eetd.lbl.gov/sites/all/files/demand_response_and_open_automated_demand_response_opportunities_for_data_centers_lbnl-3047e_0.pdf. Accessed on June 12, 2014.

[16] NIST Smart Grid Testing and Certification Committee (SGTCC). 2011. Interoperability process reference manual, Version 1.0. Available at http://collaborate.nist.gov/twiki-sggrid/bin/view/SmartGrid/SmartGridTestingAndCertificationCommittee. Accessed on May 24, 2014.

[17] Ghatikar G, Bienert R. Smart grid standards and systems interoperability: a precedent with OpenADR, in Grid-Interop 2011, Phoenix, 2011.LBNL-5273E.pdf

[18] Silicon Valley Leadership Group and Accenture. 2008. Data center energy forecast. XIII of the Energy Independence and Security Act of 2007 (Pub.L. 110–140). Silicon Valley Leadership Group. Available at http://accenture.com/SiteCollectionDocuments/PDF/Accenture-Data-Center-Energy-Forecast.pdf. Accessed on June 12, 2014.

[19] Gartner. 2008. Available at http://www.gartner.com/it/page.jsp?id=799812. Accessed on April 26, 2012.

[20] Shehabi A, Tschudi W, Gadgil A. Data center economizer contamination and humidity study. 2007. Report to Pacific Gas and Electric Company. LBNL-2424E. Lawrence Berkeley National Laboratory. Available at http://hightech.lbl.gov/documents/data_centers/economizerdemoreport-3-13.pdf. Accessed on June 12, 2014.

[21] Gartner. 2007. Available at http://www.gartner.com/it/page.jsp?id=535714. Accessed on April 26, 2012.

[22] Lawrence Berkeley National Laboratory (LBNL). 2009. High performance buildings for high-tech industries: data center server power supplies. Available at http://hightech.lbl.gov/. Accessed on July 13, 2009.

[23] Motegi N, Piette MA, Watson DS, Kiliccote S, Xu P. 2007. Introduction to commercial building control strategies and techniques for demand response. California Energy Commission. LBNL-59975. Lawrence Berkeley National Laboratory. Available at http://drrc.lbl.gov/sites/all/files/59975.pdf. Accessed on June 12, 2014.

[24] Roth K, Westphalen D, Feng M, Llana P, Quartararo L. 2005. Energy impact of commercial building controls and performance diagnostics: market characterization, energy impact of building faults and energy savings potential. TIAX LLC. Available at http://s3.amazonaws.com/zanran_storage/www.tiaxllc.com/ContentPages/42428345.pdf. Accessed on June 12, 2014.

[25] Brill KG. 2008. Special report: energy efficiency strategies survey results. Uptime Institute.

[26] Electrical Power Research Institute (EPRI). Report to NIST on smart grid interoperability standards roadmap; 2009. Available at http://www.nist.gov/smartgrid/upload/InterimSmartGridRoadmapNISTRestructure.pdf. Accessed on April 26, 2012.

[27] Wikler G, Bran I, Prijyanonda J, Yoshida S, Smith K, Piette MA, Kiliccote S, Ghatikar G, Hennage D, Thomas C. Pacific gas & electric company 2007 AutoDR program: task 13 deliverable: AutoDR assessment study. Report to PG&E; 2008. Lawrence Berkeley National Laboratory. Available at http://drrc.lbl.gov/sites/all/files/pge-auto-dr-assessment-study.pdf. Accessed on June 12, 2014.

[28] Cassatt Pacific Gas and Electric Company feature pack for automated DR installation and user guide. Available at http://www.cassatt.com/infocentral/PGE/1.0/docs/InstallingPGE/. Accessed on April 26, 2012.

[29] Motegi N, Piette MA, Watson D, Sezgen O. Measurement and evaluation techniques for automated demand response demonstration. Proceedings of the 2004 ACEEE Summer Study on Energy Efficiency in Buildings; August 23–27; Pacific Grove; 2004. LBNL-55086.

[30] Mission statements of NARUC and FERC. Available at http://www.naruc.org/about.cfm, http://www.ferc.gov/about/about.asp. Accessed on May 24, 2014.

[31] California Energy Commission (CEC), 2002 under Senate Bill 1078, accelerated in 2006 under Senate Bill 107 and expanded in 2011 under Senate Bill 2, California's Renewables Portfolio Standard. Sacramento, CA.

[32] Watson D, Matson N, Page J, Kiliccote S, Piette MA, Corfee K, Seto B, Masiello R, Masiello J, Molander L, Golding S, Sullivan K, Johnson W, Hawkins D. Fast automated demand response to enable the integration of renewable resources. California Energy Commission; 2012.

FURTHER READING

DRRC annual report. 2013. Advancing grid integration through research, development, demonstration, deployment. Demand Response Research Center, LBNL, Berkeley, California. Available at http://drrc.lbl.gov/sites/all/files/DRRC%20Annual%20Report%202013.pdf. Accessed on August 16, 2014.

NREL. 2011. Reducing data center loads for a large-scale, net zero office building. DOE/GO-102011-3459. Available at http://www.nrel.gov/docs/fy12osti/52786.pdf. Accessed on April 26, 2012.

Public Law 109-58 US Energy Policy Act of 2005. Available at http://www.gpo.gov/fdsys/pkg/PLAW-109publ58/pdf/PLAW-109publ58.pdf. Accessed on May 24, 2014.

PART IV

DATA CENTER OPERATIONS AND MANAGEMENT

32

DATA CENTER BENCHMARK METRICS

William J. Kosik
Hewlett Packard Company, Chicago, IL, USA

32.1 INTRODUCTION

It is important to note that there are currently a number of energy and performance metrics for data centers in addition to PUE™. Many go beyond the facility power and cooling systems and judge the efficacy of the IT systems. Some of these are being used by the industry, while others are still under development. The information presented here focuses on PUE, since it is used worldwide as a standard to measure data center efficiency and is not meant to minimize the importance of the other standards.

32.2 ORIGIN AND APPLICATION OF PUE AS A METRIC

In 2007, the US Environmental Protection Agency (EPA) report to Congress on server and data center energy efficiency stated: "The federal government and industry should work together to develop an objective, credible energy performance rating system for data centers, initially addressing the infrastructure portion but extending, when possible, to include a companion metric for the productivity and work output of IT equipment." This was a clear directive to the industry to develop a uniform metric. Just a few years later, the metric Power Usage Effectiveness (PUE), developed and proliferated by the Green Grid to determine the energy efficiency of data centers, was being used worldwide in the technology industry and had become a mainstream approach to determine data center energy use efficiency. While the definition of PUE is generally understood and has been in use for a number of years, the industry is still fine-tuning how PUE is computed, both theoretically and when measuring energy consumption from a live data center.

PUE takes into account how energy is consumed for all systems within the data center, cooling, power distribution, and other ancillary systems. A PUE can be developed for the power and cooling systems individually, but more importantly, as a group representing a total PUE for the entire facility. The PUE is a measure of power efficiency and is represented by the following equations:

$$\mathrm{PUE} = \frac{\sum \text{Power delivered to data center}}{\sum \text{IT equipment power use}} = \frac{P_{\text{mechanical}} + P_{\text{electrical}} + P_{\text{other}}}{P_{\text{IT}}}$$

$$\mathrm{PUE}_{\text{mechanical}} = \frac{P_{\text{mechanical}}}{P_{\text{IT}}}$$

$$\mathrm{PUE}_{\text{electrical}} = \frac{P_{\text{electrical}}}{P_{\text{IT}}}$$

$$\mathrm{PUE}_{\text{other}} = \frac{P_{\text{other}}}{P_{\text{IT}}}$$

Power is measured at the main utility transformer, mechanical switchgear, UPS, and miscellaneous load panels. Measuring the input and output power indicates efficiency of different electrical components, which becomes additional data when attempting to optimize efficiency. The document *Recommendations for Measuring and Reporting Overall Data Center Efficiency*, authored by an industry consortium

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

in 2011, provides guidance on a consistent methodology for establishing PUE.

So, how can the PUE be used to analyze alternative system types in order to arrive at a decision when building a new data center or examining potential existing facility upgrades? This is another area where the interdependencies are vital to identify. Climate, cooling system type, power distribution topology, and redundancy level (reliability, availability) will drive the power efficiency of these systems. When an analysis is performed to determine peak and annual energy use of the facility, these interactions will become obvious. An example that illustrates this point are two recent data center projects, one in Delhi, India, and the other in Zurich, Switzerland. The Delhi data center has a PUE of 1.35 and the annual HVAC energy use is 16% of the total annual energy consumption. Contrast to that is the Zürich data center that has a PUE of 1.21 and the annual HVAC energy use is 7% of the total annual energy consumption. The two data centers have the same type of HVAC system, but the differences in energy use primarily come from the climate and how the HVAC system is operated. In general, of the total energy use of the facility, the mechanical system can consume approximately 5–20% of the total (Fig. 32.1a), and of that, the power required for cooling will typically be close to 50% (Fig. 32.1b). These percentages will vary based on many factors such as data center size, lighting density, miscellaneous power requirements, and operational schedule. The design, control, and operation of the mechanical system will have a significant influence on the annual energy use of the facility but also presents an opportunity to reduce energy consumption.

The efficiency of the HVAC system depends not only on the efficiency of the HVAC components themselves but also on electrical systems that support the cooling and ventilation systems. These elements include the power transformation from medium voltage to low voltage to support the various mechanical systems, UPS power for those mechanical components requiring uninterrupted power, such as CRAC fans or chilled water pumps, and mechanical wiring distribution losses. The electrical power consumption includes all the power losses starting from the utility and includes transformers, the UPS plant, power distribution units (PDUs), with or without static transfer switches, and Remote Power Panels (RPPs) that provide power to the IT equipment. For a typical data center, the total electrical system efficiency at 100% load will range from 85 to 95% with the non-UPS system losses ranging from 2 to 4%.

Of the largest non-IT energy consumers in data centers that use mechanical cooling, chillers, or other electrically driven vapor compression, cooling equipment will expend the greatest amount of energy. One primary strategy to decrease overall energy consumption is to elevate the supply air temperature by increasing the chilled water supply temperature and/or reducing the temperature of the air moving across the condensing coil. However, the ability to incorporate this strategy will completely depend on the type of mechanical system, the climate, and the allowable supply air temperature for the IT equipment. Consider that for fixed speed chillers, every 1°F increase in chilled water temperature can increase chiller energy efficiency 1–2%. For VSD chillers, every 1°F increase in chilled water temperature can result in a 2–4% efficiency increase. Therefore, increasing the supply air temperature from 60 to 75°F will result in an average efficiency increase of the chiller of nearly 40%.

After all of the detailed measurement and analysis is complete, care must be taken when attempting to assemble the PUE. As an example, a facility may have a very efficient

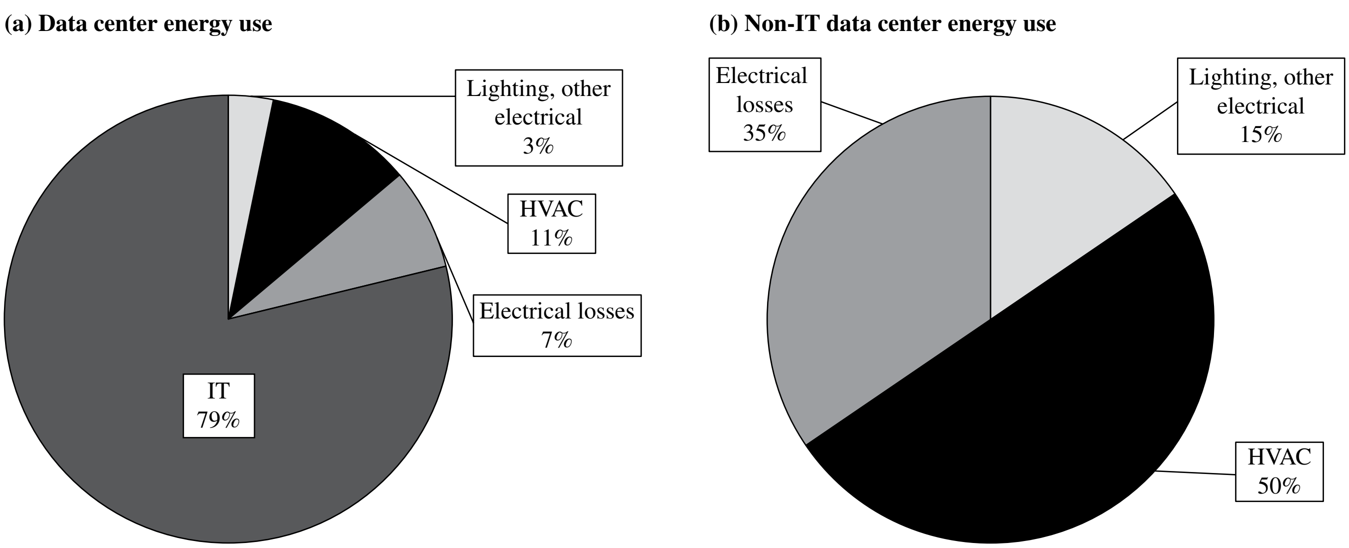


FIGURE 32.1 (a and b) The losses due to electrical system inefficiency and energy use related to cooling the data center make up a large percentage of the overall annual energy use of a data center.

cooling system, but the electrical distribution system may have very high losses. The PUE might appear to be good; but without the data on the electrical system's performance, it is impossible to develop valid efficiency strategies. Similarly, if a cooling system has an extremely efficient chiller, but the chilled water pumps are consuming an inordinate amount of energy, generating granular data is required to uncover this type of anomaly. Finally, PUE is not meant to compare the performance of different data centers. It is meant to assist in identifying areas of improvement in a single facility, to develop a personal best if you will.

32.3 METRICS USED IN DATA CENTER ASSESSMENTS

After the measurement data from the energy audit have been collected and analyzed, it is helpful to use the information as inputs for industry metrics that have been developed specifically to help data center owners benchmark and judge the overall effectiveness of their data center. Many of these metrics are still in their initial release; so, it is essential that the users of these programs provide feedback to the organizations issuing the metrics.

Since many of the programs and metrics use established protocols, it is vital to know if there are any specific criteria that must be adhered to when taking measurements, collecting data, and assembling the final analytics. As an example, when using the Rack Cooling Index (RCI), the appropriate guidelines need to be used to ensure accurate and consistent measurements. There are several types of metrics related to data center performance; not all of these metrics must be used to determine the efficiency of the data center, rather the ones that will generate the most meaningful statistics beneficial to bettering the energy use of the facility.

32.4 GREEN GRID'S XUE METRICS

Since the Green Grid released their white paper on PUE in October 2007, they have been developing a family of xUE metrics "designed to help the data center community better manage the energy, environmental, societal, and sustainability-compliance parameters associated with building, commissioning, operating, and de-commissioning data centers."

32.4.1 Energy Reuse Effectiveness

The metric Energy Reuse Effectiveness (ERE) was developed to recognize that some data centers have the ability to provide energy that can be reused in other parts of the facility or campus. One of the more common ways this can be applied is by providing low-grade heat obtained from the discharge from the servers (through a heat exchange process) to heat adjacent buildings or preheat domestic water before it enters an electric or natural gas powered water heater. The process is represented as follows:

$$\text{ERE} = \frac{\text{Annual facility energy use} - \text{Annual energy reused}}{\text{Annual IT energy use}}$$

32.4.2 Water Usage Effectiveness

The metric Water Usage Effectiveness (WUE) is analogous to PUE. The purpose of the metric is to determine the efficacy of water use in the data center, based on the energy used by the IT equipment. The formula is

$$\text{WUE} = \frac{\text{Annual site water usage}}{\text{Annual IT energy use}}$$

Knowing that by reducing site water consumption, it is possible to increase the site energy consumption, it is also advisable to look at the source water consumption that is associated with generating the electricity supplied in the data center. (This scenario comes about when, in certain locations and for certain types of data centers, air-cooled direct expansion or DX cooling equipment is used in lieu of water-cooled equipment. The amount of site water used will be much lower using the DX system, but potentially, more site energy is used. With more site energy being used, the source water—the water used at the electrical generation plant—will increase.) This way, the metric will yield a more complete picture of water that is consumed on a regional level, where it is most important. When the source water is taken into consideration, the formula changes as follows:

$$\text{WUE}_{\text{source}} = \frac{\text{Annual site water usage} + \text{Annual source water usage}}{\text{Annual IT energy use}}$$

This involves using data on water use of the regional power plants, which will most likely have to be an estimate, since exact water consumption figures may not be available.

32.4.3 Carbon Usage Effectiveness

The metric Carbon Usage Effectiveness (CUE), similar to PUE, judges the amount of carbon that is expended as compared to the annual IT energy used in the data center. Like the other metrics, the data center owner is encouraged to decrease the numerator, thereby making the CUE smaller. The CUE is defined as follows:

$$\text{CUE} = \frac{\text{Annual CO}_2\text{ emissions caused by the data center}}{\text{Annual IT energy use}}$$

This is a source-driven metric, since the CO_2 emissions come from the power plants that feed the electric grid that supplies electricity to the data center. Like the water consumption figures used in calculating the WUE, the CO_2 emissions come from data issued by the US EPA in the United States and the International Energy Agency internationally. Once the annual data center energy in kWh is determined, it is a simple calculation to determine the annual CO_2 emissions. Like WUE, these metrics are especially useful when vetting sites for building a greenfield data center.

32.5 RACK COOLING INDEX AND RETURN TEMPERATURE INDEX

32.5.1 Rack Cooling Index®

Rack Cooling Index (RCI) is a dimensionless factor developed by Dr Magnus Herrlin. It is a metric to determine the effectiveness of the air management in the data center. Using CFD modeling (or measurements in an existing data center), the mean inlet temperature is used in conjunction with the maximum allowable and maximum recommended temperature to develop a percentage effectiveness, where 100% indicates that the mean inlet temperature exactly meets the requirements. There are two metrics RCI_{HIGH} and RCI_{LOW} to reflect that the temperatures of the IT cabinet at the upper and lower levels will vary. RCI is defined as follows:

$$RCI_{LO} = \left(1 - \frac{\text{Total under} - \text{Temp}}{\text{Max allowable under} - \text{Temp}}\right)100\%$$

$$RCI_{HIGH} = \left(1 - \frac{\text{Total over} - \text{Temp}}{\text{Max allowable over} - \text{Temp}}\right)100\%$$

Based on the standard being used, the numerical value of these indices will vary.

32.5.2 Return Temperature Index

Return Temperature Index (RTI), developed by Dr Magnus Herrlin, is a metric to determine the efficacy of the air management in a data center. Similar to RCI, it judges how effective the air distribution system is at isolating the cold air meant for the computer equipment from the hot air that is expelled from the equipment. When the rack inlet temperatures are equal to the supply and return air temperatures, the RTI will be 100%, meaning there is no mixing and the supply air that comes from the air handling system is of the same temperature as what is delivered to the IT equipment. And, when the air that is being returned to the air handling system is equal to the air temperature at the discharge of the IT equipment, there must be no mixing.

$$RTI = \left(\frac{\begin{array}{c}\text{Return air temp}\\ -\text{Supply air temp}\end{array}}{\begin{array}{c}\text{Rack outlet mean temperature}\\ -\text{Rack inlet mean temperature}\end{array}}\right) \times 100$$

32.6 ADDITIONAL INDUSTRY METRICS

There are a number of metrics that focus on the IT equipment performance and energy consumption. While these currently do not directly tie into how PUE is established, these metrics impact data center energy efficiency and are vital elements of the overall energy use strategy in a data center.

32.6.1 SPEC

According to the website, "The Standard Performance Evaluation Corporation (SPEC) is a nonprofit corporation formed to establish, maintain and endorse a standardized set of relevant benchmarks that can be applied to the newest generation of high-performance computers. SPEC develops benchmark suites and also reviews and publishes submitted results from our member organizations and other benchmark licensees." The specific standard, SPECpower_ssj2008, is a benchmark that evaluates the power and performance characteristics of volume server class and multi-node class computers, providing a method to measure power in conjunction with a performance metric. This benchmark test provides vital data to engineers and operators on the range of actual power demand of a server.

32.6.2 The Green500™

The Green500 ranks the most energy-efficient supercomputers in the world using to the metric megaflops per watt (Mflops/W) where FLOPS is short for floating-point operations per second. This metric was initially developed to address the massive amounts of energy that supercomputers consumed when running workloads. Since then, in addition to computational ability, the computer's performance per unit power has become a part of the computer's overall capability. The testing can be done by the end user and must follow the Green500s testing protocol, *Power Measurement of Supercomputers*. The results are tabulated and released approximately every 6 months.

32.7 EUROPEAN COMMISSION CODE OF CONDUCT

The Code of Conduct (COC) was developed to counteract the escalating energy consumption in data centers, as well as to minimize the related negative environmental, economic, and energy supply security impacts. The primary goal of the COC is to provide the necessary information to data center operators and owners to assist in reducing energy consumption without reducing the availability of the data center. The COC consists of a number of tools to help people responsible for a data center's operations identify, plan, and implement energy efficiency programs, including a best practices handbook.

32.8 INTERNATIONAL TELECOMMUNICATION UNION

The ITU released Toolkit on Environmental Sustainability for the ICT Sector "with the following aspects of environmental sustainability in ICT organizations: sustainable buildings, sustainable ICT, sustainable products and services, end of life management for ICT equipment, general specifications and an assessment framework for environmental impacts of the ICT sector." The document was developed by the ITU in conjunction industry leaders.

32.9 CONCLUSION

It is important to understand that these metrics should be used together, providing a range of data points to help understand the efficiency and effectiveness of a data center; different combinations of these metrics will produce a synergistic outcome. As an example, when PUE is used in conjunction with WUE, it is possible to see how the values interrelate with each other and why it is good to look at the corresponding water consumption when different energy efficiency strategies are contemplated. Similarly, when analyzing different cities for a new data center build, using PUE and CUE will result in data that is influenced by the type and efficiency of the local power generation, and how the climate affects the cooling system performance. Using these metrics alone or in strategic combinations brings great value to analyzing energy use in the data center.

FURTHER READING

ANSI/AHRI 1360 (I-P)-2013. Performance Rating of Computer and Data Processing Room Air Conditioners.

ASHRAE Guideline 14-2002. Measurement of Energy and Demand Savings.

ASHRAE Real-Time Energy Consumption Measurements in Data Centers.

ASHRAE Standard 90.1-2013 (I-P Edition). Energy Standard for Buildings Except Low-Rise Residential Buildings.

Building Research Establishment's Environmental Assessment Method (BREEAM) Data Centres; 2010.

Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric, the Green Grid.

Data Center Air Management Research. Pacific Gas and Electric Company, Emerging Technologies Program, Application Assessment Report #0912; Issued September 22, 2010.

ERE: A Metric for Measuring the Benefit of Reuse Energy from a Data Center, the Green Grid.

Green Grid Data Center Power Efficiency Metrics: PUE and DCIE, the Green Grid.

Koomey JG. Estimating Total Power Consumption by Servers in the U.S. and the World; February 15, 2007. Final report.

Koomey JG. Growth in Data Center Electricity Use 2005 to 2010.

Lawrence Berkeley Lab High-Performance Buildings for High-Tech Industries, Data Centers.

PUE™: A Comprehensive Examination of the Metric, the Green Grid. International Telecommunication Union: Environmental Sustainability for the ICT Sector; 2012.

Recommendations for Measuring and Reporting Overall Data Center Efficiency Version 2—Measuring PUE for Data Centers, the Green Grid.

Singapore Standard SS 564: 2010 Green Data Centres.

US Green Building Council—LEED Rating System.

Usage and Public Reporting Guidelines for the Green Grid's Infrastructure Metrics (PUE/DCIE), the Green Grid.

Water Usage Effectiveness (WUE™): A Green Grid Data Center Sustainability Metric, the Green Grid.

33

DATA CENTER INFRASTRUCTURE MANAGEMENT

Mark Harris
Nlyte Software, San Mateo, CA, USA

33.1 WHAT IS DATA CENTER INFRASTRUCTURE MANAGEMENT?

The Data Center industry is awash with change. Since the days of the Dot-Com era, the data center has been massaged, squeezed, stagnated, and reconstituted more than once for the purposes of cost reductions, increased capacity, compliance and control, and overall efficiency improvements. In a survey of IT professionals published by Gartner in their "IT Key Metric Report" (December 2012), almost a third of all global IT budgets are spent on data center infrastructure and its operations; and surprisingly, very few companies have invested in the tools, technologies, and discipline needed to actively manage these huge capital investments.

Data Center Infrastructure Management (DCIM) is now a critical management solution for Data Centers. As a new category, the origin of the term "DCIM" is not clear, nor is the exact definition of DCIM universally agreed at the moment. That said, the initial spirit of DCIM can be summarized much in the way Gartner has expressed it: "The integration of information technology and facility management disciplines to centralize monitoring, management and intelligent capacity planning of a data center's critical systems. Additionally, DCIM is achieved through the implementation of specialized software, hardware and sensors. DCIM will enable a common, real-time monitoring and management platform for all independent systems across IT and Facilities and must manage workflows across all systems."

DCIM has transitioned well beyond simple monitoring, drawing pretty pictures and interactive eye candy, and has become the data center extension to a number of other systems, including asset and service management, financial general ledgers, etc. In a few cases, DCIM solutions have been created that have an element of control for very specific hardware environments. At the end of the day, a well-deployed DCIM solution quantifies the costs associated with moves, adds, and changes on data center floor, it understands the cost and availability to operate those assets, and clearly identifies the value derived through the existence of that asset over its useful lifespan. And true to the original spirit of DCIM mentioned earlier, these business management views span the IT and Facilities worlds.

Taking a closer look at Figure 33.1, you can see that DCIM is in direct support of modern approaches to data center asset and service management. Combining two well-known models from Forrester and The 451 Group, you can see how DCIM provides the view of the data center from the physical layer upward, whereas most IT management umbrellas in use currently are limited to a top-down logical view. According to a recent IDC report, 57% of data center managers consider their data centers to be inefficient and 84% of those surveyed have issues with space, power, or cooling that is directly affecting the bottom line. Clearly, these models must converge into a single management domain with combined views.

33.1.1 DCIM Maturity—The Technology and the User

Any new technology typically emerges after a long journey. Taking any of the recent data center examples like Virtualization or Cloud computing, it can be noted that there are several distinct periods that must be traversed before any technology is deployed in standard Production. Gartner refers to this flow as the "Hype Cycle," a broader interpretation of which is shown in Figure 33.2. What starts as an

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

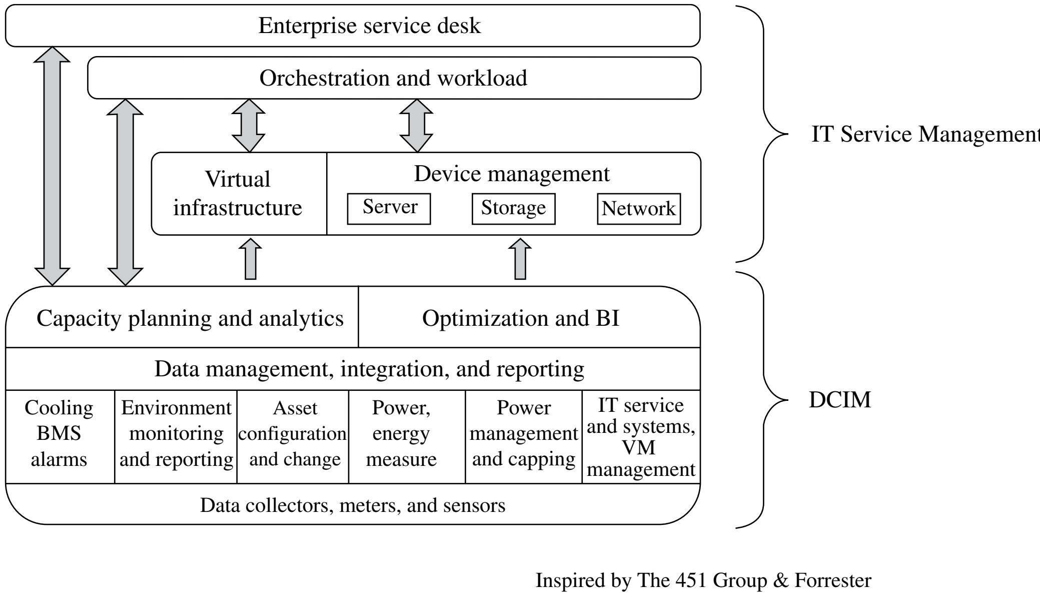


FIGURE 33.1 DCIM has been broadly defined as the management layer of physical infrastructure that supports the IT function.

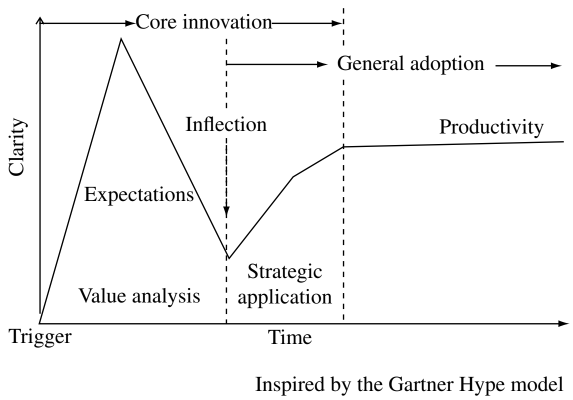


FIGURE 33.2 "**Hype Cycle**"—The maturation of any new technology follows a very predictable adoption cycle. Wild enthusiasm is replaced by reality, with the most sound ideas becoming production.

amazing innovative idea with all the promise in the world gets tested and retested over the subsequent years with a dose of reality thrown in for good measure. Technologists and business managers alike poke and prod at new inventions to determine how it could pertain to their own use cases. Over time, some inventions vanish for various reasons, and others emerge with general adoption growth rates over time.

DCIM has been following this same curve. Referring to Figure 33.3, you can see that various organizations are currently at different stages in their ability to think about their computing needs in the future and are being challenged to self-evaluate their own IT best practices of the present. After years and years of ad hoc asset management solutions for the data center and the typical unique change processes that have abounded throughout the data center industry, we find a set of points along the maturity continuum that characterizes the industry's current status.

33.1.2 DCIM Is Strategically Important to the Modern Data Center

DCIM is a resource and capacity planning business management solution for the data center. It enables the data center to leverage existing physical layer technologies including monitoring, capacity planning, configuration data bases, environmental sensing, etc., and it also enables the seamless integration into an enterprise's other business management solutions used for asset management, process management, data management, HR planning, budgetary planning, SOX compliance,[1] etc. In a twenty-first century company, information is the most strategic asset and competitive differentiator. The data center structure itself is the factory floor that produces that strategic value of processing workloads or transactions. DCIM is the manager of the data center floor in which hundreds of millions of dollars of assets and billions of dollars of information flow in nearly every fortune 500 enterprises. DCIM is the business management for this critical infrastructure and will be instrumental to the dynamic self-adjusting data center of the near future.

There are at least four stakeholders emerging as having keen interest in DCIM, and each has their own sets of needs from the adoption of DCIM: (1) the IT organization, (2) the Facilities organization, (3) the Operations' finance

[1]The Sarbanes-Oxley Act (SOX) of 2002 is a US legislation that requires company's IT in compliance of record-keeping practices to support in the event of an audit.

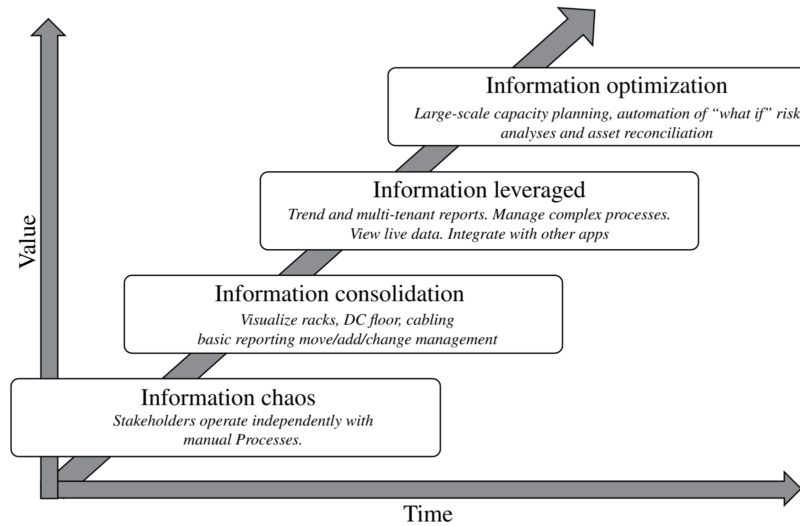


FIGURE 33.3 Existing data centers have management processes that vary widely in their maturity. With DCIM, each can be optimized.

department, and (4) the Corporate Social Responsibility individuals.

In general, the compilation of everything "IT" is becoming viewed as a single entity that has a quantifiable value to the organization and associated cost. The IT and Facilities organizations are now being tasked with a common set of goals regarding the data center, and in response, finding themselves required to behave as a single business unit, with transparency, oversight and accountability, forecasting, and overall effectiveness, all in focus. DCIM enables that focus and has become essential for the data center community.

Andy Lawrence at The 451 Group put it most succinctly: "We believe it is difficult to achieve the more advanced levels of data center maturity, or of data center effectiveness generally, without extensive use of DCIM software. Today, the three main drivers of investment in DCIM software are economics (mainly through energy-related savings), improved availability, and improved manageability and flexibility."

As the IT industry transformation continues, we see a handful of corporate IT goals that are greatly influenced, enabled, and supported by a well-conceived DCIM strategy, which puts DCIM squarely in the same level of importance as any of the other deployed business management applications:

1. IT transformation including the Cloud
2. Actively managing power needs and focus on Green IT
3. Asset Migrations and Capacity Planning
4. Operational Excellence and process reengineering
5. Deferring new capital expenditures and, in particular, Data Center builds
6. Audit and Regulatory and Legislation compliance.

33.1.3 Common Goals for DCIM

DCIM has become a general-purpose "efficiency" category where a wide range of stakeholders have voiced their data center management needs. Some of the most common goals for the introduction of DCIM into operational and strategic plans include the following:

1. Energy management has become the first priority for most data center managers and IT business managers alike. Their goal is to reduce operating costs across the board, with the starting point being much more proactive energy management. DCIM solutions can provide a highly granular understanding about energy usage and a wide range of other physical layer metrics and ultimately help to identify and control inefficiencies.
2. The need for highly accurate and actionable data and views about the current capacity and availability of their data centers and easy access to baselines associated with current operations.
3. Operational Best Practices are being redefined to accommodate much more streamlined operations and remediation. DCIM is being viewed as the tool best suited to identify, present, and manage the workflow associated with data center physical assets. DCIM essentially captures and enforces the processes associated with change.
4. With such a wide range of traditional tools in use, even within the same organization, there exists a very loose and disconnected source of truth about the data center assets. DCIM promises to coordinate and consolidate disparate sources into a single view of asset knowledge.
5. Resource availability and capacity planning is high on the list of needs. Better predictability for space, power, and cooling capacity means increased useful asset life

spans and increased timeframes to react to future shortfalls.

6. Keen insight allows large amount of raw data to be transformed into business intelligence. Enhanced understanding of the present and future states of the data center allows for better asset utilization and increased availability.
7. Corporate Responsibility goals to assure the latest innovations in IT management are being considered, investigated, business cases are being created, and as a standard practice, major technological advances are not being missed.

TABLE 33.1 Reasons to adopt DCIM include capacity and asset management as well as increased availability and utilization

Reasons to deploy DCIM	%
Better capacity management	73
Better visibility and management of assets	35
Identifying availability-threatening problems	34
Increasing the utilization of IT assets	19
Improving data center customer service	14
Staff cost reductions	3
No plans to purchase DCIM	10

Source: The Uptime Institute [1].

33.1.4 Whose Job Is DCIM? IT or Facilities?

One of the most interesting aspects of the adoption of DCIM is the audience diversity and their individual driving factors. Traditionally, data centers were built, maintained, and utilized by different distinct organizations: (i) the Facilities organization that took care of all space, power, and cooling requirements for any piece of real estate, and (ii) the IT organization that took care of the data center physical and logical build and all of the equipment lifecycles itself.

IT and Facilities now find themselves required to work together for planning and optimization processes. Decisions about equipment and placement are now being made jointly. In many companies, both organizations now report to the CIO. DCIM enables the data center resource capacity planning to be managed over long periods of time. That said, DCIM and all of its capabilities will (in most cases) be driven by the IT organization just as it has for every other type of data center management (Systems Management, Network Management, etc.) over the years. Sure there will be benefits across all groups once DCIM is in production, but the IT organization tends to have the largest role and experience in enterprise-class software selection, so it is expected that the deployment and organizational integration of DCIM is best leveraged as an extension to existing software management frameworks driving data center logical operations already. In a few years, as the DCIM implementations have become mature, this layer will play an instrumental role in matching supply and demand, assuring just the right amount of processing resources exist at each point in time.

33.2 TRIGGERS FOR DCIM ACQUISITION AND DEPLOYMENT

While the term "DCIM" has only been used in the vernacular for the past few years, the concept of asset management has been around since the inception of the data center. Traditional approaches to data center asset management were fairly straightforward extensions of the financial bookkeeping tools in use. Earlier asset management methodology simply built upon the accounting systems of the period through the addition of physical attributes and organizational ownerships. In a few cases and with a dedicated desire by a handful of IT professionals to innovate, a bit of rack and floor visualization was added to that information. As these minor extensions provided little new business value, the adoption of asset management solutions which embrace energy, visualization, and lifecycle capabilities languished. These pre-"DCIM"-type solutions remained a curiosity, a nice to have set of features, rather than a must have business need.

As shown in Table 33.1, the highly acclaimed Uptime Institute did a survey in May 2012 regarding the top drivers for the adoption of DCIM [1]. Not surprisingly, the top reasons included a desire for better capacity management, better visibility and lifecycle management of assets, support for resource availability goals, increased asset utilization, improvements in customer service performance, and finally staffing related savings either through reduction or repurposing of existing personnel. None of these reasons are surprising in the new context of running the Data Center like a business.

33.2.1 Capacity Management including Power, Cooling, and Floor Space

Power was first major DCIM trigger on everyone's list. The rising cost of power was being seen in all aspects of life, both residential and commercial. Individuals saw the price of gasoline and electricity rise, corporations saw their huge power bills become larger. IT is typically the largest single line item in a corporation, so the abnormal rise in these highly visible costs caused a stir. The CEO and CFO leaders began asking questions about true IT costs, which the CIO and their teams were unable to answer. Power is one of the first quantifiable values that are directly associated with a successful implementation of DCIM.

33.2.2 Business Process Reengineering and Operational Efficiency

The operation of a data center is quickly transforming from individual and disconnected tactical activities with an historical primary goal of "high service levels at any cost" to a planned and predictable approach with the modified metric "service at what cost." Essentially, the cost factor is being added to the equation and being tested at every step of the way. IT organizations are being asked not just to document and then automate their existing practices, but to actually consider their current approaches and determine if they are still valid and/or optimal. As such, a number of organizations are finding themselves with limited awareness of their existing practices, which is impeding their ability to create streamlined new approaches. As baselines are created for existing conditions, IT organizations will begin to author new optimized workflows and deploy new technologies such as DCIM to manage their assets over long periods of time. DCIM promises to be able to capture current business practices, and allow optimizations in workflow and labor-related efficiency to be realized.

33.2.3 Data Center Consolidation Projects

Data center consolidation is reality today for most corporations for various reasons: advancements in computing technology, mergers, and acquisitions. DCIM supports the commissioning and decommissioning of vast amounts of computing equipment typically found in data center consolidation projects.

33.2.4 New Capacity, New Data Centers

Many organizations are realizing that their core data center assets are either past their useful lifespan or are simply not able to support their organization's rapidly increasing demands for processing due in part to their inefficient practices and wasted resources. The acceleration in the adoption of new business applications was never imagined to be at the current rate, so these inefficiencies have presented themselves across the board.

DCIM promises to address the quantification of current data center capacity, with a keen eye on capacity management over time. The data center itself provides computing resources; and when a large sample of time-based usage data is studied in combination with the demands associated new corporate initiatives, highly accurate data center planning is not only possible, but expected. DCIM quantifies data center capacity and allows it to be planned.

33.2.5 Data Center Cost Reductions, Enhanced Resource Efficiency

With the era of "Green computing" came the primary goal of reducing waste. Specifically focused on energy overhead, it has become quite popular to focus the majority of data center optimization projects on their ability to allow the data center to operate at a lower cost per unit of work. Green IT has been used as a catchall phrase to describe the more efficient usage of power.

33.2.6 Technology Refresh and Architectural Changes

A good number of data centers find themselves with large-scale technology refresh projects. These projects stem from the desire for higher density computing, virtualization, Virtual Desktop Infrastructure (VDI) initiatives, or mobility. Entire infrastructures are being redesigned, and when faced with this level of change, IT professionals find themselves looking for innovative ways to manage these new designs more effectively than previously practiced.

33.2.7 Environment and Sustainability Focus

There is a great deal of interest shown by most major corporations in reducing the impact of IT on the environment. Many organizations use three key metrics proposed by The Green Grid and covered elsewhere in this book to describe their efforts toward environmental friendliness: Power Utilization Effectiveness (PUE) relating to overall efficiency in the data center, Carbon Utilization Effectiveness (CUE), which refers to the Carbon footprint associated with energy consumption, and most recently, a metric associated with water, Water Utilization Effectiveness (WUE), which represents the amount of water consumed in the production of data.

33.2.8 Regulatory and Compliance, Audit and Documentation

The executive teams within major corporations across the globe have found themselves under new levels of scrutiny regarding IT. IT as the most critical corporate asset is involved in every major function company-wide. The impact of IT has become so great and pervasive that various government and regulatory agencies are striving to provide oversight to assure that data are maintained accurately and that the environmental impact of the data center is considered.

DCIM becomes a means toward this end. DCIM allows a data center to be documented as a single system, with the intricacies of its components identified and understood. The efficiency of the operation of each component can be seen and, over time, optimized.

33.2.9 The Cloud

DCIM fits everywhere! Public and Private Clouds share a common set of characteristics: Self-Service, Quick Provisioning, and Accounting. For a Public Cloud provider, scalable DCIM solutions are required to help quickly manage assets and dynamically tune supply and demand. A well-conceived DCIM solution is essential for the Public Cloud providers to understand all capacities (across IT and Facilities) and thus allow quick-turn for remediation, for provisioning, and for decommissioning. DCIM enables the data center to run as a business with all of these costs clearly quantified and optimized. DCIM allows Public Clouds to exist, to be more responsive, more accuracy in their operations, and reduce the overhead required to provide their end-user customers' required levels of service.

Private clouds are just traditional IT infrastructures that have been operationally transformed using the principles pioneered in the Public Cloud world. DCIM solutions are proving to be one of the most significant enabling technologies for this IT infrastructure reengineering. DCIM will allow this Private Cloud transformation. Remember that DCIM is all about enabling the data center to be managed like a business: comprehensive access to all of the business metrics, costs structures, services, etc. and dynamic management of assets. A comprehensive DCIM solution is essential for the transformation of traditional IT infrastructures to a highly tuned, optimized Private Cloud.

33.3 WHAT ARE THE MODULES OF A DCIM SOLUTION?

The most mature of today's DCIM solutions include all of the necessary functions to allow a fully functional production data center to be streamlined and support all of the required material provisioning, optimization, remediation, and documentation over time. Comprehensive DCIM suites are usually created as a range of functional modules that are intended to work together seamlessly. These modules offer various means to gather static and dynamic data, store this large amount of time-specific data, correlate the associated data, and then present and leverage this wealth of data in increasingly meaningful ways. When these tightly integrated modules are driven from single data repository, the resulting DCIM allows highly impactful business decisions to be made.

33.3.1 Asset Lifecycle Requirements and Change Management

DCIM enables lifecycle management of the data center and all of its assets. It addresses the physical layer of the data center and includes the same change management and workflow capabilities found in the other Enterprise Resources Planning (ERP)-class business management solutions found in the typical enterprise. DCIM is not just a monitoring utility, although there is monitoring within DCIM solutions. The biggest value of DCIM is not a monitoring utility. DCIM is as an enabler to manage change with a keen eye on the cost structures associated with this change.

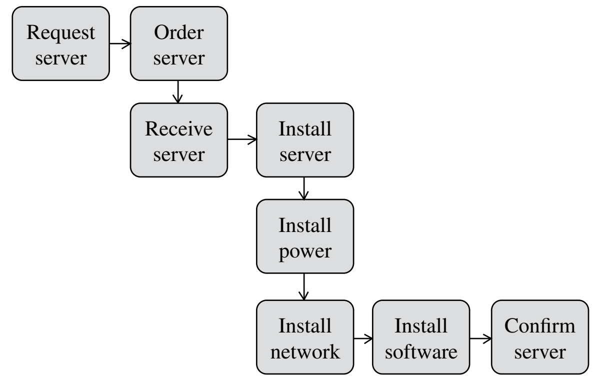


FIGURE 33.4 Leading DCIM suites offer comprehensive workflow capabilities that capture a data center's operational best practices to assure consistency.

Over the course of these years, it is estimated that at least 25–30% of the assets contained within any given data center change each year. Technology refresh cycles due to depreciation and maintenance costs, adoption of dense computing and virtualization, new networking or storage technologies, all account for huge amounts of change.

As can be seen in Figure 33.4, changing a single server seems relatively simple, imagine multiplying that effort by a thousand or ten thousand times each month! It is staggering. DCIM is the business management platform that keeps all of these Add/Move/Change cycles in order, documents the process at each step, and identifies the tasks needed to be completed in extreme detail to reduce human errors incurred during the execution of these tasks.

33.3.2 Capacity Planning, Analytics, and Forecasting

Of specific note to the DCIM opportunity discussion is its ability to consider the data center as a system, with a very specific set of metrics and capacity over time. Data centers have physical attributes and associates limitations. Whether it is space, power, or cooling, each data center has a physical set of limitations that define the limits of a data center's capacity. DCIM has already been shown to be the best way to look at these factors together, and then consider over time. With this ability to consider all resources over time, predictions can be made about when one or more of these critical resources will be exhausted and what cost will it take to bring new resources online.

The most successful DCIM offerings understand that visibility into the future is extremely valuable. It's quite easy to focus on historical data and present it in various forms, but interpreting historical data and using it to trend into the future is where mature DCIM offerings shine. Worth noting is the recent IDC findings that almost one-third of all data centers are forced to delay the introduction of new business services and more than a quarter of those data centers needed spend unplanned OPEX budget to maintain a poorly defined data center structure. These unrealized opportunity costs can be huge!

The DCIM model includes a highly granular representation of the data center, which enables it to identify where resources (power, space, cooling, and connectivity) exist and where they are being used. Over the years, many data centers have lost resources due to their inability to identify their exact location. Terms such as "stranded capacity" and "vertical white space" come into discussion when these conditions occur. Essentially, the originally designed resources become fragmented and therefore cannot be effectively utilized, or in other cases the availability of one resource is not coresident with similar capacity of another resource. A great example is a data center that wishes to deploy high-density blade chassis systems in an area with plenty of power but limited cooling. That power essentially becomes "stranded." The same types of imbalances occur across all of the data center resources. Modern DCIM solutions help by identifying when resources exist and allow balancing to recapture these resources. In some cases, this repositioning of equipment to better balance all available resources may add two or more years of useful life to existing data center structures.

33.3.3 Real-Time Data Collection

There are two major types of operational data that must be collected. The first type is the traditional "IT" devices and their virtualization components. These devices communicate most commonly using traditional networking protocols such as Simple Network Management Protocol (SNMP) or modern web-based Application Programming Interfaces (APIs) and include fairly well-defined templates that are embedded by each data collection utility which understands how to interpret the various values provided by the device itself. These devices report hundreds of data points, so the mapping of just those values needed by DCIM is critical.

The second type of device important to the DCIM solution is all of the components that form the Mechanical, Electrical and Plumbing (MEP) infrastructure. These include power and cooling devices typically found external to the data center, or those devices used to provide large volumes of power and cooling for subsequent distribution. This includes generators, battery backup UPS systems, large floor mounted PDUs, cooling chillers, and CRAC/CRAH units. These devices typically communicate with more challenging protocols such as MODbus, BACnet, LON, and in some cases older serial command lines via ASCII RS232.

In general, data center metrics useful to DCIM solutions are observed every few minutes by polling. In a few cases, there are triggered asynchronous events like doors opening, but the vast majority of this "real-time" data in a data center relates to temperature, humidity, pressure, and power, and those metrics are measured over longer periods of time, with analytics looking for trends over those same periods. Worth noting is that "real time" in the context of DCIM is not sub-second real time as the manufacturing world might define it, but instead typically deals with metrics observed over minutes or hours.

33.3.4 Integrations with Third-Party and/or Existing Management Frameworks, Web APIs

One of the requirements for DCIM solutions is their ability to connect to existing structures. Most IT and Facility organizations have deployed point management solutions over time. These solutions have formed the core of data center management for years. The strongest DCIM solutions will be those that provide connectivity to these solutions as well as a number of traditional business management applications to coordinate workflows and metrics in a meaningful fashion. These systems can provide a wealth of knowledge source, are critical to service desk and ticketing processes, and include all of the control hooks to the existing components. There are dozens of IT and Facilities systems that will be found across the many diverse corporate data centers, and DCIM vendors increasingly find their customers asking for these integrations. Integrations range from simple device access using standard protocols like SNMP or WMI to more complex web-based integrations of workflow and power-chain management. The integrations are seemingly endless and the strongest DCIM are accumulating inventories of these integration "conduits." Figure 33.5 shows just a sampling of major systems that will ultimately be connected over time to perform the DCIM function. Prospective customers should consider these inventories of "off-the-shelf" conduits from each vendor when making their choices.

33.3.5 Discovery Services and Location Services

33.3.5.1 Discovery Services. What Devices Do I Have? Discovery Services can be thought of as the LOGICAL discovery of active assets on a network. This active asset discovery can be deployed to identify or confirm the presence

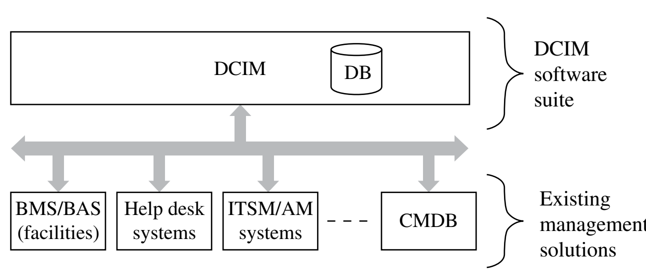


FIGURE 33.5 Modern DCIM deployments will require deep integrations across external enterprise management applications. Ecosystems of supporting technologies are forming.

of devices upon the network, and then advanced reconciliation techniques can be used to assure that the DCIM asset model matches the reality or what is physically installed and vice versa. How various DCIM vendors handle this reconciliation between what they model and what they logically discover is based upon their maturity as a solution.

Once logical addressing has been confirmed, active asset identification can occur. Since there is no single approach programmatically to determine the specific make and model and configuration of those devices, various technical approaches must be used to determine their specifics and configurations. These approaches leverage a number of protocol interfaces including IPMI, ILO, DRAC, RSA, Serial, RPC, WMI, and a multitude of virtualization protocols.

Although a cumbersome process, active asset identification has been done today. A number of DCIM enhancement start-ups have created mission statements based solely on their ability to interrogate active devices and then using a combination of table-lookup and metric retrieval to accurately identify each device and its configuration.

33.3.5.2 Location Services. Where Is Each Device Installed? Sometimes grouped with Discovery Services, asset Location Services is part of important theme of the DCIM segment. While logical detection of devices on a network has always been available using Discovery Services as mentioned earlier, there is no easy approach to detect where an asset is located physically. Essentially, modern data centers must still rely on mostly manual audit and widely diverse documentation to identify the installed locations of data center assets.

Various vendors have brought forward their versions of physical asset Location Services, each requiring customized hardware add-ons of various complexities. Some of these systems identify physical asset placement at the high granular "rack-unit" level, while others are less specific and can identify regions where assets currently exist. Worth remembering is the various low-tech approaches such as barcode technologies that have been prevalent for the past 25 years and are still in use today to track assets. In some cases, these traditional approaches have been adapted to become part of the DCIM solution and tend to offer the granularity needed for asset location tracking purposes.

Over time, there would appear to be a significant need for a standardized approach to determining specific asset location using an agreed industry standard. If such a capability becomes available, this would enable all manufacturers of IT gear to release hardware devices that have the ability to identify themselves and their placement in a structure mounting system for data centers. That said, the emerging new rack form-factor being put forward by the Open-Compute Project (OpenRack) still does not include this type of location awareness, so the wait for this capability to become reality will likely be an extended one.

33.3.6 Data Import and Export

One of the important features when implementing any DCIM solution is the ability to gather and normalize existing sources of asset and connectivity data. In a complex data center, there may be hundreds of thousands of individual pieces of data that would otherwise have to be manually entered or recreated through some means. In general, the labor cost to establish this knowledge manually without using any data import will exceed the cost of the DCIM software license itself, and in some cases may actually be twice the cost of the software license. Hence, the significance of data import innovation is a critically important part of the DCIM solution.

In response, most DCIM vendors include some means to import data sources such as spreadsheets and text files. Each vendor takes a different approach to importing and includes varying degrees of intelligence during the import process.

Most mature solutions use advanced field and pattern recognition and will even handle fairly well-defined types of problem resolution during the import process to map to existing source files. These files vary in format, and the error corrections available during the importing process may include missing information lookup, sequential missing data replacement, asset field de-duplication, proper handling of structured cabling range conventions, and a general ranking of data fields based on overlapping sources.

Data import is typically a critical component of any DCIM solution at the time of deployment. It is most commonly used once, and then the previous means to track and maintain asset knowledge are abandoned in lieu of the production DCIM solution. The most effective DCIM implementations allow the DCIM suite to become single source of truth about assets, once put into production.

In a related topic, some DCIM solutions also enable the EXPORT of data to industry standard file formats such as CSV or XLS. These exports may include some or all of the DCIM database information, and tend to be large files used for ad hoc analysis to feed into other systems transitionally. DCIM solutions that include the export functionality can usually recreate the entire main database using this same file as an import.

33.3.7 Materials Catalog and Library

All DCIM solutions are designed to manipulate asset material lifecycles, their placements, and their connectivity. In the creation of the physical structure, various types of devices must be selected from this catalog and then used throughout the DCIM modeling process.

Most vendors of DCIM solutions supply material libraries with 5000 IT devices or more. It is the extent and means to enhance this library that will define the success and ease of use when attempting to articulate the current complexion of the data center faithfully.

The materials catalog includes representations of devices and includes the manufacturer's specified parameters for each device. These parameters typically include high-resolution renderings of the front and back of the devices, power requirements, physical dimensions, weight, connectivity, etc. In the case of complex devices, the material catalog also includes options that may be installed (i.e., power supplies, and interface cards). All of these materials must be supplied by the DCIM vendor, or must be created manually by the end-user, which is a huge undertaking. Some vendors offer the ability to request these new devices to be added "on demand" (and typically within a week or two), while other DCIM vendors require the user to create these special new devices themselves. In a few cases, the DCIM vendor provides both mechanisms to enhance the materials catalog.

33.3.8 Rack Planning and Design

One of the most visual features of any DCIM solution is the ability to create faithful representations of equipment racks and their installed gear and associated connectivity. In fact, it is the visual representations of these racks elevations that typically attract some of the most enthusiastic initial interest by data center managers regarding DCIM. When considering DCIM vendors, a great deal of weight is given to the level of fidelity and absolute accuracy of the racks created by a given DCIM solution, and each offering is judged by its ability to most closely represent their real-world counterparts.

Most DCIM solutions use the aforementioned materials catalog as building blocks for rack design as shown in Figure 33.6.

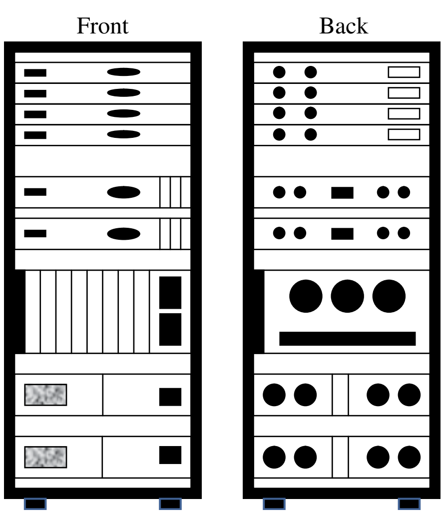


FIGURE 33.6 Rack planning tools allow a highly accurate representation of all installed assets, providing front and back views as well as the cabling interconnection of those devices.

33.3.9 Floor Space Planning

The floor of the data center is essentially an X–Y coordinated grid used to identify the actual location of equipment racks and other free-standing data center gears. Floor planning is a critical reference process as the floor of the data center must be designed to mimic the actual geometries in each data center.

Unfortunately, the data centers in use today are not always in simple rectangles. There are many types of construction and various obstacles which influence the placement of equipment and racks with the data center. The DCIM solution's floor planning component must allow these geometries to be captured accurately as they influence nearly all other aspects of modeling and planning within the data center if desired. Accurate positioning for every device and rack is a core requirement to realize the maximum benefit of DCIM.

Most DCIM offerings also include the visual representation of the data center at large using the floor planning component. Shown here in Figure 33.7 is an example of where we see various top-down representations of the data center, with the floor-tile systems, racks, CRACs, and other components shown in precise detail. These top-down views are also able to present metric data and aggregations using color-coded scales. For instance, they can represent the number of available rack-units, or total power consumption in each given rack. Using these visual representations of the data center, capacity can be visualized and new projects can be created based on the actual complexion of the data center as it sits currently.

33.3.10 Reporting, a Critical Part of the DCIM Story

One of the most valued capabilities for any DCIM solution is its reporting. Reporting is the way in which the raw information is correlated and then presented in a business impactful fashion. These reporting systems may also include a library of standard data center management reports and can typically distribute any of these desired reports to specific user(s) in an automated fashion. Other DCIM vendors simply include data store definition schemas and rely on their customers to design their required reports and then use industry leading reporting packages such as Microsoft Reporting Services, Business Objects, or SAS to create these desired reports.

33.3.11 Dashboards. A Picture Is Worth a Thousand Words

Dashboards tend to a special case of reporting and can be considered the "At-a-Glance" report. Dashboards have the ability to present vast amounts of information in easy-to-read displays suitable for desktop or operations "command center" consoles. Even though dashboards could be considered by some to be a "cosmetic" attribute of the DCIM suite, they are one of the first considerations new DCIM prospects look for when selecting a DCIM solution. Remember, the amount of

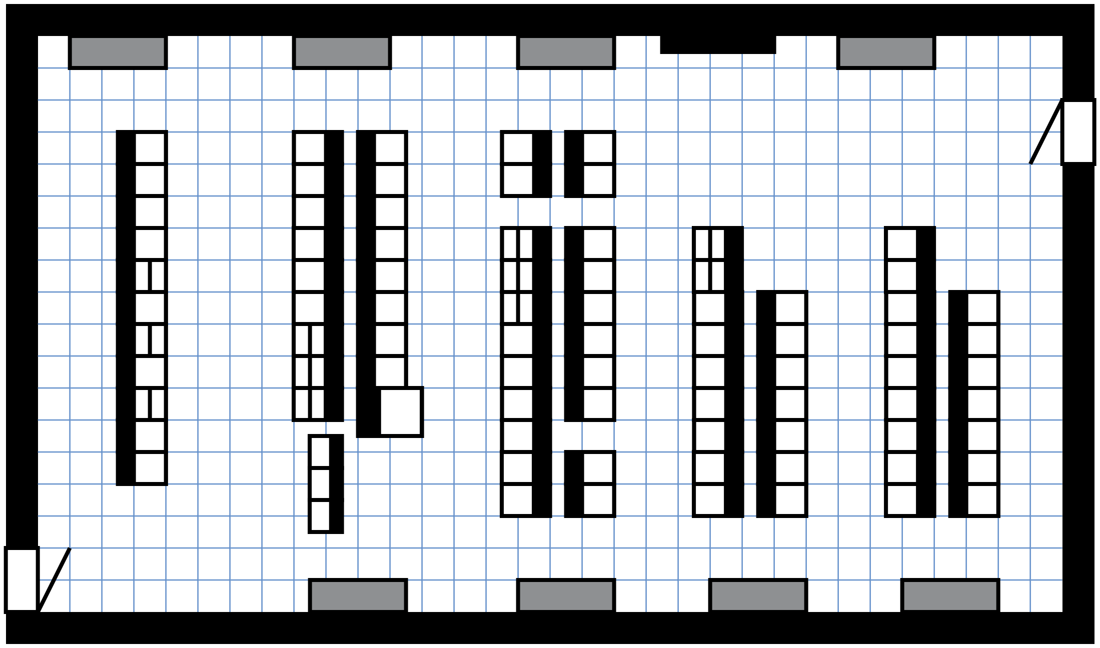


FIGURE 33.7 Data center floor planning enables efficient placement of racks and cabinets, accounting for service allowances and obstacles.

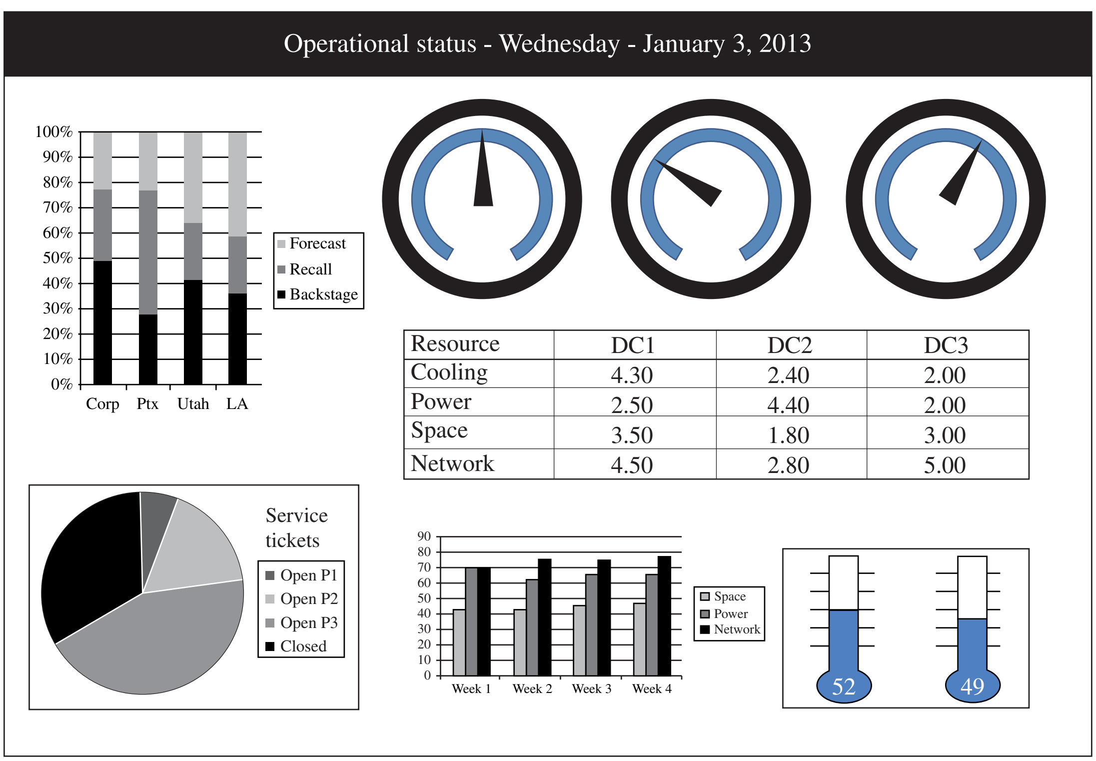


FIGURE 33.8 Critically important to the success of DCIM deployments, dashboards enable vast amounts of data to be visualized easily and in near real time.

data available within a DCIM solution can be enormous, and the ability to distill large amounts of this raw data into meaningful information that can then be presented using easy-to-understand visual dashboard elements is key to overall success. The ability to quickly access actionable information is a major value for DCIM deployments.

The vendor's included dashboards (Fig. 33.8) are a critical presentation of the DCIM function. With so many stakeholders in the proper operation of the data center, each will have a set of metrics that they hold themselves accountable to. Each has a set of needs which can be measured and derived from data found within the data center. Operations and Finance look at costs of equipment, depreciation, warranty, etc. Facilities professionals look for trends in power and cooling consumption.

33.4 THE DCIM SYSTEM ITSELF. WHAT TO EXPECT AND PLAN FOR

As we've seen, DCIM implementations provide a comprehensive set of asset management capabilities in the data center. The assets themselves have an enormous set of individual identifiers which are unique to each asset, ranging from physical characteristics and location, business owner attributes, and service information. DCIM technologies allow the Data Center assets to be organized in a variety of ways, which then allows solid business decisions to be made. In addition to these static attributes, the data center can provide a wealth of dynamic information that is derived from complementary technologies, sometimes referred to as "DCIM-Specialist" solutions. DCIM-Specialist solutions are available from over a hundred vendors today.

33.4.1 The Platform's Architecture

In this section, we'll describe the DCIM platform as well as the instrumentation layer that supports it. As stated previously, the DCIM data model is a living 3D model of assets, and as a general rule, the more connected this model is to the wealth of real-world instrumentation available, the higher the realized value from DCIM.

These early single-user developments were not bad choices in the pre-2006 timeframe, but most of these early DCIM applications are now being augmented or entirely replaced by well-behaved modern web-based versions that are deployed upon enterprise-class and IT-maintained business servers. "The Web" and all of the advanced communications and presentation technologies have provided a huge opportunity to create complex management applications that can be easily scaled and widely accessed from anywhere on the Internet. These modern DCIM offerings typically scale by spanning across multiple server engines, with each engine serving various user, data collection, storage, and analytics/reporting functions.

33.4.2 The Platform's Data Storage Model

A core component of the DCIM suite is a robust data storage model. Large amounts of data will be sourced from a wide variety of sources, much of it being time-series in nature, and all of it being required to be readily accessible for complex analysis and presentation needs. The data model itself must be robust enough to store data in a way that very complex analytics can be used across the data set interactively.

It is critically important for DCIM suites to have data models that are designed for interactive retrieval. Large volumes of data will be stored over time, and one of the key attributes of a strategic DCIM solution is its ability to present interpretations of vast quantities of raw data into meaningful metrics. While it may sound like a technical detail, the choice of storage approaches will directly affect the usability of the entire DCIM solution. Users will not tolerate the slow performance caused by data retrieval in a DCIM solution. Complex searches across gigabytes of data would take unacceptable amounts of time if the wrong storage technology was chosen. Imagine that every time you wanted to use an application on your smart phone, there was a 30-s delay before the first screen. You would likely NOT use the smart phone. The DCIM storage model, if chosen poorly, has the potential to have the same effect.

33.4.3 The Platform's User Interface

Modern DCIM suites are most commonly web-based and utilize the latest web-based access methods being adopted in common business management applications. The visual presentation of DCIM is complex and varies from vendor to vendor, but each shares a common goal of allowing easy navigation across a vast field of data by multiple users across the Internet.

The Graphical User Interface (GUI) can be considered one of the key attributes of a DCIM solution, as customer adoption many times is directly related to the intuitive nature of the GUI. A great example of an intuitive GUI interface is Google's Earth application, which allows the untrained user to start with a map of the entire planet and within seconds zoom into a view showing the house where they live.

The user interface for DCIM is critical. These applications must be highly intuitive. Large populations of any combination of IT and Facilities equipment spanning over thousands of square feet must be quickly accessible. DCIM enables the relationship between components to be clearly articulated in great detail.

33.4.4 Instrumentation: Sensing the Physical Components in Real Time

Modern data centers can provide a wealth of information about the current status of everything from the power chain and cooling status to the performance of the servers and virtualization layers. It's all fit into a category that is leveraged by DCIM called "instrumentation." Instrumentation is essential for a DCIM solution to be effective, and it includes a wide range of technologies and protocols, each intended to gather a specific portion of the entire infrastructure. The most mature DCIM suites expect many of these subsystems to exist and the DCIM systems themselves deal with the normalization and presentation of this instrumentation data. A point of reference regarding scale is worth noting here, as the magnitude of data gathers using various means of instrumentation can be massive. In a typical small data center with 100 racks and 1000 servers, tens of thousands of data points per MINUTE can be generated!

33.4.4.1 Environment Instrumentation: Temperature, Humidity, and Airflow Sensors One of the earliest arrivals on the journey to DCIM has been the environmental sensor vendors. For years, environmental sensors in the data center were considered merely as a "nice-to-have" tool by data center operators. As such, their usage was limited to a relatively small population of these operators. Some of the reasons given for low adoption included the perception as a relatively high-cost solution to an otherwise simple set of needs, nonspecific use cases, installation cabling complexity, and lastly the associated costs and limited pre-DCIM business management value. Environmental sensors were not considered a strategic source of knowledge within the data center.

The "American Society of Heating, Refrigerating and Air-Conditioning Engineers" (ASHRAE) has published guidelines over the last several years in support of new ways of looking at and optimizing data center cooling. Their recommendations on sensor placement have provided resurgence in sensor innovation and, in fact, a handful of startups have been formed to meet these ASHRAE-inspired needs for easy to use environment sensors to support DCIM deployments. These stand-alone environmental monitoring systems are available in wired and wireless variants. These purpose-built systems fulfill the need to understand the temperature and humidity of a data center and can do so at the granularity recommended by ASHRAE, if desired.

Wired environmental systems were the first entrants into the data center market and usually consist of purpose-built micro-PC hardware with some form of micro-operating system within, and all of the necessary analog I/O hardware to monitor temperature and humidity, and perhaps read and control dry-contacts and relays, listen to or emit sound and alarms, sense light, etc. These devices are connected to a LAN port anywhere in the data center and all interactions with these devices are done using standard web and IP-enabled protocols. This "connected" approach requires significant deployment complexity where cabling can become costly and prohibitive.

A second type of environmental sensing system has emerged, which addresses implementation simplicity through the use of wireless. Wireless systems can be either AC-powered and commonly use 802.11 (WiFi) or battery-powered using 802.15 (e.g., Zigbee) or Active-RFID technologies. Powered wireless devices operate as long as AC power exists and due to their physical power connectivity tend to behave much like a wired sensor solution, gathering and communicating larger amounts of sensor information much more frequently. Powered wireless devices allow network connection wirelessly, but the requirement for AC power itself makes these "wireless" solutions, something less than truly wireless. Worth noting is a continued concern by many data center operators prohibiting the use of WiFi in the data center for security reasons, since opening WiFi channels for instrumentation also allows any other type of network access using the same WiFi channels via PCs and handhelds alike.

The new generation of battery powered wireless monitoring devices that can be quite impressive in their ease of deployment and true to their name are truly "wireless." These battery-powered devices are highly engineered to consume less power by significantly limiting the amount of data transmitted by employing highly intelligent data manipulation and de-duplication, reporting only changing sensor values each period of time. These low-power battery-powered wireless devices tend to be unidirectional, reporting changes upward, but do not receive data of any type. Monitoring can be viewed as an upstream activity, so this approach works perfectly in the majority of DCIM supporting roles. Some of these devices have claimed battery lives in excess of 3 years!

Working together, a data center may include hundreds or thousands of wired and wireless sensors of temperature, humidity, and air pressure or flow. Each of these systems has its pros and cons, and ultimately DCIM installations will likely find use for a combination of these systems to support various portions of their environment.

33.4.4.2 Power Instrumentation: The Rack PDU The most basic build-block in a data center is the rack or cabinet which houses active equipment. Each rack may contain up to 40 or more active devices that require power (and the associated cooling) and connectivity. The number of racks in a data center may range from a small handful to well into the thousands. The scope of each data center varies widely, but what remains constant is the requirement for power in these racks. The means to deliver power to these devices is an appliance referred to as a "rack-based" PDU (or in some cases referred to as a power strip or plug strip).

While supplying power to the active equipment is mainly a function of the power capacity and number of outlets available, modern energy management focuses data center operators on maximizing the ways in which power is used and the efficiency in doing so with the ultimate goal to reduce costs. Elaborate power distribution strategies have been devised over the past 10 years to move power within a data center more effectively, taking advantage of some of the modern electrical utility's new approaches to supply raw power.

There are two optimization opportunities related to power: (1) Distribute power more efficiently through higher voltages and higher currents in smaller spaces and shorter distances, and (2) measure and monitor usage at a highly granular level, allowing individual components to be studied and analyzed over time.

DCIM provides the means to visualize these power chains and allows granular business decisions to be made. DCIM allows these power distribution approaches to be deployed, studied, and then actively monitored, to assure that loads are properly balanced and demand for available power and cooling is matching available supply.

33.4.4.3 Hidden Instrumentation: Server Intelligence DCIM solutions have the ability to associate large amounts of operational data for any asset to allow business decisions to be made. Various protocols are used to extract this information from servers, including IPMI, SNMP, WMI, and each of the virtualization vendors' own APIs. Nearly all active data center device support the use of one or more protocols to report their operational status.

In a typical modern server, storage or networking device, a wealth of knowledge is being made available today to external applications upon demand. This not only includes the more traditional logical operating parameters and performance metrics (such as CPU and I/O rates), but also typically includes physical device metrics, such as power consumption, power supply status, operational status, internal fan speeds, multiple temperature readings within each device, security lock status, etc. DCIM suites provide the unique opportunity to consider all of this physical and logical information together.

33.4.4.4 Building Instrumentation: Building Management Systems and Mechanical Equipment Usually referred to as MEP equipment, or simply "Facilities" equipment, a typical data center has a long list of equipment that becomes a critical part of any well-implemented DCIM solution. This includes the power generation and distribution devices, cooling components, and all of the control systems that have been deployed over the years to control these systems. The devices may be networked already, or they may be stand-alone.

The true promise of DCIM is to join the world of IT and Facilities, which allows all of the equipment required to provide computing services to be a useful part of the DCIM structure. Only with a complete picture of the IT and Facilities styles components can the maximum value be derived from DCIM. For example, a "power chain" consists of many links; the server's power supply, the in-rack PDU, the floor-mounted PDU, the data center UPS, the breaker panels, and the generators. Each of these forms a component of the power structure that is considered when making business decisions in the data center.

Today, Building Management Systems (BMSs) are a midpoint aggregation level and source of metrics for DCIM. Typically installed to control cooling resources, these "BMSs" can be fairly simplistic in nature and tend to have relatively few points of sensing, which in turn causes relatively macrochanges to the environment. In general, these systems are rigid in deployment and change very little over time. These systems can become a wealth of great information when integrated into a DCIM solution and, in fact, will allow DCIM suites to quite easily control cooling resources. Looking ahead a few years, we will see existing BMS systems augmented or supplanted by DCIM foundations combined with a layer of control and orchestration. Like BMS systems in the past, triggering events will make changes to the environment, only in the world of DCIM enablement, these triggering points will number in the thousands, and the types of control actions will be highly granular.

33.4.5 The Rack, the Most Basic Building Block of the Data Center

Data center racks themselves are the physical building block for the IT task. Typically, a physical cabinet is made of steel, and each rack is typically 6-ft tall, 2-ft wide, and slightly longer than 3-ft deep. Commonly, 42–48 devices may be housed in each rack, with larger or smaller numbers being seen in specific applications.

As standard building blocks, most DCIM offerings understand these mechanical designs and use very accurate templates for the selected rack(s). The size and shape of each rack is well understood and when coupled with floor-tile systems, allow an extremely accurate representation of the data center to be modeled. DCIM offerings use these building blocks as the basis for their high-fidelity physical topology representations, and rely on this ordered approach when depicting location and relative placements.

33.4.6 Remote Access and Power State Management

Related to DCIM has been the notion of remote access to systems. In fact, long before the DCIM marketplace emerged, the concept of managing the infrastructure was left primarily to two constituents: (1) Facilities managers who visualized power and cooling using purpose built BMSs and control panels and (2) systems administrators who used hardware and/or software tools to access their equipment remotely to power cycle or reconfigure operational settings.

The Facilities manager's ability to manage their power and cooling infrastructures has become quite mature. Highly advanced and completely customized visualization, dashboard, and control mechanisms have been created by the large building automation vendors. These BMSs are tailored individually for each deployment and tend to be quite functional, albeit extremely rigid. BMSs and their more advanced counterparts Building Automation Systems (BASs) have high price tags, and must be defined at the time of building construction in extreme detail by the Facilities

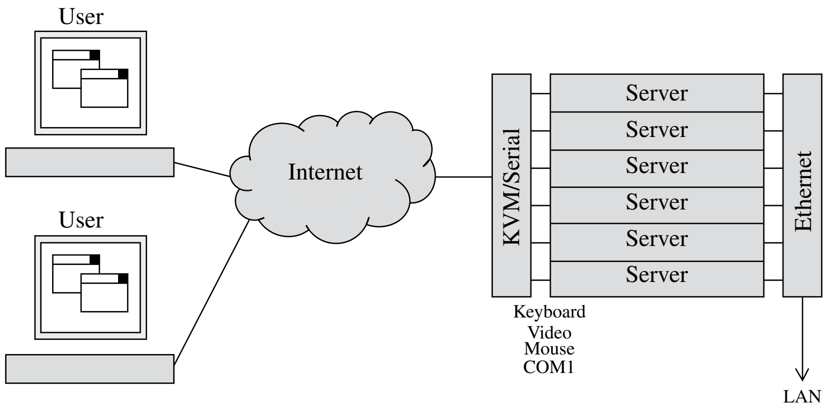


FIGURE 33.9 For years, physical infrastructure management was limited to Remote Console and KVM access technologies alone.

engineers who manage the building power, cooling, security, and lighting systems. Building engineers and mechanical designers work in concert to create these infrastructures, and then BMSs/BASs are tailored to reveal the inner workings and control capabilities. These systems tend to change very little over time and only when major Facilities construction changes occur do these needs get reevaluated and capabilities updated.

For the IT world, some of these remote management technologies are referred to as “power cycling,” “KVM,” or simply “console” and can be seen in Figure 33.9. Essential physical management of IT systems and data center devices was based upon the one-user-to-one-device approach using one or more of these technologies. This brute force approach to IT device management was directed at a single server, switch, or other type of IT system as a stand-alone management entity. The servers and other devices had no notion of placement and relative location or required resources, and basic metrics for energy consumption and temperature usually did not exist. Remote management technologies can be considered some of the most basic and primitive means of device management used for the past dozen years. The need for this type of remote access has been greatly reduced by hardware and software maturity found in enterprise-class equipment and today is most commonly found where a specific mission-critical device (and single point of failure) is deployed for specific functions. For these applications, power cycling and associated system reboot is the most common use of these remote management technologies.

Although not strictly required for a successful implementation of DCIM, these suites can usually take advantage of remote access technologies already deployed. These DCIM systems can allow traditional system administrators to share the user interface found within DCIM suites, and navigate to the system where remote operator access is desired, addressing configuration or reboot requirements.

33.5 CRITICAL SUCCESS FACTORS WHEN IMPLEMENTING A DCIM SYSTEM

The approach you use to implement a DCIM can make or break the long-term success of your project. As DCIM is a relatively new category, much of the waters that you and your team will be navigating will be unfamiliar territory and you’ll find a number of surprises along the way. Above all, you’ll need to keep reminding yourself about the goals of the project to not let it get away from you.

Here is a list of some of the critical factors that should be considered today to increase the likelihood of success in your DCIM project. While your mileage may vary, and every organization is different, there is a common set of steps found in the most successful DCIM deployments. Your DCIM journey will have many of these same steps:

1. Do your research. Read and talk to your peers at other companies that have invested in DCIM
2. Get buy-in from all of the Stakeholders. The four critical organizations to include in this journey are IT, Facilities, Finance, and the corporate social responsibility team.
3. Be realistic with setting Scope and Timing. Since DCIM technology is new to most people involved, there may be a tendency to over-simplify the complexity and resource requirements actually needed for deployment. Remember successful DCIM implementation require process changes, cultural changes, and training which is no small matter in larger organizations and those organizations that may have many different people involved.
4. Document Your Existing Processes and tools. Capturing the current operation of ALL of your data centers is the place to start. You’ll find a great deal of diversity from person to person, and center to center. The challenge is compounded when similar groups begin to negotiate

regarding the desired outcome of new processes that will be enforced by the DCIM solution.

5. Audit and inventory the assets that you already have installed. There are usually a number of sources of electronic data available (like spreadsheets, text files, and CAD drawings) which describe much of the current structure. It can be treated as a starting point, and through a combination of electronic documents and some good-old fashioned manual audit efforts, the structure currently in place can be easily described.
6. Determine your Integration requirements. Successful DCIM solutions are not stand-alone. They connect to the other Data Center management frameworks (such as building, asset, and work flow management) which may already be in place. Consider it all.
7. Establish a roster of users and associated security policy. Strategic business management solutions usually have many users since their value affects many organizations. When DCIM is implemented as a strategic component, larger numbers of users receive benefit. Finance, Technical, Facilities, Asset planners, and others all need access to the DCIM solution to enable the modernization of their previous tasks.
8. Determine each stakeholder's required Outputs (Dashboards, reports, etc.). The biggest mistake that could be made in a DCIM deployment is to assume one-size fits all. The needs of each user may be very different, so the reports and dashboards must reflect their individual needs. Even users that have common job descriptions may find their specific areas of interest to be unique.

Two important factors worth more discussion and are as follows.

33.5.1 Selecting the DCIM Vendor

Look for a DCIM vendor with a vision that matches your own. Obviously, every data center strategy is unique across the industry, but there will be well-formed thoughts about capacity planning, operational excellence, energy management, disaster recovery, etc. that must be discussed prior to choosing a DCIM vendor.

Take into account how long the vendor's solution has been available and how many installations each vendor has. Obviously, more is better as it supports and defends the vendor's approach to DCIM and will ultimately help steer your choice. Each vendor's installed base will have provided a priceless resource of users requirements from similarly situated users who have walked down the same paths you are GOING to walk down. Vendors should be able to share existing customer names and contact details or arrange for discussions with these customers on request.

Consider each vendor's recommended platform, architecture, and integration capability. Will the new solution be able to be integrated with the other systems that you have in place today? Can the vendor cost-effectively deploy it at the scale of your IT structure? How do they handle many users, many assets, and many data centers? How do your geographically dispersed data centers affect the DCIM suite's performance and real-time monitoring capability?

Look for vendors that can provide NEW levels of visibility and analysis, in ways previously not available to you. You are not looking for a prettier way of doing what you can already do, you are looking for new business management insight to allow yourself to make more informed decisions, respond more quickly, etc. Visibility down to the device level is just the start of a solid selection, and what the vendor's offering does with that level of granular information is where the magic comes from.

Once you have a short list of vendors, you should require out-of-the-box, demonstrable capabilities. The DCIM marketplace is relatively new, and nearly all vendors want to please prospects. When looking for a DCIM vendor, you really want to consider which of these capabilities they can deliver today, and avoid the more theoretical discussion about what they could do given enough time and enough money. Engineering projects create orphaned installations, and can diverge so much from any commercial offering, that customers will be abandoned at the onset, and will not be able to take advantage of the selected vendor's future releases of their software. As a rule of thumb, if they can't show you specific DCIM features, they probably don't have them built yet. Be cautious here as this will determine your long-term costs for DCIM.

Last, ask about pricing models. Be specific. Software vendors are notorious about turning what was presented as a product into a time and materials project. This can be a costly approach. Choose a mature DCIM vendor that details their cost structure, which pieces are off the shelf, and which are custom. They must also articulate ongoing maintenance costs.

33.5.2 Considering the Costs of DCIM

This is one of the most misunderstood topics when discussing DCIM, since the definition of DCIM is so diverse. As mentioned earlier, there are a handful of management software suites that comprise the top-level DCIM functionality. This enables the business management aspects of the deployed solution, and it is the most common interface that users will interact with. All of the rest of the offerings in the DCIM space are actually subcomponents of the total solution. Gartner refers to these vendors as "DCIM-Specialists" or simply "enhancements" to the DCIM solution. These enhancements include hardware and/or software that provides real-time data about power, or environments, allows deeper analytics or customer presentations, or even the ability to discover and identify various assets and their locations.

Today, there is no single pricing scheme for DCIM. The DCIM software management suites are priced in a wide range of schemes based on size or capacity, with perpetual and subscription licenses further complicating the process. Although different pricing schemes exist, for comparison purposes we can use a unit of measure at the "rack" or cabinet.

DCIM enhancement components on the other hand are much more straightforward in pricing. Sensor vendors, for instance, can tell you exactly what a thousand sensors would cost; and if you plan to use four sensors per rack, you can do the math to determine what these 250 racks would cost to outfit with a DCIM-Specialist vendor's sensors.

So, what does it cost when you are looking to budget DCIM for an upcoming project? As a general rule of thumb, based on the value of dollars in 2014 DCIM Suites and their natural DCIM-Specialists (enhancements) should be budgeted at about US$1000 per rack. This will include core DCIM core functionality, basic integrations with common systems, real-time sensors, installation, and training. (Intelligent rack-based PDUs will add another $1000 per rack, if intelligent power metrics are desired.)

33.5.3 Other DCIM Considerations

DCIM solutions are coming of age and are almost at a level of maturity that large and small organizations alike can begin to take advantage of this new area of data center management. It has never been easier or timelier to create an extended view of the data center infrastructure, by extending the logical views already deployed with physical layer extensions found in modern DCIM offerings.

A few points worth considering as you begin to investigate and then formulate your DCIM plans:

- What is your adoption timeframe, or can you afford to do nothing?
- What are your existing sources of truth and other documentation used in production today?
- Who will be the owner of this project and what resource are committed to it?
- Once DCIM is deployed, where do all the existing IT Support people go?
- Acknowledge your DCIM needs and capabilities will evolve over time.

33.6 FUTURE TRENDS IN DCIM

The DCIM marketplace is rapidly progressing as a management category. Whereas most efforts underway today allow a highly granular means to maintain and present an accurate representation of the existing IT and Facilities infrastructures (complete with real-time metrics), the future of DCIM will include (i) consolidation and/or rationalization of vendors solutions, (ii) new leveraging features including automation and control, and support for asset location and auto-discovery technologies, and (iii) Ecosystems approach where specific cross-vendor integrations will be formed using more standardized approaches to integration across these related infrastructure management solutions.

33.6.1 Consolidation and Rationalization of Vendor Solutions

DCIM has been an emerging technology of increasingly high interest since the mid-2000s. Referring back to Gartner's "hype cycle," we saw the peak in vendors of any type of DCIM during the 2010 timeframe where more than 100 vendors self-declared their participation in the DCIM marketplace. Partly because DCIM continues to be poorly defined across the industry, and partly because it is viewed as an emerging greenfield for management vendors with few incumbents, the potential DCIM customer has been bombarded by these vendors all making overlapping and, in many cases, unsupportable claims. There has always been a bit of "yellow journalism" afoot in the DCIM space.

It is clear this cannot continue. The value of DCIM is too great to allow the market to continue to be fragmented and confused. Recently, The 451 Group has attempted to provide some guidance into this confusion and have begun to try and define various aspects of the value possible with DCIM by showing how it relates to Service Management. They have introduced the term Data Center Service Optimization (DCSO) which can be thought of as a benefit-oriented superset of DCIM. From their definitions, DCSO systems are used to plan and optimize datacenter resources and services for availability, agility, and financial, operational, and energy efficiencies. Physical and virtual resources include critical systems, assets, power, compute and IT services, and applications.

Other analysts will surely follow this trend and begin getting much more specific about the business value of DCIM solutions rather than focus primarily on product naming.

33.6.2 Automation and Control

Control is a broad topic that will transform DCIM suites from visibility and analysis solutions into well-orchestrated business and workload management solutions that focus on dynamically adjusting ALL resources that are required to meet computing demand. While there are a few functional DCIM systems that offer a level of hardware control across a chosen physical reference design, the promise for DCIM is that true multivendor automation will occur across the entire hardware and software platforms found in the common data center, which will rely on a highly functional DCIM layer itself. DCIM with control plane capabilities will come of age over the next 10 years. While there are a number of startups today that focus on these automated approaches to dynamic capacity management (cooling and processing), the overall market has not yet embraced rallied behind these concepts as mainstream, and there are several active community discussions which describe commercial data centers which are operational in the year 2020 and beyond as being completely self-healing and dynamic in capacity, where automation will precisely align supply and demand for computing, along with all of the physical resources required to do so.

33.6.3 Asset Location, Physical Discovery

Asset location is one of the Holy Grails of the DCIM market. Today, there is no standardized way to determine where an asset is physically. In the logical world, active devices can quite easily be detected and interrogated to determine where they are on the network, the type of device, and the services which are running. In 2014, the Open Compute Project proposed a new rack platform, the first new rack design in more than 25 years. This platform once again does not include the built-in ability to physically identify down to the physical location slot where devices are populated. While the Open Compute industry is still working on their final designs, only time will tell if/when asset location will become a reality in any form-factor. Until then, innovative startups will continue to look for more ways to retrofit standard racks to enable this capability through the use of wired, wireless, and optical interrogation technologies.

33.6.4 Ecosystems and Integration "Standards," Linkages to Other Systems

The DCIM marketplace is poised for strong partnerships to form. Potential customers are looking for the DCIM vendor community to consider all of the strategic pieces required to demonstrate core data center management value, and then seek out those portions which they do not make themselves. Prospective customers of DCIM are looking for the "heavy lifting" for integrations to be done by the DCIM vendors involved. It is no longer enough to hide behind standard statements that speak about "protocols" such as "SNMP" or "WebAPI" as their sole approach to integration. Experienced IT professionals understand that general purpose support for standard interfaces is a far cry from systems that can seamlessly work together. These potential adopters of DCIM are looking for strong Ecosystems to form.

33.7 CONCLUSION

Real DCIM is available now. Whether purchased as an on-premise perpetually licensed software offering, or via a cloud-based SaaS variant, DCIM is available and affordable at any scale today. The management of physical aspects of the data center has been a fragmented and poorly understood science for the past decade; and as such, this physical layer of management has been ignored altogether, or addressed historically by over-provisioning ALL resources. The general guideline in the past was to simply create such an abundance of Data Center resources so that the upper limits would never be tested. Only with the recent and dramatic rising costs of power and the rapid movement to virtualized dense computing has the attention to the gross inefficiency associated with over-provisioning been scrutinized. Building data centers that are severely over-provisioned is no longer considered a strategic plan, even in the context of availability and uptime. As it turns out, the shareholders and stakeholder are demanding that their resources be used wisely with a level of defendability for each action.

As such, the CFO/CIO and even CEO executives are searching for the next phase in cost-effectively managing the data center, to include a coordinated IT and Facilities costing and service delivery model. While the Cloud and Virtualization technologies have their own unique impacts and opportunities to their data center strategy, the entire hybrid structure will benefit from DCIM in very tangible ways. DCIM is here to stay, and the most competitive organizations will begin to execute aggressive plans to start leveraging these new capabilities today. Those same organizations will quickly realize that DCIM solutions provide one of the most foundational means to support their business agility needs and enable new applications to be more quickly deployed.

REFERENCE

[1] Uptime Institute. May 2012. Data center annual survey. Available at http://uptimeinstitute.com/2012-survey-results. Accessed on May 27, 2014.

FURTHER READING

Azevedo D, Belady C, Patterson M, Pouchet J. September 2011. Using CUE™ and WUE™ to improve operations in your DataCenter. The Green Grid.

Belady C. October 2007. The green grid data center power efficiency metrics: PUE and DCiE. The Green Grid. WP #06. Available at http://www.thegreengrid.org/Global/Content/white-papers/The-Green-Grid-Data-Center-Power-Efficiency-Metrics-PUE-and-DCiE. Accessed on May 27, 2014.

Belady C. 2010. Carbon usage effectiveness (CUE): a green grid data center sustainability metric. The Green Grid. WP #32. Available at http://www.thegreengrid.org/~/media/WhitePapers/CarbonUsageEffectivenessWhitePaper20101202.ashx?lang=en. Accessed on May 27, 2014.

Blackburn M. January 2010. THE GREEN GRID data center compute efficiency metric: DCcE. The Green Grid. Available at http://www.thegreengrid.org/~/media/WhitePapers/DCcE_White_Paper_Final.pdf?lang=en. Accessed on May 27, 2014.

Cappuccio D. March 2010. DCIM: going beyond IT. Gartner ID: G00174769.Gartner Inc., Stamford, CT.

Cappuccio D, Cecci H. June 2012. Cost containment and a data center space efficiency metric. Gartner ID: G00235289. Gartner Inc., Stamford, CT.

Clark J. October 2011. The price of data center availability. Data Center J. Available at http://www.datacenterjournal.com/design/the-price-of-data-center-availability/. Accessed on May 27, 2014.

Clark J. November 2011. The next data center real estate boom. Data Center J. Available at http://www.datacenterjournal.com/facilities/the-next-data-center-real-estate-boom/. Accessed on May 27, 2014.

Cole D. June 2011. Data center energy efficiency—looking beyond PUE. No Limits Software. Available at http://www.nolimitssoftware.com/docs/DataCenterEnergyEfficiency_LookingBeyond.pdf. Accessed on May 27, 2014.

Cole D. May 2012. Data center knowledge guide to data center infrastructure management. No Limits Software. Available at http://www.datacenterknowledge.com/archives /2012/05 /22/guide-data-center-infrastructure-management-dcim/. Accessed on May 27, 2014.

Data Centre Specialist Group. May 2012. Data centre fixed to variable energy ratio metric. Available at http://dcsg.bcs.org/data-centre-fixed-variable-energy-ratio-metric-dc-fver. Accessed on May 27, 2014.

EPA. June 2012. Annual energy outlook with projections to 2035. DOE/EIA-0383. Available at http://www.eia.gov/forecasts/aeo/pdf/0383(2012).pdf. Accessed on May 27, 2014.

Fichera D, Washburn D, Belanger H. November 2012. Voice of the customer: the good, the bad, and the unwieldy of DCIM deployments Forrester Research, Cambridge , MA.

Fry C. January 2012. Green data center: myth vs. reality. WWPI. Available at http://www.wwpi.com/index.php?option=com_content&view=article&id=13817:data-center-infrastructure-management-myth-vs-reality&catid=210:ctr-exclusives&Itemid=2701757 (iTracs). Accessed on May 27, 2014.

Harris M. June 2012. DCIM value: two halves make the whole!. Available at http://dcimexpert.com/2012/06/two-halves-make-the-whole/. Accessed on May 27, 2014.

Harris M. 2009. Taxonomy of data center instrumentation. Mission Critical Magazine Available at http://www.missioncriticalmagazine.com/ext/resources/MC/Home/Files/PDFs/WP-Taxonomy_of_Data_Center_Instrumentation-Mark_Harris.pdf. Accessed on May 27, 2014.

Howard C. February 2012. Hybrid IT: how internal and external cloud services are transforming IT. Gartner ID: G00231796 Gartner Inc., Stamford, CT.

IBM Global Technology Services. February 2012. Data center operational efficiency best practices. Ref: RLW03007-USEN-01. IBM Corp., Whiteplains, New York..

Intel. February 2012 Moore's law inspires Intel innovation. Intel Corporation. Available at http://www.intel.com/content/www/us/en/silicon-innovations/moores-law-technology.html. Accessed on May 27, 2014.

Kaplan J, Forrest W, Kindler N. July 2008. Revolutionizing data center energy efficiency. Available at http://www.ecobaun.com/images/Revolutionizing_Data_Center_Efficiency.pdf. Accessed on May 27, 2014.

Kumar R. July 2008. The six triggers for using data center infrastructure management tools. Gartner ID: G00230904. Gartner Inc., Stamford CT.

Mell P, Grance T. September 2011. The NIST definition of cloud computing. NIST. Pub #800-145. Available at http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf. Accessed on May 27, 2014.

Neaves R. September 2011. Moving the data center from chaos to control. Nlyte Software. Available at http://www.nlyte.com/german/doc_download/15-moving-the-data-center-from-chaos-to-control-white-paper. Accessed on May 27, 2014.

Ortiz Z. February 2012. The green grid monthly members webcast: productivity proxies, economizer survey, and forum 2012. Available at http://www.thegreengrid.org/en/events/Forum2012PreviewWebcast.aspx. Accessed on May 27, 2014.

Pultz JE. December 2011. Net IT out: Data Center Infrastructure Management (DCIM): new tools to monitor, manage and control power. Las Vegas Conference. Gartner Inc., Stamford, CT.

Pultz JE. February 2012. More than half of data center managers polled will likely be using DCIM tools in 2013. Gartner ID: G00231803. Gartner Inc., Stamford, CT.

Ravo K. 2012. A new power measurement standard, UL 2640, to reduce data center costs. Available at http://www.ul.com/global/eng/pages/corporate/aboutul/publications/newsletters/hightech/vol2issue3/4par/. Accessed on May 27, 2014.

Schreck G. December 2009. Put DCIM into your automation plans. Forrester Research, Cambridge, MA.

34

COMPUTERIZED MAINTENANCE MANAGEMENT SYSTEM IN DATA CENTERS

PETER SACCO
PTS Data Center Solution, Inc., Oakland, NJ, USA

34.1 INTRODUCTION

We're living in a fast-paced society fueled by technology with an acceleration rate that's nearly impossible to grasp. If a company isn't constantly improving and updating itself, then it is doomed for failure.

Dissatisfactory maintenance service can be a costly problem for any company. Performing tedious repairs on equipment because of improperly filed work orders is not only a time waster but also a money waster. In addition, so is buying extra parts because of unlogged/lost inventory.

As companies grow, it becomes harder and harder to keep up with operational maintenance in the data center. When does this machine need service next? When was this machine last serviced? Is the warranty about to expire? Through the successful use of a Computerized Maintenance Management System (CMMS), these questions will not only be answered but never will be asked again.

34.1.1 Why CMMS

The CMMS provides a complete and integrated managed solution for the operational maintenance of data center sites, from stand-alone data center facilities to tenant-space computer rooms and their related support infrastructure. CMMS is a cost-effective solution to maintain and/or track information technology (IT) and facility assets in the data center. The goal of the service is to help organizations increase asset life, track maintenance details, prevent equipment failures, predict equipment failures, improve labor productivity, reduce equipment downtime, minimize investment inventory, and lower the total cost of maintenance. The solution is a web-based application that tracks all assets by type, manufacturer, and age and manages the assets against standard manufacturer-based maintenance schedules. All schedules, tasks, dates of service, problems, issues, resolutions, and other maintenance-based required items are tracked for each asset through the software system.

CMMS software is the most efficient way to schedule maintenance, create and maintain an accurate and detailed inventory, and store important documentation. Paper records can be long and tedious and can create wasted company time, trying to search through boxes of files for a particular document. An effective solution to this is CMMS. This web-based software can store your entire company history, and documents can be searched for with ease. CMMS gives a user the capability to store information that can aid in quick and successful decision making. Other benefits include the following.

34.1.1.1 Improve Operations Using a CMMS solution can achieve a high efficiency in operations by minimizing failures and maximizing operating time.

34.1.1.2 Maximize Asset Performance Strategic asset management can maximize the performance of all assets.

34.1.1.3 Control Maintenance Costs Through the use of asset management, life cycle management, and preventive and predictive maintenance, a company can control their maintenance costs by maximizing asset life and minimizing downtime.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

34.2 CMMS BASICS

34.2.1 Architecture

CMMS is a service-oriented software application with a user interface for "real-time" maintenance management. The user can track and schedule maintenance, manage asset inventory, write and review various reports and documents, and process work orders.

34.2.2 Configuration

CMMS solutions must be configurable to meet the perpetual demands placed on the software. When entering asset information, companies need the ability to add fields that are relevant to that asset and hide the ones that aren't.

34.2.3 Operations

CMMS provides the means for its users to record and track their past, present, and future maintenance work as well as equipment documentation, warranties, and service contracts. Operational costs throughout an asset's life cycle can be lowered through the use of asset management, life cycle management, preventive maintenance, and predictive maintenance.

34.2.3.1 Asset Management Asset management records and manages asset data from installation to decommissioning. It allows a user to log equipment information such as asset type, model number, serial number, manufacturer, install date, commissioned date, start-up date, and asset remote connectivity links. The user can also log the location of an asset, spare parts for inventory control, warranty and service contracts, purchase details, and service history.

34.2.3.2 Life Cycle Management Life cycle management starts from the moment a company's needs and requirements are analyzed.

Life cycle management is the process of managing the entire life of a piece of equipment from the moment it is purchased to its disposal. This is useful when deciding to replace a particular piece of equipment.

The key goals of life cycle management are to cut costs and increase efficiency.

34.2.3.3 Preventive Maintenance Preventive maintenance can maximize asset performance/life and reduce unplanned downtime. Typical examples include changing air filters, replacing light bulbs, and wearing components based on life cycle analysis.

34.2.3.4 Predictive Maintenance The key goal of predictive maintenance is to prevent unexpected equipment failures through the use of condition monitoring. Statistical data is used to determine the future maintenance trend of a particular piece of equipment. Common examples include oil analysis, thermography, ultrasonic, vibration analysis, etc.

34.3 CMMS MODULES

CMMS applications are typically available as client-/server-based programs or as a hosted service. In addition, they are sometimes module based with each section including a core group of functionality that seamlessly integrates with the other modules.

34.3.1 Asset Management

Effective asset managing begins with tracking assets and ends with managing asset data throughout their entire life cycle. Consistent asset management includes location definitions, past and future work assignments, as well as cost history tracked over time to maximize productivity and extend asset life.

34.3.1.1 Location At the heart of asset management is an understanding of where the asset is physically. The most effective location management toolsets define not just a country, state/province/region, city, and/or street address but also details about floors, rooms, and even particular specific site notes that may exist unique to the asset itself.

34.3.1.2 Contacts Just as important as understanding physical proximity of an asset is an understanding of the individuals both within and outside of the organization that interface with the specific asset including:

- Who "owns" the asset from an organization management standpoint?
- Who is financially responsible for the asset?
- Who is responsible for contract management for the asset?
- Who are the various contractors that provide technical and maintenance support of the asset?
- Who are the individuals from the original equipment manufacturer (OEM) that supplied the asset?

34.3.1.3 Hierarchies The ability to organize assets into hierarchies allows a manager to roll up costs, performance, dependencies across systems, subsystems, and locations.

Location Hierarchy As previously stated, it is important to understand the location hierarchy for an asset. This is not only useful to understand where an asset is, but is also useful as part of effective staff management by being able to perform multiple tasks or services for a given region and/or area.

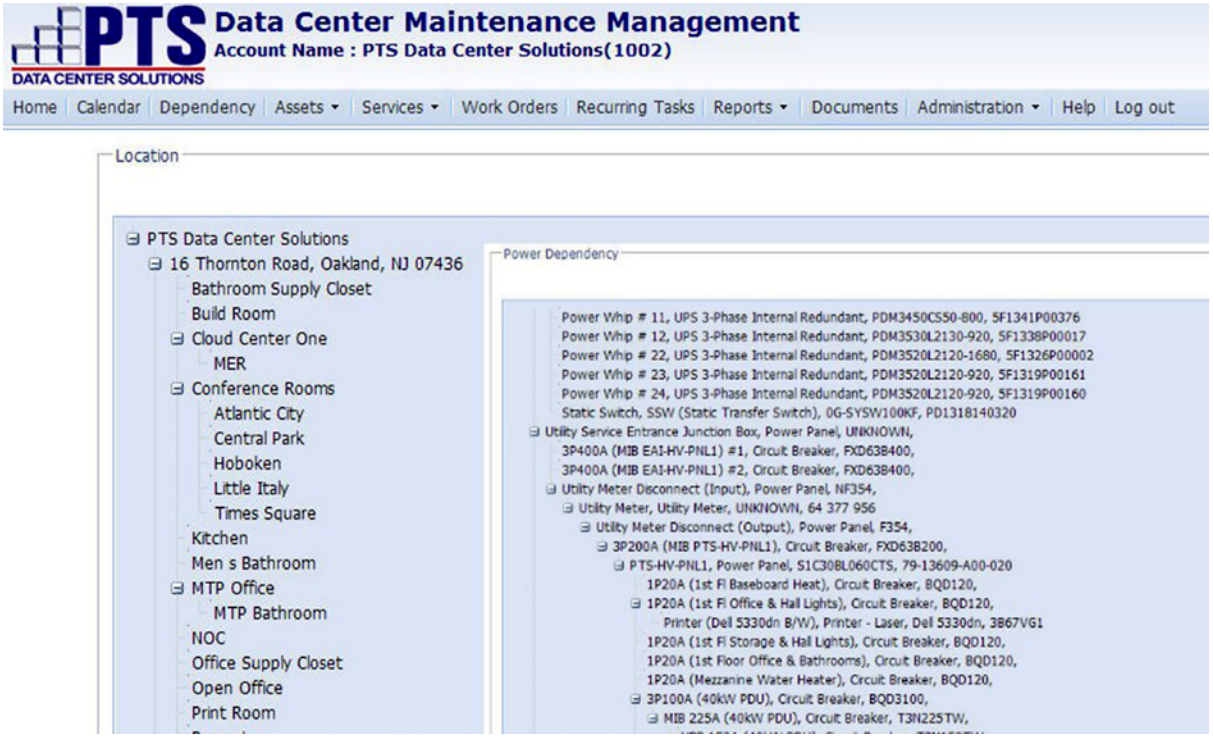


FIGURE 34.1 Location and power dependency. Courtesy of PTS.

Power Train Hierarchy For similar reasons, it is often useful to organize assets in terms of power train dependency (Fig. 34.1). In this way, an individual can visually understand how interruption of power of one device will cascade and affect all of the devices subservient to it. With this functionality, an asset manager can better minimize the affect maintenance activities will have on downstream devices as well as communicate potential disturbances to the "owners" of those devices.

34.3.1.4 Assets Assets are the various pieces of equipment that make up the supporting infrastructure for which maintenance is being managed. In actuality, assets can be of various types including facility-based infrastructure, data center supporting infrastructure, and even the IT infrastructure utilized throughout the organization. Regardless of the assets being tracked, it is important to gather accurate characteristics for all of the equipment. Each of these designations makes it easier to search and find specific information.

Asset Type Asset type is a designation that defines the equipment class in which an asset belongs. In the most useful CMMS application, a standard list of asset types is furnished. However, it is important to have the functionality to add new classes on an as-needed basis. Some asset types are as follows:

- Computer Room Air Conditioner (CRAC)
- Computer Room Air Handler (CRAH)
- Air-Cooled Condenser (ACC)
- Dry Cooler/Fluid Cooler
- Pumps
- Uninterruptible Power Supply (UPS)
- Generator (standby)
- Automatic Transfer Switch (ATS)
- Power Distribution Unit (PDU)
- Power Strip

Specific Asset-Type Information In addition to defining the asset class, it is important to gather specific configuration information unique to each individual asset type (Fig. 34.2):

- CRAC
  - Tonnage
  - Cooling Configuration (air, water, glycol, chilled water)
  - Operating Voltage

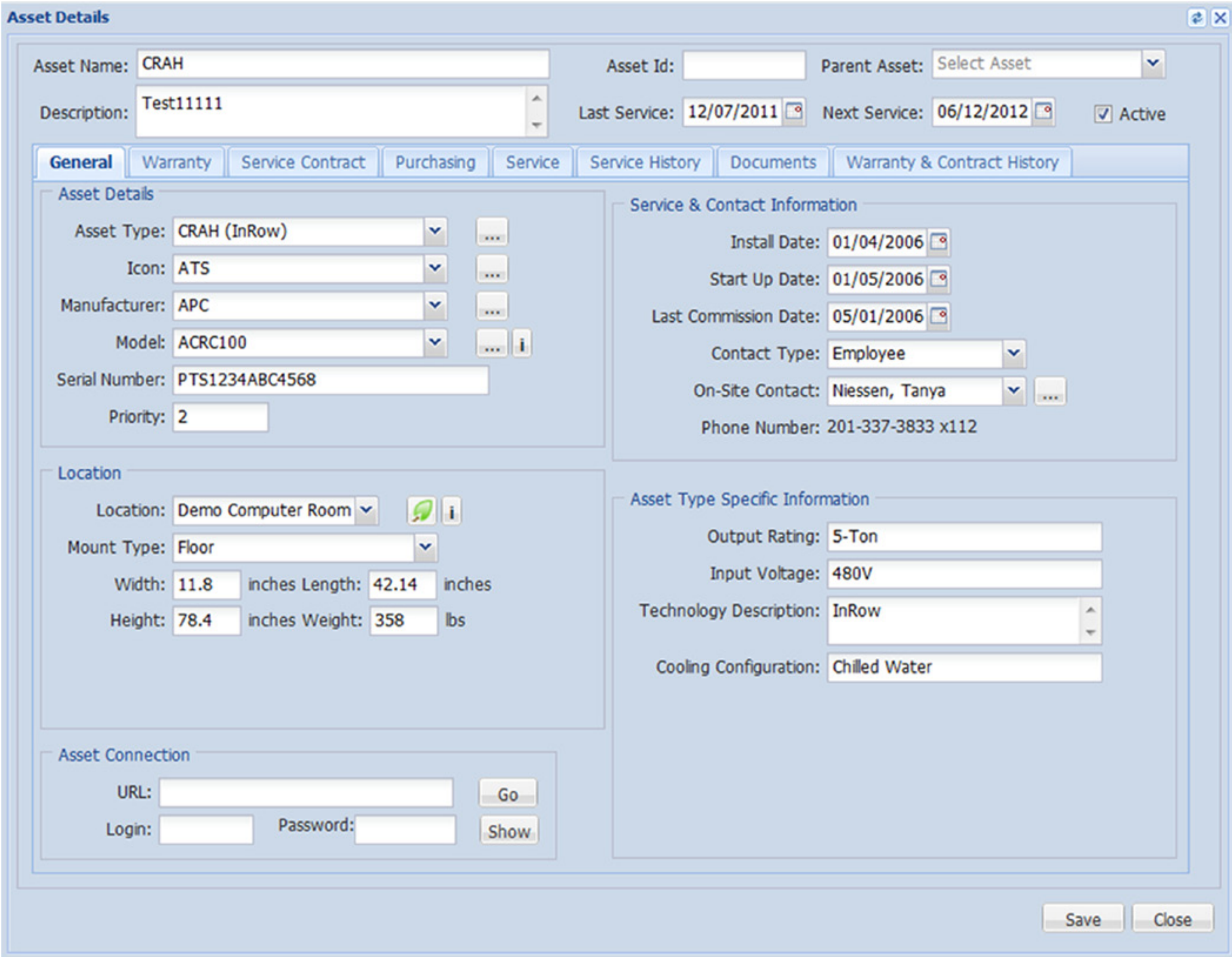


FIGURE 34.2 Asset details. Courtesy of PTS.

- UPS
  - Power Capacity
  - Input Voltage
  - Output Voltage
  - Topology

Manufacturer, Model Number, and Serial Number The manufacturer is the company that built the equipment. They are often referred to as the OEM. The model number is the identification used by the OEM to identify the specific equipment and configuration for the equipment ordered. The serial number is the OEM's unique identification code for the specific equipment ordered. Remember, while this number may be for a specific OEM, that does not mean it may not be an identical number for some other device from some other OEM. As such, it should not be utilized as the unique asset identifier in the application. Company asset tag with bar code could be used as identification number.

Install Date The installation date is the actual date in time the equipment was installed. This date is fairly subjective in that different individuals may interpret what "install date" actually means. Some people consider it as the date purchased, and some the date delivered, installed, started, turned on, actually used, etc. In any case, consistency should be utilized throughout the information gathering process. The best practice is to utilize the date the OEM considers as the warranty start date. This ensures an accurate account of the warranty start date. Again, each OEM has its own rules as to when the warranty starts, so the information should be sought out.

Start-Up Date As important as the install date is the date the equipment was started. The start-up date is a date in time that the OEM and/or the OEM's agent officially recognizes that the equipment was installed correctly, energizes, is configured properly, and is in good working order.

Commissioned Date The commissioning date is often different from the start-up date. Commissioning is the act of testing the equipment under real or simulated operating conditions to make sure the equipment is ready for its purpose in a live production environment. It is not uncommon for equipment to be recommissioned if substantial upgrades and/or improvements were made to the asset. As such, this information should also be logged.

Asset Remote Connectivity Links Asset remote connectivity links can mean a number of things including:

- A Bar Code Scan Code used to track the device using bar code scanning equipment
- A web URL and/or TCP/IP such that the equipment status can be accessed via an IT methodology remotely
- Some other physical or virtual link being utilized to collect and/or report live monitored data to some management platform

34.3.1.5 Material/Inventory/Spare Parts Management Material or inventory management is the process of recording all transactions, allowing for real-time knowledge of material status.

Inventory Control Useful functionality includes the capability of providing inventory transaction tracking. This streamlines parts and material management, which decreases costs by eliminating excess and obsolete inventory. Additional useful functionality includes inventory optimization and/or planning. This enables a manager to stock the right level of inventory to meet maintenance demand to ensure the right parts are available at the right location.

34.3.1.6 Contract/Entitlement Management Contract or entitlement management gives a maintenance manager complete control over vendor contracts. A useful functional tool is one that allows the correlation of an asset contract to a specific service level agreement (SLA) as detailed in the vendor's contract. In addition, a terms and conditions library is also useful to ensure consistent and standardized policies between contracts. It is desirable to have functionality that provides automatic notifications and alerts to aid in avoiding penalties and getting the most value out of every contract. Finally, it is useful to have payment schedule support functions to streamline workflow efficiency and strengthen vendor relationships.

Warranty A warranty is a written guarantee given to the purchaser of new equipment by the manufacturer or sometimes supplier, usually specifying that the manufacturer will make any repairs or replace defective parts free of charge for a stated period of time. As such, a warranty has important elements that need to be recorded.

WARRANTY DETAILS The specific limitations of the warranty need to be accurately captured including the start and end dates, as well as the exact terms of coverage. Different types of warranty coverage can include:

- Parts and labor
- Parts only
- A prorating parts schedule (one whose value diminishes over time such as is often the case with battery warranties)

WARRANTOR DETAILS The warrantor is the company that is actually providing the guarantee and who should be notified in case of need. As such, it is useful to have the ability to capture specific contact information for the administrator of the warranty.

WARRANTY HISTORY Warranties are usually for a specific term and can expire. Therefore, it is desirable to have the ability to retain information on old warranty contracts as well as the ability to assign new and/or extended warranties for each asset. This provides cost and term data retention, which is useful in the contract renewal negotiation process.

WARRANTY EXPIRATION ALERT NOTIFICATIONS As mentioned earlier, the ability to have automatic notifications and alerts aids in avoiding penalties and getting the most value out of every warranty contract.

Service Contracts A service contract is an agreement whereby a contractor supplies time, effort, and/or expertise instead of a good (tangible product) usually specifying that the contractor will investigate issues and make any repairs for a stated period of time. Sometimes, the warranty and service contracts are combined under a single contract. As such, a service contract has important elements that need to be recorded.

SERVICE CONTRACT DETAILS The specific limitations of the service contract need to be accurately captured including the start and end dates, the exact terms of coverage, and the cost model for coverage. Different service contract cost coverage models are available including:

- A fixed-fee contract for an unlimited level of technical support and on-site service

- A block of hours contract with a scope definition as to what maintenance services will be performed and over what period of time
- A time-and-material contract usually on a predefined rate schedule to perform services on an as-needed basis

In addition, different terms of coverage are available including:

- Best endeavor coverage
- Next business day coverage
- Next day coverage
- Maximum number of hours coverage including:
  - 24 h
  - 8 h
  - 4 h
  - 2 h

Service Contractor Details The service contractor is the company that is actually providing the services and should be notified in case of need. As such, it is useful to have the ability to capture specific contact information for the administrator of the contractor.

Service Contract History Service contracts are typically for a specific term and can expire. Therefore, it is desirable to have the ability to retain information on old service contracts, as well as the ability to assign new and/or extended service contracts for each asset. This provides cost and term data retention, which is useful in the contract renewal negotiation process.

Service Contract Expiration Alert Notifications As mentioned earlier, the ability to have automatic notifications and alerts aids in avoiding penalties and getting the most value out of every warranty contract.

34.3.1.7 Procurement Management Procurement management functions provide support for all phases of enterprise-wide procurement, including direct purchase requirements and inventory replenishment. In addition, processes for approved vendor setup and/or performance tracking tools eliminate costly off-contract purchases. Also, it is helpful to have functionality for automated materials requisitioning based on maintenance schedules. Finally, some applications provide automated interval, real-time metering, or event-driven purchasing capabilities to ensure the maintenance staff never finds itself without adequate parts needed during the maintenance procedure (MP).

34.3.1.8 Service History Obviously, tracking the service history for every asset is vitally important and is one of the key tenets of utilizing a CMMS tool for maintenance management. Key features for capturing service history are as follows:

- What asset was worked on?
- Where was it located?
- When did the work occur?
- Who performed the actual service work?
- What was the scope of services performed?
- Was the service work performed under an existing warranty and/or service contract coverage?
- What parts were utilized during the execution of service?
- Did the maintenance process incur any operational downtime?
  - If so, how much downtime was incurred?
- Was there a cost realized as a result of the service?
  - If so, how much cost was incurred?
- How long did it take to perform the service, and was it as expected?
- Who approved completion of the service work?
- Was all required documentation completed and retained?
- How has the completed service affected the statistical probability of individual component failure as well as the potential failure of any system including this component?
- How has this service affected any applied SLA?

34.3.1.9 Life Cycle Management Asset life cycle management (Fig. 34.3) is the set of business practices that join financial, contractual, maintenance, and inventory functions

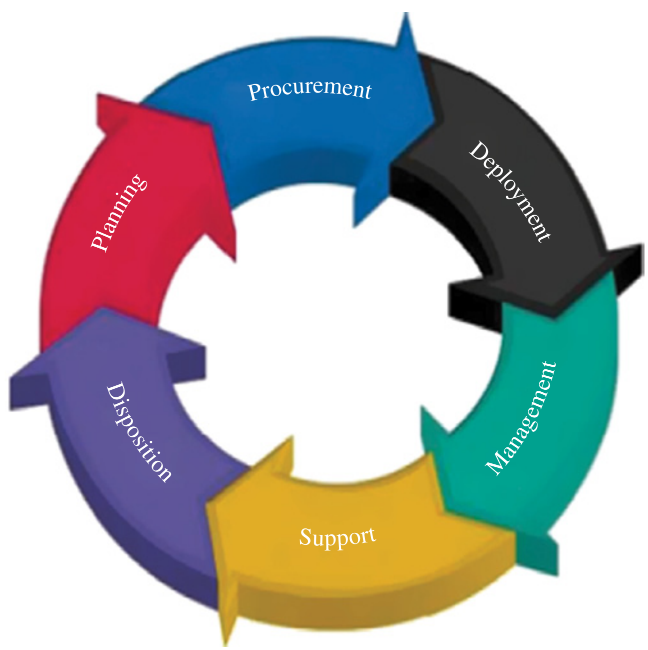


FIGURE 34.3 Asset life cycle management. Courtesy of PTS.

to support life cycle management and strategic decision making for the data center environment. Effective life cycle management is used to make decisions about repair versus replace purchases and redistribution. Life cycle management ensures organizations manage their systems more effectively and save time and money by eliminating unplanned maintenance, early replacement purchases, and, most importantly, forecast when future replacement purchases are prudent.

Two important elements in determining when an asset should be replaced are as follows: (1) what is the predictability that a component failure will occur within a certain time period, and (2) on average, how much time will it take to repair the asset should a failure be realized?

MTBF The Mean Time between Failure (MTBF) is the predicted elapsed time between inherent failures of a system during operation. MTBF is typically calculated as the average time between failures of a system as measured in hours. The MTBF is usually part of a model that assumes the failed system is immediately repaired MTTR; see below, as a part of a renewal process.

For a repairable asset, failures are considered to be those out of design conditions, which place the system out of service and into a state for repair. Failures that occur that can be left or maintained in an unrepaired condition and do not cause downtime are usually not considered failures. In addition, units that are taken down for routine scheduled maintenance are also not considered within the definition of failure.

As such, a CMMS tool that aids in identifying a typical MTBF for a certain class of an asset, tracks component and systems MTBF, and/or provides the probability of failures of the component and related systems is desirable.

MTTR The Mean Time to Repair (MTTR) is a basic measure of the amount of time required to repair a failed component or device. It is typically expressed in hours for the complete repair. It normally does not include lead time for parts.

MTTR can be a part of a maintenance contract. As such, an asset with an MTTR of 8 h is generally more valuable than one for 24 h given their respective MTBFs are the same.

However, in the context of a maintenance contract, it would be important to distinguish whether MTTR is meant to be a measure of the mean time between the point at which the failure is first discovered and the point at which the equipment returns to operation (usually termed "mean time to recovery") or only a measure of the elapsed time between the point where repairs actually begin and the point at which the equipment returns to operation (usually termed "MTTR"). As an example, an asset with a service contract SLA guaranteeing a mean time to "repair" of 8 h but with additional part lead times, administrative delays, and technician transportation delays adding up to a mean of 24 h is not any more attractive than another asset with a service contract SLA guaranteeing a mean time to "recovery" of 24 h.

As such, the ability to track these attributes with the CMMS application is very useful as a performance metric.

34.3.2 Service Management/Service Routines

Service management is most effective when utilizing predefined service routines that detail the scope-of-service tasks thus allowing end users to create service requests via work orders as well as track and update open service requests (Fig. 34.4).

Service management can include functionality for defining SLAs, which specifies the expectation between maintenance organization and the business units. Ultimately, SLAs help align service levels with business objectives. As such, it becomes important to have functionality to monitor SLAs and thus proactively monitor performance against metrics to avoid missing service-level commitments.

In addition, it is important to have functionality for escalation management to ensure the proper management of resources to achieve service levels.

Finally, it is beneficial to organize service routines into service catalogs to improve organizational communication.

34.3.2.1 Scope-of-Work Management A Scope-of-Work (SOW) Management tool inside a CMMS application usually consists of a service routine work engine that allows for the construction of a step-by-step task definition of the work to be performed. It is helpful if there is ample space to describe the work as well as separate areas for safety and precaution designations. In addition, parts usage assignments on a per-task basis are also helpful.

The service routine should include a detailed scheduling section that allows the definition of the frequency of the group of tasks.

Generally speaking, service routines must be created for every type of service for every asset. Some CMMS packages come with a standard library of general and specific service routines and SOW.

Templates by Asset Type Service routines are usually developed for individual assets but can also be created by asset type. For instance, a template version of a service routine might be created for a three-phase UPS system that can then be copied and customized for specific use for a particular model, configurations, and specific OEM.

OEM-Specific Scope of Work Additionally, some OEMs provide standard SOW for their particular equipment including:

- Start-up procedures
- Commissioning procedures

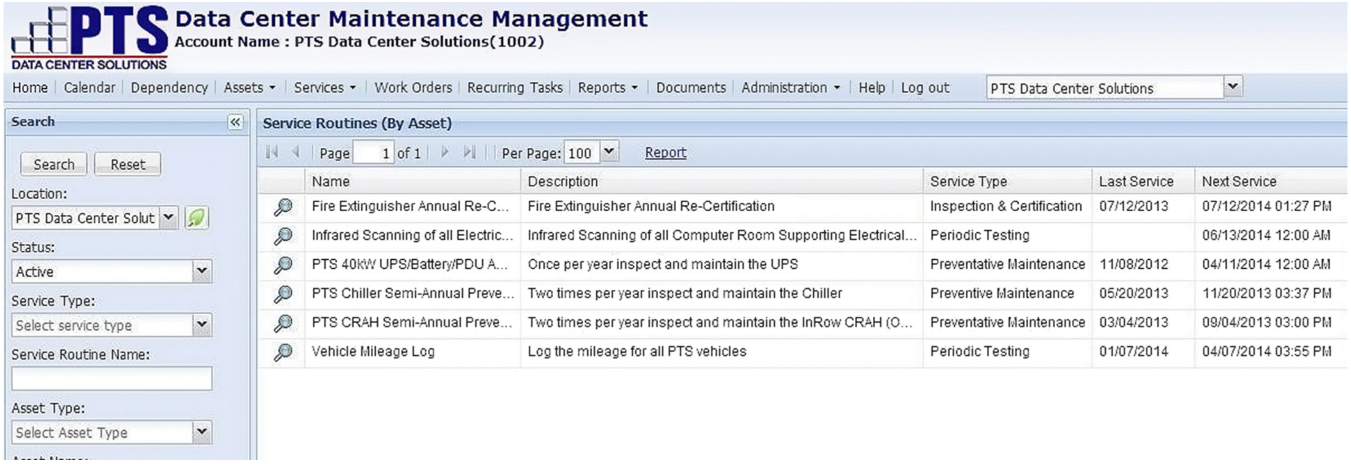


FIGURE 34.4 Service routine details. Courtesy of PTS.

- Various levels of preventive MPs
- Standard operating procedures (SOPs)
- Emergency operating procedures (EOPs)

Time-Based Scheduling Typically, service routines are created for a specific definitive frequency. For instance, a general visual inspection of a preventive MP might be created and utilized on a weekly basis as opposed to a more comprehensive service that includes changing filters, which is done on a less regular basis.

34.3.2.2 Planned Service Planned service is service whereby it has been previously anticipated and scheduled. As such, it is done with regularity. Therefore, its impact to its operational state and to that of all its related systems is well understood. In general, planned service is much more favorable than its unplanned cousin.

Planned services can be created not just for specific assets but also non-asset-specific locations.

Asset-Based Services Asset-based services are conducted on specific equipment, at a specific location. An asset can typically be considered as any equipment that has a serial number.

Location-Based Services Often is the case that maintenance services must be performed on non-asset-specific items and/or areas. In particular, this is the case for locations. For instance, cleaning a computer room is a standard maintenance activity. However, the room itself is not a serialized asset. As such, a CMMS program that allows for the creation of service routines for non-asset-specific locations is highly useful. In the case of location-based service routines, a SOW can be defined for a particular space or area. Some examples of location-based maintenance services are as follows:

- Computer room cooling
- Lighting maintenance
- Roof maintenance
- Parking lot maintenance

Preventive Maintenance Preventive maintenance allows schedules to be put into place with a goal of reducing unplanned downtime and reactive maintenance.

Effective preventive maintenance is conducted by qualified and trained personnel with the purpose of maintaining equipment and facilities, already in satisfactory operating condition, by providing for predefined inspection, detection, and corrective action of failures before they occur or before they develop into a major problem.

The actual maintenance can include cleaning, functional tests, measurements, calibrations, adjustments, and parts replacement, all performed specifically to prevent faults from occurring.

Predictive Maintenance Predictive maintenance helps determine the condition of in-service equipment in an effort to predict when maintenance should be performed. This approach offers cost savings over routine or time-based preventive maintenance, because tasks are performed only when warranted.

The goal of predictive maintenance is to allow convenient scheduling of corrective maintenance and to prevent unexpected equipment failures. The key to success is to have the right information at the right time. By understanding which equipment needs maintenance, service work, resources, and spare parts, predictive maintenance can be better planned. As such, what would have been unplanned maintenance interruptions are transformed to shorter and fewer planned service events.

34.3.2.3 ***Unplanned Service*** Generally speaking, unplanned service is not preferred. In the ideal case, the only services that are desired are those that are planned and whose impact to the operation is benign and predictable. However, such is not always the case. Inevitably, equipment fails regardless of how well preventive services are performed. Therefore, it is important to have the information well organized on how to execute unplanned and/or emergency services.

Break–Fix The most common unplanned service is that of the typical repair service. In general, it is not practical to write service routines and/or scopes of service for every potential failure. As such, what is important is to understand how to respond to a failure of an asset. In addition, it is equally important to have information pertinent to the asset readily available such as:

- Service manuals
- Service history
- Warranty contracts
- Service contracts
- SLAs
- Service technician contacts

Emergency Service Operating Procedures As a response to unplanned service, it is prudent to have established operating procedures on how to respond to an individual component and/or system failure. This can include:

- Contact procedures
- Escalation procedures
- Alternate operations procedures
- Bypass operations procedures

34.3.2.4 ***Recurring Tasks*** Recurring tasks are any tasks that do not quite fit into a regular service category. They can be events that are not typically tied to any particular asset and/or location, but are required, regularly scheduled events. Some common recurring tasks might be:

- Password change reminders
- Personnel reviews and/or training events
- Management meetings

34.3.2.5 ***Conditional Service*** Conditional service is service performed in reaction to a real-time monitoring event. It too allows for proactive maintenance and decreased unplanned downtime. As previously discussed, asset-based services are typically scheduled on a time-based schedule. However, it is often additionally preferable to prompt for service based on real-time events. For instance, a generator might have multiple preventive maintenance service routines involving periodic testing, filter and fluid changes, etc. However, often is the case that a data center manager might call for service as a result of having just responded to an emergency whereby the generators were utilized for an extended period or perhaps they have run for 100 h prior to the regularly scheduled service, and therefore, an impromptu maintenance service is needed. Therefore, the ability to call for service based on, for example, the number of hours of operation for a generator plant or the number of hours of continuous use is very beneficial. As such, a CMMS application with real-time data integration and the ability to prompt for service on event-driven data is desirable.

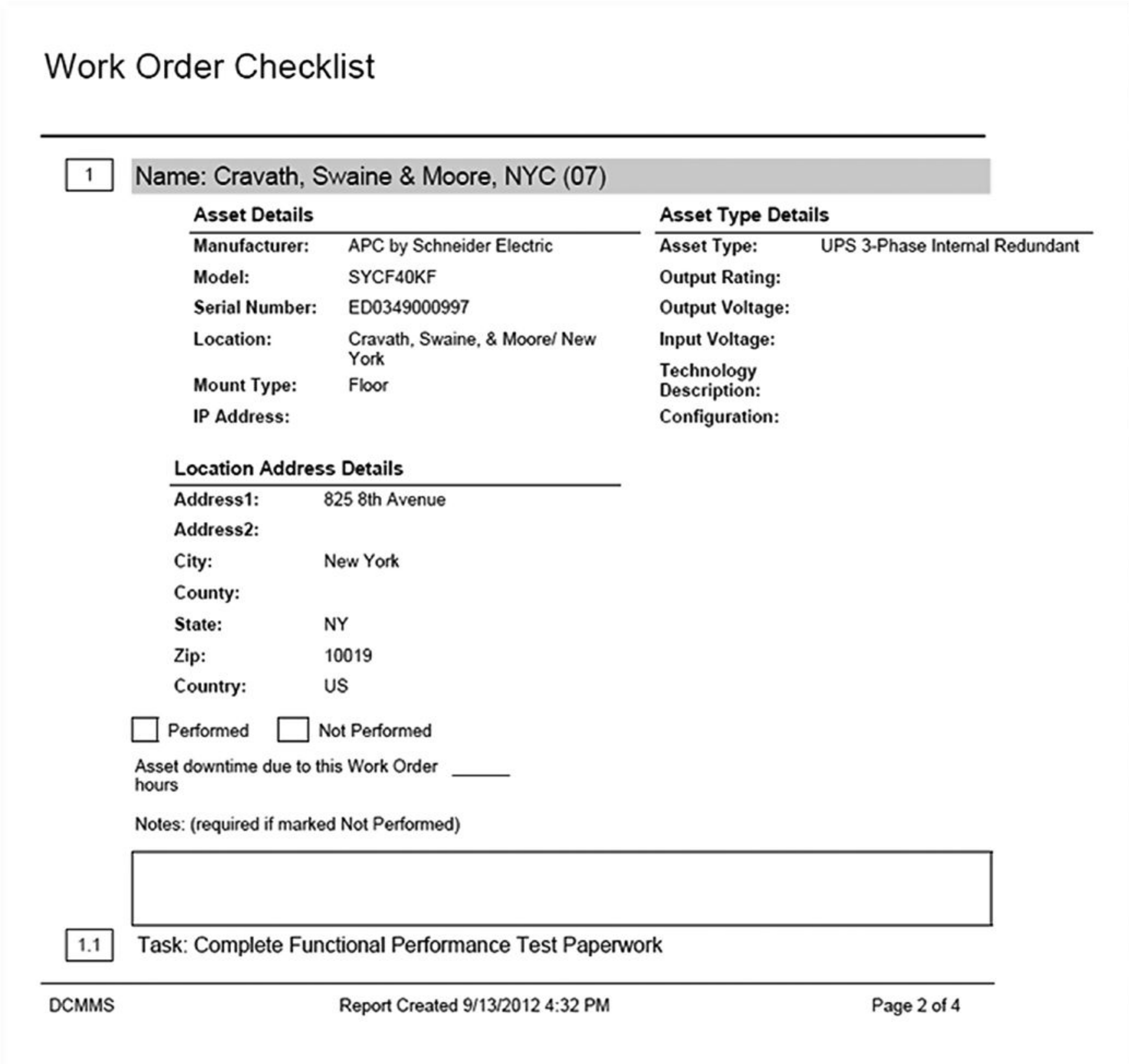
34.3.3 Work Management

Work management includes supporting both planned and unplanned maintenance activities, from the initial work request and work order generation through completion and recordings of the actual work performed. It should leverage the use of tools that enables detailed analysis of resources, materials, and equipment usage and costs—all helping to decrease labor/material costs.

34.3.3.1 ***Work Orders*** A work order is a request to complete a scope of services on a particular asset, group of assets, and/or location that is made either by the client or their agent responsible for managing the data center site (Fig. 34.5).

In most CMMS applications, the work orders are submitted via email. However, they usually can also be printed out. Work orders can include additional documentation such as instructions, site requirements, safety notices, paperwork to be completed, and so on.

34.3.3.2 ***Time and Region Management*** Graphical assignment management functionality ensures that the right person with the right skill set is assigned to the right job. For example, if a service manager knows that a SOW will be executed at a certain location on a certain day, then it makes perfect sense to conduct a search of upcoming scheduled services that are in the same general area that might be moved up to allow for efficient use of time and resources.



Work Order Checklist

1 Name: Cravath, Swaine & Moore, NYC (07)

Asset Details

Manufacturer:	APC by Schneider Electric
Model:	SYCF40KF
Serial Number:	ED0349000997
Location:	Cravath, Swaine, & Moore/ New York
Mount Type:	Floor
IP Address:	

Asset Type Details

Asset Type:	UPS 3-Phase Internal Redundant
Output Rating:	
Output Voltage:	
Input Voltage:	
Technology Description:	
Configuration:	

Location Address Details

Address1:	825 8th Avenue
Address2:	
City:	New York
County:	
State:	NY
Zip:	10019
Country:	US

☐ Performed ☐ Not Performed

Asset downtime due to this Work Order ______ hours

Notes: (required if marked Not Performed)

1.1 Task: Complete Functional Performance Test Paperwork

DCMMS Report Created 9/13/2012 4:32 PM Page 2 of 4

FIGURE 34.5 Work orders. Courtesy of PTS.

34.3.4 Calendar

This functionality within a CMMS platform allows users to conveniently view data center maintenance details in a calendar view. Optimally, the events that should be viewable within the calendar are:

- Work Order Dates
- Server Routine Schedules
- Warranty Expirations
- Service Contract Expirations
- Recurring Task Due Dates

Additionally, it is useful if the calendar is viewable in day, week, and month views. Also, it is helpful if the events can be filtered down by client and/or location.

34.3.5 Report

The report functionality is a vital part of any CMMS application. It is the primary method for viewing and analyzing data, be it lists, schedule, or performance based.

The hallmark of any report-generating tool is the ability to easily filter and tabulate data into meaningful ways in accordance with the needs of the user. As such, it is important for the reporting tool to have certain key capabilities including:

- A standard array of the most typical and useful reports
- The ability to generate and save custom or customized reports
- Extensive search and filter capabilities
- The ability to print and/or export in any number of formats

Typically, reports are presented as two types, management and operational. Operational reports are usually detail oriented and show the latest up-to-date records. Operational reports are used by stakeholders for short-term tactical decision making. Management reports look at summary data over a longer time horizon and are used for strategic decision making. Each type is discussed in greater detail in the following text.

As mentioned, operational reports are detailed reports of the latest, most up-to-date data. They list the details of the current state and all the short-term events. They are utilized to make day-to-day decisions across a broad array of maintenance management activities.

34.3.5.1 Lists Simply stated, lists are tabulations of the most basic data needed to conduct activities day to day. They include:

- Customers
- Contractors
- Assets
- Spare Parts
- Resources
- Service Routines
- Contracts

34.3.5.2 Completed Activities Completed activities include the activities and/or the services provided over a defined period of time. Typically, stakeholders require an understanding of completed events over a 30-day period. However, this can be any time period that best accommodates the organization.

Work Orders Recently completed work orders define what services were completed, by whom, at a particular location, and for a particular asset or series of assets. In addition, it is ideal to capture the amount of time that was taken to complete the work (and thus at what labor cost), how much downtime was realized (if any), and what spare parts were utilized.

Recurring Tasks Likewise, completed recurring tasks relay information about non-asset- and/or non-location-specific events that may be tracked on a regular basis.

Warranties Knowledge of the currently renewed or new warranty coverage for assets is important for maintaining appropriate asset replacement and repair cost control.

Service Contracts Similarly, knowledge of currently renewed or new service contract coverage for assets is important for reigning in otherwise costly time-and-material repair costs.

34.3.5.3 Scheduled Activities An understanding of upcoming activities is equally as important to best align and plan resources in coordination with them. Once again, each organization typically establishes a time period that best works for its own resource planning. For most, it is usually between 30 and 90 days.

Work Orders Upcoming work orders define what services will be completed, by whom, at a particular location, and for a particular asset or series of assets. In addition, they predict the amount of time that will take to complete the work (and thus at what labor cost), how much downtime may be realized (if any), and what spare parts will be needed. This information is then used against the actual data for use in process improvement.

Recurring Tasks Likewise, upcoming recurring tasks relay information about non-asset- and/or non-location-specific event that are being tracked on a regular basis.

Warranties Knowledge of the upcoming renewals of warranty coverage for assets is important for maintaining appropriate asset replacement and repair cost control.

Service Contracts Last, knowledge of the upcoming renewals for service contract coverage for assets is important for reigning in otherwise costly time-and-material repair costs.

34.3.5.4 Management Reports As mentioned, management reports look at summary data over a longer time horizon. They are utilized to make strategic decisions to improve the operational and/or maintenance processes. At the heart of management's long-term planning is a deep understanding of three factors: the operational performance of the data center, the financial impact of operations, and the effectiveness of resource utilization.

Performance Data center operational performance is a direct result of the effectiveness of the maintenance provided to its supporting infrastructure. The more effectively maintenance is performed, the greater the availability of the data center. The measure in which "effectiveness" is measured is the time the data center was fully available for operations versus the amount of time it was not—usually expressed in terms of hours.

UPTIME Data center uptime is the amount of hours the data center was fully operational. This means the data center's critical load operated without incident and/or disruption despite what events may have been affecting the data center itself.

In addition, it's also useful to track the uptime of individual assets even though an event that might cause the disruption of continuous operation of this asset may not cause a disruption of the system's availability as a result of redundancy.

DOWNTIME The antithesis of uptime is downtime. In this instance, data center downtime is a measure of the amount of hours the data center did not operate fully. The probable cause of the downtime can often be directly traced to the failure of an individual asset or a complete system, despite redundancies.

Once again, it is useful to track individual asset downtime even though it might not have caused a failure of the overall data center availability.

FAILURE IMPACT ANALYSIS The most important information to be gleaned from any data center downtime is its cause. As such, failures should be fully analyzed for their root cause. Furthermore, an analysis of the root causes can lead to improvements in operational and maintenance practices and processes.

Financial The second metric used by management to gauge the effectiveness of data center maintenance is its financial factors. In particular, it is vital to track the labor cost of all maintenance activities, the cost of all spare parts, and the loss of revenue as a result of failures, which may or may not be as a direct result of failures in effective maintenance.

LABOR There are a number of labor costs that should be considered throughout maintenance executions. For instance, the labor cost of all direct maintenance management personnel including executives, managers, technicians, and maintenance staff should be tracked.

CONTRACTS In addition to labor is the cost for all maintenance contracts regardless of whether warranty and/or on-site service-based needs are accounted for.

SPARE PARTS Another financial impact is the cost for all spare parts regardless of whether they are in service or in inventory waiting to be used.

In addition, it is also useful to track the turnover of all spare parts inventory to ensure excessive capital is not being wasted on unused inventory.

LOST REVENUE DUE TO FAILURE Lost revenue as a result of failures that cause the interruption of data center service is often the most detrimental. Although often difficult to quantify, doing so provides an excellent cost basis to justify maintenance expenditures.

DEPRECIATION The final, and often most overlooked, financial impact is that of depreciation. Depreciation is the decrease in value of assets (fair value depreciation) and the allocation of the cost of assets to periods in which the assets are used (depreciation with the matching principle). More simply stated, it is the cost to replace a particular asset over a certain period of time reflective of the assets effective useable lifespan:

$$\text{Annual depreciation expense} = \frac{\text{cost of fixed asset} - \text{residual value}}{\text{useful life of asset years}}$$

Resource Utilization The final metric used by management to gauge the effectiveness of data center maintenance is tracking how effectively it is utilizing its resources. In particular, it is vital to track the labor cost of all maintenance activities, the cost of all spare parts, and the loss of revenue as a result of failures, which may or may not be as a direct result of failures in effective maintenance.

34.3.6 Document Management

Documents can be associated with an asset, asset type, location, manufacturer, or model. Documents associated with a specific asset will only appear for the asset. Documents associated with an asset type will appear as a related document for all assets that are of that asset type. Documents associated with a location will appear as a related document for all assets that are in that location. Documents associated with a manufacturer will appear as a related document for all assets matching that manufacturer. Documents associated with a model will appear as a related document for all assets matching that model.

34.3.6.1 Document Types The types of documents that need to be tracked throughout maintenance management are nearly endless. However, a CMMS tool should include any number of standard types including:

- Entitlements
- Purchase Orders
- Quotes
- Safety Documentation
- SOW Documents
- Completed Service Paperwork
- Warranty Documents

- Service Contracts
- Products Technical Specifications
- Facility Drawings

34.3.6.2 Document Associations In addition to tracking document types, it is equally important to track document associations. The most common associations for documents are:

- Locations
- Asset Types
- Assets
- Manufacturers
- Models

34.3.7 Administrative Functions

As is the case with any software tool, there must be a management and/or administrative set of tools available. Administrative tools are typically broken into the following areas:

- Tools to globally change and/or replace data in bulk
- Tools to administer system functions such as users, roles, passwords, email, etc.
- Tools to manage pick list information within the program

34.3.7.1 User Management The administrator can modify user details such as username, display name, password, and role. The administrator can also add, edit, or delete a user.

34.3.7.2 Role-Based Management The administrator can control a user's access rights by assigning none, read only, or full access to any number of program areas.

34.3.7.3 Audit Logs Audit logs allow a user to see the actions performed within CMMS by whom they were performed, the date, and a description of the action. The user can also search by date, by whom the action was performed, and/or the program segment affected.

34.3.7.4 System Settings System settings can include any number of things from default values for fields to email settings, to calendar standard settings, and more. In the case of PTS's DCMMS, there are seven system values that can be set:

- Dashboard Recurring Tasks—any recurring tasks due within the specified number of days will be displayed on the dashboard.
- Dashboard Service Contracts—any service contracts that will expire within the specified number of days will be displayed on the dashboard.
- Dashboard Service Routines (by location)—any service routines (by location) that are due within the specified number of days will be displayed on the dashboard.
- Dashboard Service Routines—any service routines due within the specified number of days will be displayed on the dashboard.
- Dashboard Warranty Contracts—any warranties that will expire within the specified number of days will be displayed on the dashboard.
- Dashboard Work Order—any work orders due within the specified number of days will be displayed on the dashboard.
- System Message—a message that will be displayed on the login screen.

34.3.7.5 Pick List Management Most CMMS programs offer an array of predefined data throughout the program that can be tailored and/or customized to suit the particular case. These pick lists improve data integrity by making sure classifications for records are the same throughout the program.

34.3.8 User Portal

Finally, one of the most functional parts of any CMMS tool is the user portal. Here, critical data is presented as tables, lists, charts, graphs, or whatever format best relays the actions required.

34.3.8.1 Coming Due/Past Due Windows Coming due/past due windows allow a user to see recurring tasks, service routines by asset and location, and work orders that need to be performed. Also, it allows a user to see service contracts and warranties that are about to expire. Last, it allows a user to see what parts need to be ordered.

34.3.8.2 Weather Some CMMS applications provide access to external websites and/or RSS feeds. One particularly useful tool is access to a weather map or service. Since foul weather often precedes the use of supporting infrastructure, maintenance on this equipment is often a standard practice following the event. Additionally, most thunderstorms are short in duration and come into and go out of an area usually within 30 min. As such, certain precautionary maintenance activities can be enacted if a storm is being tracked.

34.3.8.3 Failure Probability and Statistical Analysis Another great performance indicator that is suitable for any portal probability and statistical analysis predictions and/or trends. As discussed earlier, MTBF and MTTR are predictors to overall system availability. Likewise, preventive maintenance activities, or the lack thereof, also have an impact on failure/downtime probability. As such, some programs may provide numerical guidance as to the statistical probability of failure at the component, at the system, and/or even at the site level. This is often contrasted against the actual performance data.

34.3.8.4 Actual Performance Data The actual performance data is typically denoted in two important ways. The first is the amount of continuous hours of uptime without a disruption and/or failure. The second is the inverse of this, the amount of downtime experienced over a certain period of time. Once again, it is important to understand this data in comparison over what time period it was analyzed (e.g., day, week, month, year, all time) as well as the actual data at the component, system, and/or site level.

34.4 CONSIDERATIONS IN SELECTING CMMS

Selecting an appropriate CMMS for an organization will allow its facility management staff to monitor all enterprise assets, their conditions, and their work processes. It can result in lower downtime, lower cost of operations, and better planning and control. As such, the factors for evaluating and selecting the most appropriate platform can be broken down into two categorizes: (1) Product Feature Sets and (2) Enterprise Integration Factors.

Ultimately, the organization has to choose a tool that includes the feature set best suited to the needs of the organization. As discussed in Chapter 2, the features fall into the following categories:

- Asset Management
- Maintenance Management
- Life cycle Management
- Inventory Management
- Work Order Management
- Calendar/Work Scheduling
- Reports and Portal

In addition to software features, there are a number of other factors that will impact the organization that must be considered before making a final decision. The remainder of this chapter discusses factors in detail.

34.4.1 Implementation Process

Just like effective maintenance performance, effective CMMS implementation starts with a sound process. And, like every process, it has a beginning, a middle, and an end.

34.4.1.1 Preimplementation Assessment The first step in the process is to determine how information will be gathered, by whom it will be gathered, and how long and how much it will cost. Many CMMS software vendors provide on-site and remote assessments to aid organizations in this process. In addition, even if in-house staff is up to the task, a CMMS supplier may serve as an invaluable consultant during the process.

34.4.1.2 Implementation The actual implementation and use of the CMMS tool is another area where the organization needs to make a decision as to whether it will be more cost-effective to use in-house or outsourced resources. Once again, most CMMS vendors offer an array of services to aid an organization in implementing its platform. The actual implementation will take the most time of all the steps. As such, a timeline should be created to define all the steps that must take place including:

- Information Gathering
- Data Entry
- Maintenance Process mapping
- Document Scanning
- Training Regimen Development

34.4.1.3 Training Learning a new software can be conducted in a number of ways including self-paced by reading manuals, via off-site instructor-led classes, via on-premise instructor-led classes, and via self-paced and/or structured online trainings and/or tutorials. Typically, CMMS vendors incorporate one or more of these methods into their offerings.

34.4.1.4 Postimplementation Feedback Once a CMMS tool is finally implemented and being used, it is important to analyze its usefulness. Here again, CMMS vendors usually offer a host of services to evaluate among other things:

- Staff adoption rates
- Cost savings analysis
- Process improvement
- Resource improvement

34.4.2 Staff Considerations

In the end, it's people that are using these tools to improve how they perform their jobs in keeping the data center facility operating at its peak performance. As such, when considering CMMS, an organization needs to consider how the tool will be integrated by people into the normal process of maintaining the data center.

34.4.2.1 Information Gathering Any tool is useless if it doesn't have an accurate and complete data set to work with. Therefore, consideration must be given as to how information will be gathered from the facility and entered into the software.

Generally, there are two ways of gathering data—manually and automatically.

Manual Information Gathering The biggest problem with manual information gathering is accuracy. People are fallible. As such, they make mistakes. Studies have shown that accuracy failure rates for rudimentary manual asset inventory gathering can be as high as 10%[1] (Watson and Fulton, 2009).

Accuracy of manual information gathering can be improved by utilizing standard scripts as well as by utilizing tools such as bar coding. As such, any CMMS tool that integrates such features and/or allows information gathering to be conducted in a mobile manner should be considered favorably over those that do not.

Automatic Information Gathering The payback for automatic information gathering is attractive and obvious. However, it is also difficult and expensive to realize. The reason is that automatic data collection relies on instrumentation and data management. Typically, devices can be connected to the corporate network and polled for their characteristics. For instance, a server can report its current state as well as its logical address on the network using various services and/or protocols such as SNMP, IPMI, WMA, and more. This is also true for supporting infrastructure such as UPS, CRAC, and more.

34.4.2.2 Instrumentation Properly instrumenting a data center has benefits beyond those of improving maintenance management data gathering efficiency. Real-time device management can enable more efficient and cost-effective operation. However, outfitting data center support infrastructure with metering, monitors, and management platforms is a project unto itself—and an expensive one at that. Doing so at the very least provides the base data from which actionable plans can be based including maintenance management, performance management, and capacity planning.

34.4.2.3 In-House Skill Sets The skills and capabilities of the organization's in-house facility and management staff must be considered in selecting the most appropriate CMMS tool. As mentioned earlier, ultimately, effective maintenance is performed by people. As such, an organization should assess its in-house capabilities to ensure it meets the minimum integration requirements of whatever tool it's considering.

At a minimum, the in-house staff should be well versed in the use of software-based tools. In addition, they must be able to grasp the concept of how information flows as well as process management.

[1]Computer Associates technology brief Striving to Achieve 100% Data Accuracy: The Challenge for Next Generation Asset Management.

34.4.2.4 Staff Time Consideration In addition to staff skills, the time availability of the staff to complete the project must be considered. Presumably, everyone is already busy doing the normal course their jobs demand. Therefore, the additional time it will take to perform the integration efforts including information gathering, software training, and process building is vitally important.

34.4.2.5 Training Effective training for the use of any CMMS tool is critically important. However, the really purposeful training goes beyond the act of learning how to just use the software. Often, training must encompass what it means to effectively maintain the data center in order to best utilize the tool set of the software. For instance, if the staff is not accustomed to using formal work orders to dispatch and instruct service technicians, then that must be included in the training regimen. Likewise, if the organization does not have written processes and procedures for standard operations, emergency operations, and maintenance, then not only does the staff have to be educated in their use, they may also have to establish them in the first place.

CMMS integration failure for even the best software tools can be directly traced to a lack of adoption by the staff for which it was intended. Often is the case that staffers lose faith in a new CMMS tool as a direct result of the inability of an organization to implement an effective training program. Studies have shown that learning new software is best accomplished via immersion in using the tool over time. As such, the time element must be considered. To that end, the CMMS suppliers often offer a number of training techniques including:

- Offering user manuals
- Online demo sites and/or tutorials
- Off-site and/or on-site instructor-led classes

34.4.2.6 Technical Support Like any software package, CMMS tools often have annual support contracts for ongoing maintenance and support. As such, the support program of the CMMS supplier should be fully vetted and at the very least include the following:

- Software upgrades and/or patches
- 7 × 24 × 365 user technical support
- 7 × 24 × 365 administrator technical support
- Online available manuals and documentation
- An online and user-accessible knowledge database

In addition, it is often desirable that the CMMS vendor has at its disposal an entire array of optional services including:

- Program customization
- Preimplementation consulting

- Implementation and/or implementation consulting
- Postinstallation consulting
- Maintenance process and procedure development consulting
- SOP development consulting

In addition, some CMMS vendors even offer complete outsourcing of facility maintenance management using its own software tool as a basis of its service.

34.4.3 IT Requirements

Depending on the method of how the CMMS is deployed, the implementation may carry with it a number of IT requirements. CMMS providers of client-/server-based applications will typically make available a minimum and recommended IT requirements.

34.4.3.1 Platform Most CMMS suppliers offer their application in one of two formats: (1) a web-based, software-as-a-service (SaaS) offering and (2) a traditional client-/server-based application. However, some vendors offer both types, which is best for an organization that has a number of dependencies including:

- Availability of adequate, self-directed compute, storage, software, and support resources
- Availability of sufficient IT support staff
- Concerns over data security and/or regulatory compliance issues

Client/Server Based The traditional client-/server-based application is normally provided by disc or download. It requires a server running a particular operating system and/or having certain services enabled. In addition, the application usually requires some capacity for data storage as well as a number of software services required for operations. The organization is typically responsible for supplying its own hardware/software on which the CMMS application will reside, including often the database front end needed by the application. This means the client will be responsible for operating and maintaining an MS SQL, Oracle, or MySQL environment as required by the application.

The actual installation of the software can usually be done by the organization itself assuming it has sufficient resources to do so or can be purchased as an optional service from the software vendor.

Software as a Service In more recent years, the SaaS model has become more prevalent among CMMS providers. In this approach, the SaaS provider typically operates a data center in which the application runs. Access is granted to customers via a web-based, secure interface. In some cases, the CMMS provider may offer stand-alone versions of the hosted application, which are unique to the subscriber and certain not to be accessible by other clients of the provider. In other cases, the CMMS provider may offer a common platform from which each new client is given secure access to only “their” stuff.

The benefit of the SaaS approach is that the user does not have to worry about the IT platform other than its own ability to provide its users web access. In addition, the client does not have to worry about application upgrades and/or patches. Conversely, the SaaS client has to concern themselves with data security, user Internet bandwidth capacity, and the provider’s financial viability to stay in business.

34.4.4 Usability/Ease of Use

Another important factor in comparing different CMMS platforms is their overall usability, or stated another way, how easy they are to use. Obviously, a software tool’s usability can be very subjective. However, in general, look for a program that organizes maintenance management in a manner similar to the organization’s approach.

34.4.4.1 Simplicity The best software tools are the ones that are so intuitive that they are almost usable without training or referring to a manual as a reference. In general, an organization’s current facility staff has a sense about how they expect to be able to do within the software. As such, an array of potential users should be involved in the decision-making process.

34.4.4.2 Mobility One of the more common operational features desired in CMMS tools are those of portability and mobile use. CMMS tools that are mobility enabled mean they have adapted their software for use on the usual array of tablet and/or smart phone platforms.

Mobile users typically will have access to all of the information they would normally have in any desktop environment. Similarly, they also can have all the functionality of one too, including:

- Adding, deleting, or editing asset attributes
- Accessing and executing Work Orders and related documents
- Searching service histories
- Viewing and/or utilizing schedules
- Accessing Service Routine SOW and related documents

It should be noted that CMMS applications commonly have views that contain vast amounts of data about an asset. Since mobile platforms are essentially scaled-back versions of desktop processing environments and they usually have smaller screens, users should fully test CMMS functionality on the particular mobile platform they intend to use to make sure that usability and functionality have not been compromised to attain the mobile functionality.

Bar Coding/Reading Some mobile CMMS platforms leverage the use of bar coding technologies. This feature often enables unique bar codes to be generated from the CMMS tool itself and affixed to individual assets. Likewise, they have mobile hardware with bar code readers that can be used to immediately identify previously tagged equipment and instantly access information about it.

Asset Location Management Another more recently available functionality is asset location tracking. This is most often provided via the CMMS supplier partnership and/or integration with an existing radio-frequency identification (RFID) system provider. With this technology, unique ID tags are deployed on individual assets and can be read/detected via radio-frequency scanners. In some CMMS platforms, this tag can be used for both RFID and bar code scanning as previously described. Since these are typically different suppliers offering these platforms, potential users should fully test the integration to make sure it meets their requirements.

34.4.5 Cost

The costs of implementing a data center CMMS solution fall into two overall categories: (1) the cost associated with the software solution itself and (2) the cost of the services associated with planning, implementing, and maintaining the solution.

The first cost driver is the decision on which software strategy fits best—purchase a standard CMMS package, customize a standard CMMS package to tailor it to the organization's particular needs, or architect and build one from the ground up.

The second cost driver is to determine the level of in-house services that can be leveraged from implementation and more importantly what services will have to be procured from outside vendors.

34.4.5.1 Strategy As with any software project, the result of "build" versus "buy" is easily answered since designing a CMMS from the ground up results in a perfectly tailored solution that will meet the needs of the organization. However, getting there is a lengthy and costly process. In buying a standard CMMS package, the costs and speed of acquiring are greatly compressed but at the cost of an inexact fit to the needs of the organization. Conversely, customizing an existing CMMS application lies somewhere in between with respect to cost and meeting needs.

Build New Undertaking a software design project is a daunting task. It requires a considerable amount of in-house resources and time to establish the software requirements. Then, a programming team needs to be either insourced or outsourced for the code development. Once that is complete, extensive functional testing is performed before the program is released. Generally speaking, an organization can expect to spend approximately $1 million, and the entire process could take over a year.

Customization Due to the sheer expense in money and time, an organization that has to have certain customizations that are not available in existing off-the-shelf programs may be able to find one that almost meet the requirements and then contract the vendor to perform the necessary code changes. Some typical customizations can include:

- Reskinning the program to have the organization aesthetic and logo
- Integration with existing in-house applications for various functions including:
  - Financial
  - Change control
  - Configuration management databases
  - Contact database
  - Calendar and scheduling
  - Project management

The cost of these customizations varies greatly as one can imagine. However, it is not uncommon for a series of customizations to cost in the hundreds of thousands of dollars and take 6 months or more.

Prebuilt Solution Preexisting CMMS solutions are by far the most cost-effective and quickest to deploy. Generally, the existing platforms are mostly for generalized facility maintenance, but a few vendors have solutions specifically designed for specific industries including healthcare, utility service providers, and data centers.

CMMS solutions range from applications offered in a base configuration with various modules for added functionality to applications that are fully functional.

34.4.5.2 Software The cost of standard CMMS offerings varies greatly among the major suppliers. The major attributes that typically affect price are:

- The specific industry being served
- The number of assets to be managed
- The number of users that have access to the system
- The number of facilities to be managed
- The size of the facility being managed
- The term of access to the software

As previously mentioned, purchase options can include a onetime purchase or subscription-based SaaS options.

Purchase Model The typical client-/server-based application model is usually offered as a onetime fixed-fee purchase for a specific feature set.

The software cost is often tied to one or more of the variables described earlier.

Often is the case that the purchase can be scaled to the features and/or capacity needed by the organization today while maintaining the option to increase features and/or capacity on an as-needed or periodic basis.

SUPPORT CONTRACT As part of the purchase model, it is recommended that an annual support contract is purchased along with the software.

A support contract ensures access to two important services: (1) updates, patches, and upgrades to whatever modules may have been purchased and (2) technical support for both users and administrators.

The cost of the support contract varies, but an annual cost of 15–20% of the software's purchase cost is not unheard of.

Subscription Model The subscription-based purchase model is typically offered on an annual or sometimes monthly fee basis. In contrast to the purchase model, subscribers do not have to purchase a separate annual support contract in that those services are normally included as part of the periodic fee.

The subscription cost is also frequently tied to one or more of the variables described earlier, similar to the purchase model.

Especially in the subscription model is the case where access and functionality can be scaled and changed on an as-needed basis.

34.4.5.3 Services As previously described in the section "Implementation Process," implementation can be broken down into four major service categories: (1) preimplementation, (2) implementation, (3) training, and (4) postimplementation. The purpose for all of them is to provide adequate resources and experience to make the CMMS integration into the organization as smooth as possible.

In general, the overall cost of implementation and training services often equals or exceeds the cost of the software itself.

Preimplementation Assessment Service A preimplementation assessment is a consultative-type service with the goal of:

- Identifying the current state of the data center
- Identifying the location of quantification of all assets
- Identifying details on the facility itself
- Identifying and obtaining all asset and facility documentation including:
  - Drawings
  - Warranties
  - Service Contracts
  - Technical Specifications
- Identifying and documenting all known facility procedures and processes including:
  - SOPs
  - EOPs
  - MPs

The cost for preimplementation services is often billed on a time-and-material basis given the often unknown conditions. For a small data center facility, this might equate to only a single person for 1 day; but for larger facilities, this can easily become multiple people for multiple days on-site.

In addition, many CMMS service providers may also charge for the services of creating and developing missing documentation such as:

- "As-built" conditions drawings
- Layout drawings
- SOPs, EOPs, and MPs

Obviously, the costs for such services vary greatly but directly correlate with the amount of effort needed to produce the results.

Implementation Services If adequate information was gathered during the preimplementation service phase, then it doesn't need be repeated here. However, if no outside consultants were utilized in such a way, then all of that information has to be gathered for this phase.

The implementation itself involves performing the data entry of all the information gathered into the newly acquired software platform. This is either done manually or, oftentimes, the CMMS vendor has bulk scripts and database formats than can be uploaded into the software.

Once again, the cost for these implementation services is often billed on a time-and-material basis given the often unknown conditions. As for the data entry portion, sometimes, this work can be performed by the CMMS vendor off-site or remotely so as to realize a lower rate for the project.

Training Services As previously discussed, training services can be conducted in a number of ways including:

- Self-paced by reading manuals
- Off-site instructor-led classes

- On-premise instructor-led classes
- Self-paced online trainings and/or tutorials
- Pretaped instructor-led trainings and/or tutorials

Generally speaking, the most expensive training formats are instructor-led courses with those involving traveling being the highest cost. Self-paced training courses offer an excellent training value. However, the fallible nature of people often precludes this from being the most effective learning technique. As such, instructor-led courses, like school, offer the best format for true immersive learning.

As mentioned before, the cost for training services is often time and material based and commensurate with the length and depth of the training as well as the class size.

That being said, effective and complete training is an absolute necessity to realize a successful CMMS implementation. As such, at least 10% of the overall budget should be allocated to training services. In addition, training should be ongoing—not just for newer employees but offered as a constant stream of education in the proper use of the tool and combined with the more important overall goal of effective data center maintenance management.

Postimplementation Feedback Services Feedback on the effectiveness of the CMMS tool is invaluable in making substantive improvement to the overall data center facility maintenance management process.

As has been the theme throughout, the cost for postimplementation services is directly proportional to the time-based effort required to research and analyze the information being sought. To that end, such efforts are usually billed on a time and material basis; but if a SOW is well established, it may be possible to procure such services on a fixed-fee basis.

34.5 CONCLUSION

Success will not happen overnight. The CMMS must be implemented correctly and used by a well-trained staff.

By implementing a CMMS solution, an organization can keep track of virtually everything that happens in the data center facility. Long gone are the days of paper records of service reports.

Management will be able to determine how much time is spent on certain tasks, see what maintenance activities are scheduled and completed, and know what materials have been purchased. In addition, they can manage technicians' time more effectively. Also, it may be possible to create a work order over the web when maintenance is needed so that response to these orders can be done more efficiently.

Of course, all of this is only possible if the right CMMS solution is purchased and it is fully implemented and integrated into the organization and the organization maintains an effective, immersive, and ongoing training regimen.

Finally, one of the most important things to remember when purchasing a CMMS solution is to include a representation from all the departments that will be impacted by the implementation of the solution and include them in the decision-making process. The groups that should be represented and the roles they perform are as follows.

34.5.1 Facility Management Staff

Facility management staff bears the overall responsibility for data center facility maintenance operations.

34.5.2 Maintenance Technicians Staff

Maintenance technicians are the field technical experts that perform rudimentary, thorough, complex maintenance tasks.

34.5.3 Facility Engineering Staff

Facility engineering staff members are experts in understanding complex support infrastructure systems and their maintenance requirements.

34.5.4 Executive Management Staff

The executive management personnel bear the responsibility for the overall strategic vision of data center operations and ownership.

34.5.5 Data Center Operations Staff

Data center operations staffers are responsible for $7 \times 24 \times 365$ data center availability on a day-to-day basis.

34.5.6 IT Management Staff

IT management personnel are responsible for managing how IT can serve as an enabler in imbedding a CMMS tool into the organization.

34.5.7 Financial Staff

Financial staffers take financial ownership of the project.

34.5.8 Human Resource Staff

Human resource staffers are responsible for continuing education and training on an ongoing basis.

34.6 TRENDS

Some of future trends of CMMS in data centers include the following.

34.6.1 Growth

CMMS is growing as fundamental software in the data center. Companies don't just want it, they need it.

34.6.2 Lowering Costs

Lowering costs is always a prevalent concern for any company, even more so with all of today's economic downturns. A CMMS can be cost-effective for a company by creating cost control for maintenance costs.

34.6.3 Conditional Maintenance

The ability to effectively and immediately respond to real-time events and emergencies is a desirable feature within the CMMS application. As previously discussed, it allows for proactive maintenance and decreased unplanned downtime, which creates reliability.

34.6.4 Interoperability

Interoperability is the seamless integration of multiple related or unrelated business applications.

FURTHER READING

Capehart BL. *Handbook of Web Based Energy Information and Control Systems*. Lilburn: The Fairmont Press; 2011.

IBM Corporation. *IBM Maximo Technology for Business and IT Agility*. IBM Corporation; 2010.

Predictive Maintenance Program. Available at http://engineer.jpl.nasa.gov/practices/ops13.pdf. Accessed on May 23, 2014.

Thomas S. 2011. CMMS Best Practices Study. Available at www.reliabilityweb.com. Accessed on May 23, 2014.

PART V

DISASTER RECOVERY AND BUSINESS CONTINUITY

35

DATA CENTER DISASTER RECOVERY AND HIGH AVAILABILITY

CHRIS GABRIEL
Logicalis Group, London, UK

35.1 INTRODUCTION

"Anything that can go wrong will go wrong." This is the profound and universal truth, set out very succinctly in a series of laws known variously around the world as Reilly's and Strauss' and most famously in many Anglo-Saxon countries Murphy's law.

This law states that anything of a mechanical nature will fail no matter how well designed, well constructed, and well used it is. And, most importantly, in Murphy's law, it will go wrong at exactly the wrong time, always.

Murphy's law was coined by Edward Murphy, an engineer working at Edwards Air Force base in the United States in the late 1940s on various projects including g-forces and their impact on the human body during massive deceleration events. His son later stated that his father's approximate phrase was something like, "If there's more than one way to do a job, and one of those ways will result in disaster, then somebody will do it that way."

When the *anything* in question is the physical repository for an organization's business applications and business data, the data center, this law poses some significant and troubling questions to those that have designed and built the physical structure, those that designed and built the enterprise IT architectures that reside within it, and those that operate both of these business critical assets on a day-to-day basis. Those that face the most troubling question are those that rely on the data center to support their core organizational activities—the business.

In today's IT-centric business world, it could be said that while the accountants still have hold of the organizational purse strings, the data center has hold of the organizational heart strings. If the heart stops beating, the pulse of the organization starts to slow rapidly; and for some organizations, that could mean business processes or the business itself stopping terminally.

This chapter aims to give a sense of the two key design, planning, and process approaches to maintaining the required level of service from the data center and the enterprise architectures residing within it: Disaster Recovery (DR) and High Availability (HA).

Before we continue, we should define what we mean by the terms DR and HA and also what outcomes we wish to achieve by investing in either of these two strategies. We turn, as we will on many occasions without apology to Wikipedia for our definition source:

> **Disaster recovery** (DR) is the process, policies, and procedures that are related to preparing for recovery or continuation of technology infrastructure, which are vital to an organization after a natural or human-induced disaster [1].

> **High availability** is a system design approach and associated service implementation that ensures a prearranged level of operational performance will be met during a contractual measurement period [2].

DR is the discipline of planning for something to go wrong, for downtime of varying length to happen, and a process to initiate recovery in an orderly and timely fashion. DR assumes loss of something; those losses could be of the physical data center, loss of connectivity to that data center, loss of the physical infrastructure, and most importantly a loss of data between two points (the point of disaster to the point of recovery).

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

DR planning is having a spare wheel in the boot of your motor car in case of a puncture. There is acceptable and recoverable downtime, after an acceptable loss of service. A motor car owner should service their car to manufacturer's guidance and can decide to extend their DR approach by acquiring additional roadside breakdown cover, but these only reduce the risk of failure; they do not mitigate it, and breakdown cover is assuming failure will happen, and this investment is made to reduce the time of recovery when things go wrong.

HA is proactively designing systems to mitigate a system failure, to minimize downtime to a negligible amount, and to ensure and assume no data loss. HA planning is having totally redundant flight control systems on a commercial airliner. Uptime is the only thing that counts. And while aircraft systems fail all of the time, many times while the plane is in the air, very few planes crash.

Aircraft are designed to cope with single- or even multiple-system failures. Pilots are trained to respond with predefined checklists, designed to standardize recovery from failed systems or to revert to backup or alternate systems.

The modern aircraft design incorporates the best of HA planning from the very second its design hits the drawing board. As a passenger, it is acceptable to have physical component failure or even system failure; it is not acceptable to have complete service failure.

DR and HA come with different approaches to architectural data center and system design, different physical and design complexities, and most importantly for many different levels of cost. The physical redundancy on a motor vehicle, putting spare wheel in the boot, costs far less than the system redundancy on a commercial airliner.

No motor car that I know of comes with two engines in case one fails; in fact, the only duplicate system motor vehicles are designed with is duplicate lighting in case one is broken. In a recent survey by a motor warranty insurance company, the engine failure rate of the best manufacturer in their league table was 0.29% or 1 in 344 vehicles, and the worst manufacturer averaged an engine failure rate of just 1 in 13 engines [3].

All commercial airlines come with at least two ($N+1$) engines. That is a huge amount of reliability built into the system, especially when you consider that Airbus, one of the world's leading aircraft manufacturers, now claims that on average an aircraft engine on a modern commercial airliner will have a malfunction shutdown once every 30 years [4]. And while every pilot will be trained and tested on their ability to cope with such a shutdown, statistics shows that a pilot starting work for an airline today will likely never have to deal with an engine malfunction shutdown.

This difference is down to two key considerations.

While it is true that both are forms of transport, motor cars and commercial aircraft are designed for two totally different service and operational models. The risk profile of an aircraft failure is demonstrably different than that of the motor car.

As we will review in this chapter, the critical the things that can cause a data center outage are broadly split into seven key categories—from the physical building itself to the core IT infrastructure:

1. Building
2. Environmental (Power and Cooling)
3. Fire Systems
4. Access Control
5. Security
6. Service Provider Connectivity
7. IT Infrastructure

The physical aspects of data center protection, the physical building, environmental systems, fire systems, access control and security, and connectivity are well covered by many decades of development of best practice standards.

However, as with the investment in physical redundancy of aircraft systems, data centers are more likely to be brought down by pilot error than any physical malfunction or natural disaster—either pilot error in operating the physical environment or, more likely, pilot error operating the IT infrastructure that resides within it.

As statistics show that over 50% of all airline crashes are attributed to pilot error of some sort [5], many observers of the data center industry believe that some 75% of all data center "crashes" are brought about through some form of operator error [6].

Ultimately, in the data center, as with these two forms of transport, the approach to and levels of investment in DR and HA planning come down to acceptable *business risk*. The approach to reducing or removing this risk is then designed into the physical data center, building infrastructure, platforms, systems, processes, and people responses to potential failures.

35.2 THE EVOLUTION OF THE DATA CENTER AND DATA CENTER RISK

In the days of the early mainframe, the data center was simply the climate-controlled safe house for the gigantic and power-hungry early computers. The data center and its contents were firmly in the hands of a few special souls in the IT department. Most organizations only had one or two large central computers, so housing them, powering them, and securing them were relatively simple. In fact, in these early days, the output of the data center went no further than its four walls. The data center processed data fed into it locally, processed locally, and outputted it locally.

As we moved from green screen of the mainframe to the first green letters on the PC, the shift of importance from data center to desktop altered the balance of power in computing terms from the center to the periphery. With desktop computers running their own local disks, operating systems and applications, and servers distributed down to business units, wiring closets, and remote offices, the loss of the data center became less of a disaster and, for many, simply an irrelevance.

As the deployment models for IT systems changed from centralized to decentralized and back again, the data center's importance to the health and agility of an organization fluctuated. Lines of business decided they could build and operate servers and storage capacity more quickly than the centralized and often seemingly slow IT department. This often led to what are sometimes known as Distributed Data Center under Desks (DDuDs) with no centralized control, and more importantly, little proven operational processes, and minimal backup or DR capabilities available or used. This uncontrolled expansion of distributed computing spread the workload and spread the risk of a total data center (DC) or DC system failure. Who can remember a major headline story in the 1980s or 1990s about a data center or major IT system failure?

In the 2000s, post-Y2K hysteria, and as the decentralized IT models were recognized for the inherent business risks they posed and as regulation of IT systems by regulators increased the pressure on business to treat IT with the same levels of corporate governance they would other critical business disciplines, consolidation and control of IT moved the data center and the infrastructure it housed back into focus as a critical organizational asset.

This led to a refilling of the data center with legacy and new IT systems. From housing one big box, the mainframe, the data center became a place of computing complexity. Mainframes, midrange systems, x86 systems, storage systems, networks, Internet gateways, security appliances, email systems, print and file servers, and a veritable technology smorgasbord of platforms, systems, and applications found its way back into the data center. Many of these systems were interconnected, but most were not. Islands of technology were operated by different teams, and most technology strategy and acquisition decisions were made in splendid isolation.

From housing one important box to housing nothing of importance to being the center of the computing universe again, the importance of the data center grew more quickly than many organizations' ability to keep its operational integrity under control. No better example of this was data center cabling becoming the epitome of the lack of control in data center physical infrastructure design (Fig. 35.1)—so much so that many websites were dedicated to this phenomenon [7] and subsequently the industry of "structured" cabling in the data center became one of the hottest in the IT industry.

Over the past 5 years, we have seen yet another change in the data center.

With the advancements in server, storage, and networking technologies, the data center has become interwoven and enhanced with the logical and operational integrity of virtualized resources.

Physical servers have been consolidated and virtualized, with one physical server housing multiple virtual server instances or blade systems supporting hundreds or thousands of virtual machines (VMs) in a single system.

Storage platforms have been virtualized, so rather than connecting one server to one piece of storage, multiple servers access pooled storage, reducing the complexity of the storage environments, placing more load and more responsibility on larger storage environments.

And networks have been virtualized, so instead of individual switches supporting individual servers, VLANs subdivide security domains onto an aggregated network access layer.

The new enterprise architectures that lie within the data center today are inherently more robust and resilient. Physical resilience is now augmented by virtual resource redundancy, and it is now relatively simple to overcome the operational risk of a specific hardware component outage. However, while the spreading of risk across virtual systems reduces the risk of physical outage, the complexities of running pools of highly integrated systems bring with it its own challenges: operational and human.

Finally, the word "cloud" has entered the lexicon of the data center. Private clouds, built and operated within the clients' own data center, offer a new paradigm in data center service provisioning. Promising on-demand, flexible, scalable, and changeable server, storage, and network resources, private clouds combine physical and virtual resources alongside a new and more complex management and operational stack that create new efficiencies and complexities in data center enterprise infrastructure design.

But Murphy's law states that however robust and well designed our physical data center or data center architecture is, things do go wrong.

In today's world of IT, we need to address the loss of physical buildings, systems, and components of enterprise architectures and now have to embrace the loss of virtual servers, storage, and network entities that could be *virtually anywhere* and in any *current configuration*.

And a growing number of what would be considered "data center services" will not come from the organizations' own data center estate, but instead be piped like a utility from a contracted or possibly transient service provider. These piped-in cloud services could well provide the DR or HA service designed under the Business Continuity Plan [8].

In today's world of DR and HA, assessing, addressing, and coping with a mix of facilities, physical, virtual and now

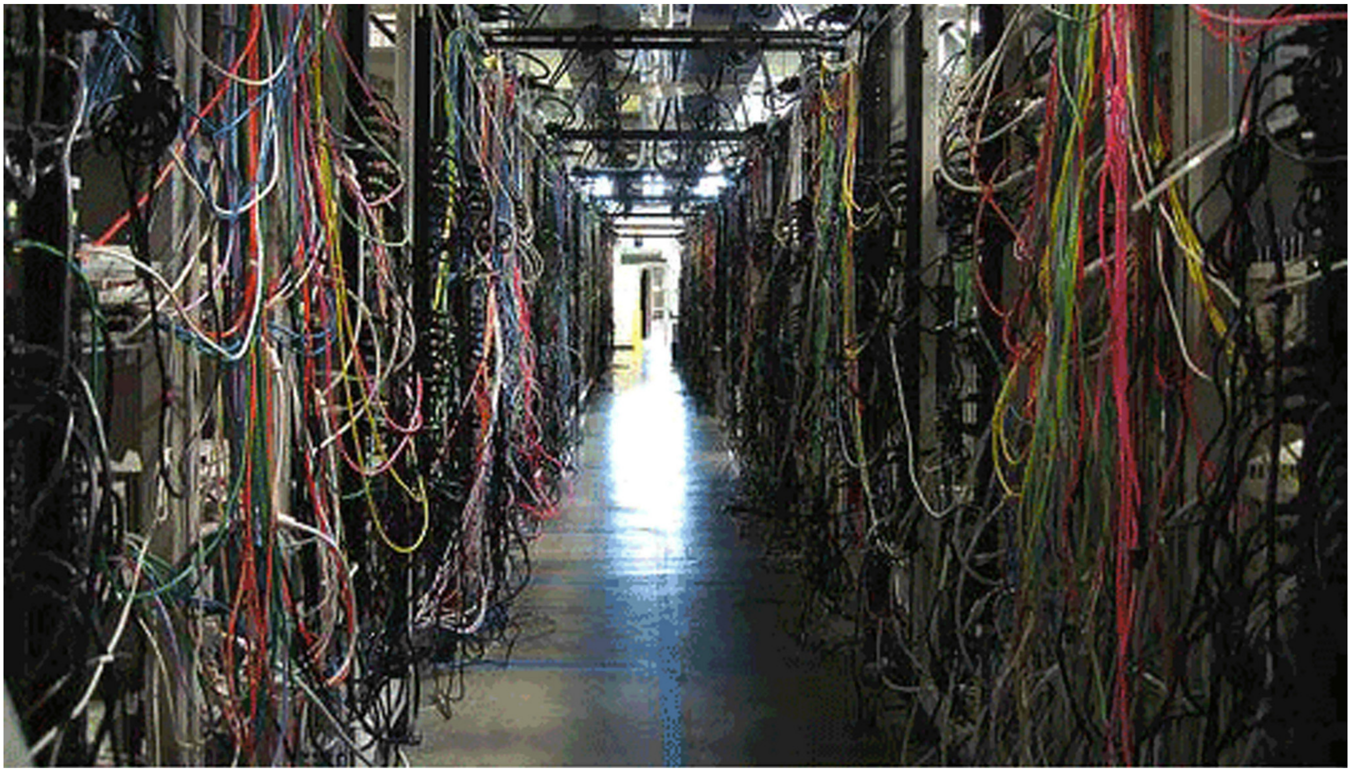


FIGURE 35.1 Data center wiring complexity. Courtesy of GeekAbout.com.

"as a service data" center services is either more complex and more challenging than anything the CIO, and IT department has faced before or much more simple and straightforward than we could have ever imagined; it depends on which perspective you are looking from.

The availability of high-performance bandwidth means that data centers no longer need to be sited within the same building, campus, city, or even country than the users it services. If your organization is in an area at risk of flood, hurricane, or other natural disasters, you simply site it elsewhere.

The historic complexity of a data center running many individual systems, relying on discreet server, storage, and network components, managed in different ways by varying degrees of operational maturity can be resolved through aggressive virtualization and the implementation of highly automated private clouds.

If you cannot afford to build a highly reliable and resilient data center for your own business or feel that owning the bricks and mortar isn't a core activity of your own business model, the choices of hosting provider able to spend many millions of dollars building a high-tier hosting facility grow daily in every region of the world.[1]

Ultimately, you can negate many issues of complexity and cost in building and operating your own data center or enterprise system infrastructures by outsourcing the provision of core infrastructure, operating systems or even applications to cloud providers, and consuming Infrastructure as a Service (IaaS), platform as a service (PaaS), or software as a service (SaaS) from a service provider who has the scale and operational capabilities most individual organizations can only dream of.

But while all of these options exist, this chapter recognizes that for many companies, designing and operating an internal data center infrastructure is still the only option and that we should look to mitigate against Murphy's law as much as we can, bringing the design, control, and ultimately destiny for data center DR and HA planning firmly into a business aligned and robust strategy.

We must assume that things can go wrong will go wrong, some of those things we can plan to cope with, and some of those things could have devastating consequences to our organizations.

The options to how we now address today's challenges of ensuring our organizations' heart does not stop and, that if it does, we use robust operational processes, procedures, and planning to recover have never been greater. Which ones we can afford or justify will come down to your organizations' individual circumstances and risk assessment and business continuity planning.

[1] http://uptimeinstitute.com/TierCertification/certMaps.php

TABLE 35.1 DR and HA strategy

What are the business drivers?	What is the nature of the business behind the infrastructure, system, or application that necessitates DR or HA capabilities?
What is the financial impact of an outage?	What does an outage cost? The cost of an outage justifies the number of layers of availability that are built into a physical environment or system.
How long can the facility, infrastructure, or system be down at any given time?	The answer to this question is usually: "It depends." However, the question is aimed at the busiest time, for example, during month-end close.
How long can the facility, infrastructure, or system be down in a year?	Since many of the availability metrics are normalized over a 1-year timeframe, accumulated annual downtime is often a measure of interest.
What exactly is an outage?	This question is more difficult to answer than it appears. The answer to this question gets to the heart of any availability measurement scheme
Should the system design revolve around the accumulated annual outage or the worst-case single outage?	This choice is made based on the nature and criticality of the application running on the highly available system. Further, a system designed to handle the worst-case single outage will have a substantially higher price tag.
Should the system design revolve around the accumulated annual outage or the worst-case single outage?	The answer to this question lends some focus to this exercise. How high is it highly available? This is the most important question to answer in building a customized availability model. The answer to this question sets a goal for the system designer. If there is no stated goal, then there is no way of knowing if any system is highly available. All the above.

35.2.1 Assessing the Impact and Cost of Data Center and IT System Downtime

The investment in any DR and HA strategy will be primarily driven by business and not technology imperatives.

An example of this is shown in Table 35.1, and there are numerous variants available from many online and professional sources such as the Business Continuity Institute.[2]

A survey by the power system manufacturer Emerson [9], completed in 2011, uncovered a number of key findings related to the cost of downtime (Fig. 35.2). Based on cost estimates provided by survey respondents, the average cost of data center downtime was approximately $5600 per minute.

Based on an average reported incident length of 90 min, the average cost of a single downtime event was approximately $505,500. These costs are based on a variety of factors, including but not limited to data loss or corruption, productivity losses, equipment damage, root-cause detection and recovery actions, legal and regulatory repercussions, revenue loss, and long-term repercussions on reputation and trust among key stakeholders. The two single biggest losses are business disruption and loss of revenue.

In the end, every business will suffer a different degree and cost of loss when a disruption occurs.

Loss can be viewed in many different ways, monetary loss, reputational loss, employee productivity loss, customer loyalty loss, and in many cases combinations of all of these.

[2]http://www.thebci.org/

A simple way of calculating your own potential lost revenue through data center outage is

$$\text{Lost revenue} = \frac{\text{GR}}{\text{TH}} \times I \times H$$

where GR = gross yearly revenue, TH = total yearly business hours, I = percentage impact, and H = numbers of hours of outage.

The loss of the data center or the services that reside within it will cost something, but the investment in the robustness of the data center you choose to build or rent space in, the investment in processes and systems to operate the facility, and the investment in IT infrastructure that has different degrees of resilience will also cost money.

The business case for building highly resilient data center facilities or IT infrastructure must be balanced against the risks you assess as relevant to your organization.

35.2.2 What Can Go Wrong Will Go Wrong: Why Do Data Centers Fail?

The cause of data center outages can be classified in two broad categories.

Natural disasters, such as floods, hurricanes, tornadoes, or earthquakes, are disasters most likely to come to mind when thinking about major outage occurrences. It is easy to imagine the devastating impact of a hurricane, a major power outage, or an earthquake. They are easiest to imagine and statistically most unlikely to happen.

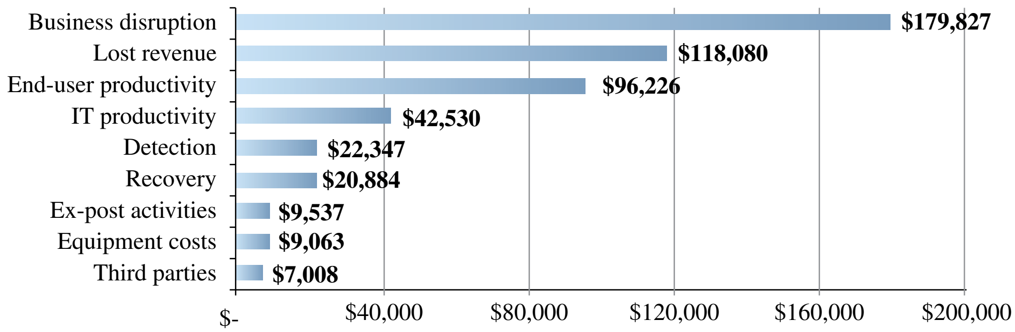


FIGURE 35.2 Average cost of unplanned data center outages from 41 benchmarked data centers. Courtesy of Emerson Network Power.

When they do, they can strike at the heart of the most well thought through business continuity planning.

In late October 2012, the Eastern United States of America and Canada braced itself for Hurricane Sandy, the most destructive Atlantic hurricane on record, measuring a diameter spanning 1100 miles. As New York City began to prepare itself for the storm, many data center operators began to implement their plans to ensure smooth operations for their customers. One such operator, Peer 1 Hosting, had positioned the generators for their 13,000 ft^2 hosting data center on the 17th floor of a lower Manhattan office block.

As the waters hit, the power generator, perched safely above ensuing disaster below, started up perfectly.

As the waters filled the basement and ground floors of the office block, the power to the main building was switched off by the local power company, leaving 20,000 gallons of fuel, also situated on the ground floor without an operating fuel pump. With no power to pump the fuel in the daily tank, and using 40 gallons an hour, the supply of fuel was fast running low.

A perfect plan to keep a data center running, hosting it hundreds of feet above street level and away from water, had failed, because "anything that could go wrong had gone wrong." In a Herculean effort to keep the hosting center that kept their own businesses working, customers assisted Peer 1 staff to carry gallons of fuel up the 17 flights of stairs until power could be restored.

In a highly fluid situation, a customer pointed out that this new and definitely untested DR plan was in effect working well. "It looks like we have *operationalized* this to the point where we can make it work—I can't honestly believe it."[3]

Sometimes, it is impossible to predict or plan against many of the arrows nature will throw at you. Finding willing clients to convey buckets of fuel up 17 floors is probably also an unrealistic strategy in most DR planning documents.

This example is simply a good indication of how even the best design and business continuity planning can be usurped by something exceptional. The Uptime Institute, a think tank and professional service organization based in Santa Fe, New Mexico, produced the Natural Disaster Risk Profiles for Data Centers,[4] which included advice to operators on the data center consequences expected from different types of severe weather. This document also includes a number of Natural Disaster Risk Location Maps for the United States (Fig. 35.3).

The second category of disaster is definitely within Murphy's law: man-made disasters.

Unless you are operating a "lights-out"[5] DC, which limits the intervention of human beings by keeping them away from the facility, the chances of an individual being the primary source of data center outage are a more likely cause.

A human being is perfectly designed to create data center outages. Brains and hands working together can spill something, turn off a switch, incorrectly configure a part of the enterprise architecture, or ignore well-designed processes created methodically to keep things running smoothly.

In extreme circumstances, although there are no recorded physical attacks that I can find, terrorists may target critical data center infrastructure as an act of commercial- or government-inspired terrorism. While the terrorist or activist threat mainly presents itself through activities such as remote Distributed Denial of Service (DDOS) attacks, governments and companies must assess and include data center infrastructure in their business/national critical infrastructure plans.[6]

While the overall threat of commercial- or government-supported cybercrime and cyberterrorism is unquestionably increasing, a review of the Top 10 Data Center Outages of 2012 by Data Center Knowledge Magazine (Fig. 35.4)

[3] http://www.computerworld.com/s/article/9233136/Huge_customer_effort_keeps_flooded_NYC_data_center_running

[4] http://uptimeinstitute.com/component/docman/doc_download/11-natural-disaster-risk-profiles-for-data-centers

[5] http://www.techopedia.com/definition/26965/lights-out-data-center

[6] http://www.cpni.gov.uk/documents/publications/2010/2010006-vp_data_centre.pdf?epslanguage=en-gb

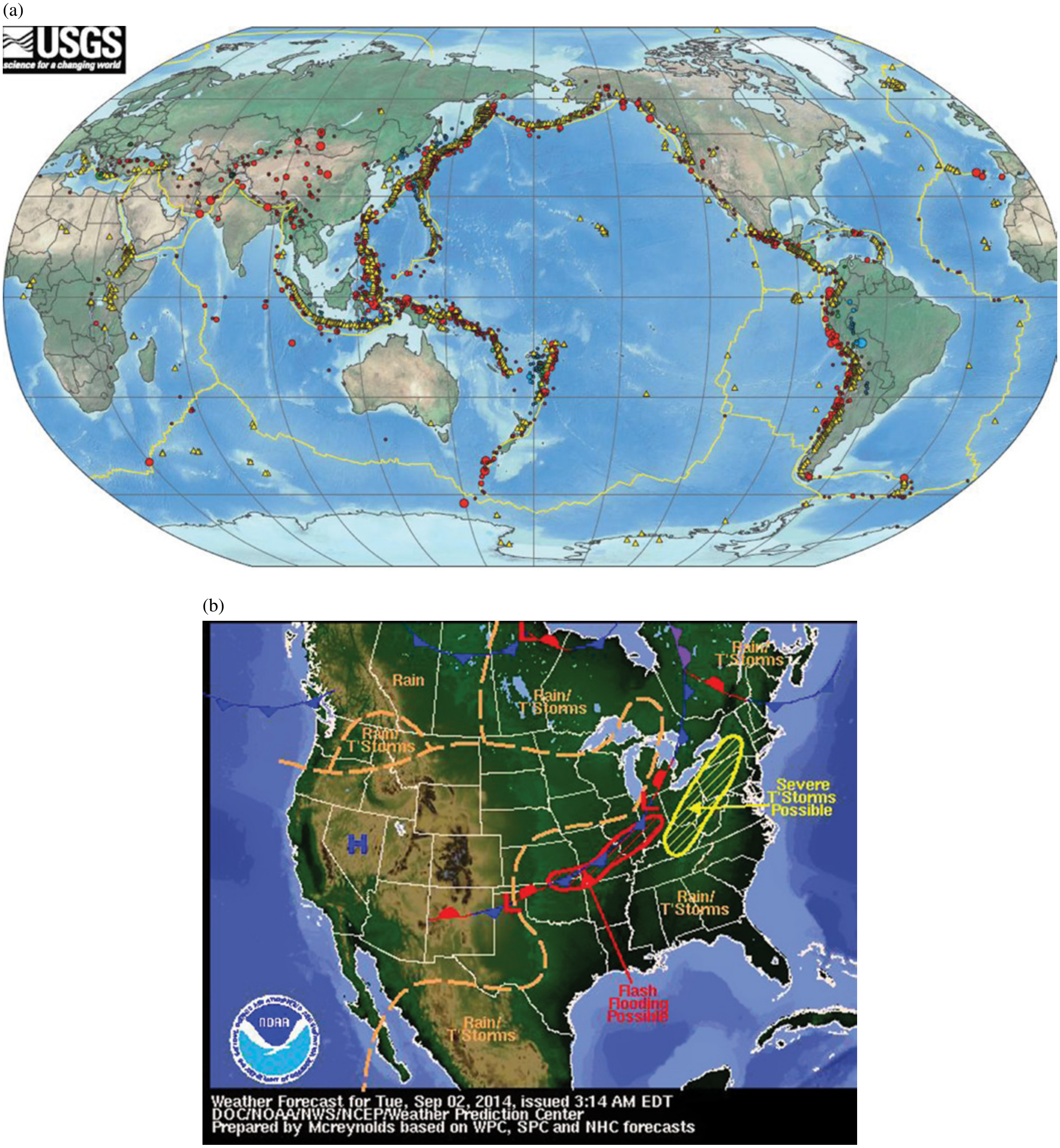


FIGURE 35.3 (a) World seismic map. Courtesy of USGS. (b) U.S. weather forecast. Courtesy of NOAA.

demonstrated that accidental human error is more likely than human terror [10].

As is apparent from this unscientific list of highly publicized data center outages, over half were caused by configuration error—whether that was in the programming of systems, manual maintenance of a data center subsystem, or simply the configuration of a piece of the internal data center computer architecture.

Ranking	Data centre outage	Explained or suggested cause
1	Ney York city (November 2012)	Hurricane Sandy
2	Go daddy DNS outage (September 10)	Internal events corrupted router data tables
3	Amazon outage (June 29–30)	Power outage
4	Shaw communications, calgary data centre (July 11)	Fire
5	Australian airline reservation system (July 1)	Date-related system
6	Windows Azure cloud (February 29)	Date-related security certificate
7	Saleforce.com (July 10)	Power outage
8	Syrian internet blackout (November 29)	Government or terrorism cited
9	Windows Azure cloud (July 28)	Traffic management configuration
10	Hosting.com (July 28)	Maintenance human error on UPS

FIGURE 35.4 Top 10 data center outages of 2012—Source: Data Center Knowledge Magazine.

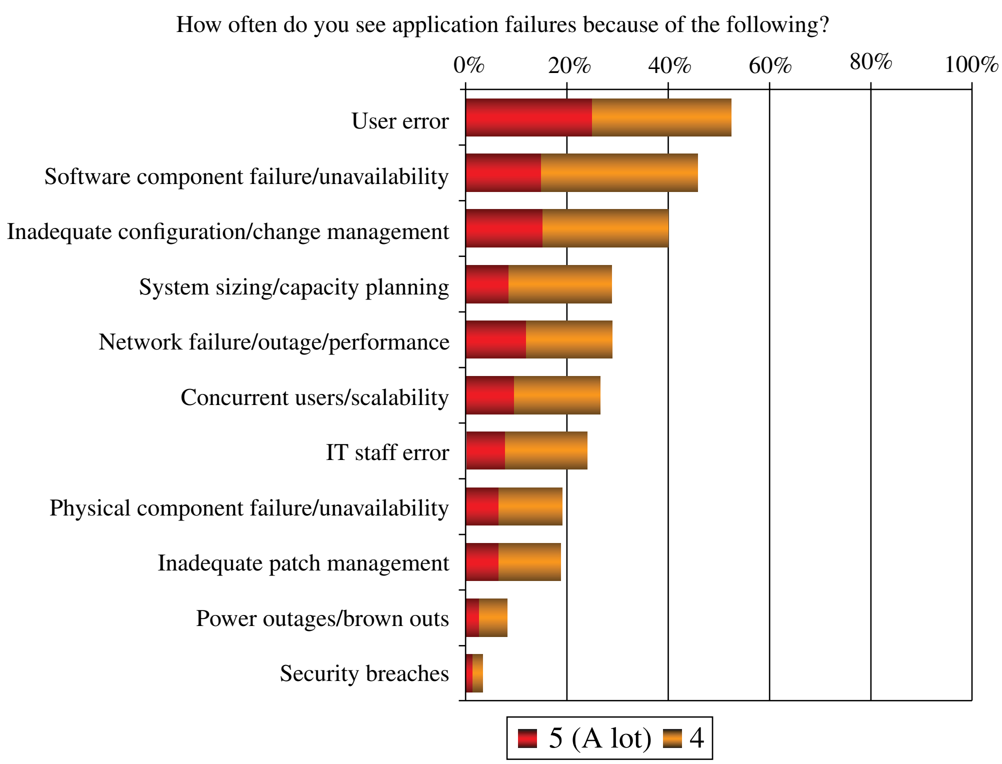


FIGURE 35.5 Application outages survey, March 2008. Courtesy of Freeform Dynamics, March 2008.

Only one category of these outages was attributed to potentially poor design of the data center physical infrastructure or subsystems, and those are grouped together under the event headed Hurricane Sandy.

Given the scale of many of these data centers, we can also assume that most of the systems residing in them were designed to HA specifications; yet they still suffered from outages.

In 2008, in the Freeform Dynamics survey of why application failures or outages occur, human failure placed top of a list that factored power outages as the least likely cause (Fig. 35.5).

In fact, the survey points to a series of human errors in software outages, user error, inadequate configuration/change management, system sizing/capacity planning, IT staff error, or inadequate patch management, and one could argue security breaches are all application outages caused in some way by human error. Out of the 11 outages listed, only 3 are physical failures, and they come in at number 5 (network failure), number 8 (physical component failure), and number 10 (power outages/brownouts).

In 2012, a major U.K. financial intuition suffered a system failure to one of their primary banking systems. The consequences of this outage became a national news story, an external public relations nightmare for the organization, and a lesson to all involved in IT operations.

You can build as much redundancy and resilience into IT systems, but the simple application of a corrupt upgrade to a system can result in chaos. The Wikipedia entry for this outage (so famous it has a Wikipedia entry) read, "Completion of new home purchases were delayed, others were stranded abroad, another was threatened with the discontinuation of their life support machine in a Mexican hospital, and one man was held in prison."

What is essential is that your organization carries out a DR risk assessment, a process of evaluating and rating all of the threat, the probability of them happening, and the impact they will have on your organization. An example courtesy of TechTarget (Table 35.2) gives some sense of the range of risks you may need to assess depending on the range of internal and external threats most likely to occur to your facilities or infrastructure.

And never forget that a chain of events is most likely to create outages and damage.

The chances of a terrorist incident directly destroying your data center are incredibly small if not negligible. The chances of a terrorist incident closing down a part of a city, interrupting transport links, and stopping your data center or infrastructure operational staff from getting to work are for many cities a highly possibly scenario.

The odds of a plane landing on your data center is virtually nonexistent, but the potential for a bird flu pandemic to deter your staff from wanting to leave their homes or congregate together is growing as the strains of this disease continue to mutate around the world.

However, the statistics shows that the current probability of the threats and the probability of threat such as flooding, civil disobedience, pandemics, terrorism, or extreme weather conditions to bring down your data remain for most business as unlikely to happen as winning the lottery.

The chances that at some point, somebody flicking or a switch or miskeying a configuration change will take out your IT systems for minutes, hours, and possibly days are pretty certain and something most gamblers would see as a sure bet.

35.3 PHYSICAL DATA CENTER DESIGN AND REDUNDANCY: TIERS AND *N*+ WHAT?

Engineering and best practice design of the physical data center has not been left to Mother Nature or chance.

Over many years pioneer by the Uptime Institute, much work has been undertaken to standardize the design of the physical data center and describe the redundancy engineered into it and its underlying support systems.

The simplest is a Tier 1 data center defined by the Uptime Institute, which is basically a server room to the most stringent level, which is a Tier 4 data center, designed to host mission-critical computer systems, with fully redundant subsystems and compartmentalized security zones controlled by biometric access control methods. The Uptime Institute has further defined and refined its own four Tiers,[7] which are now used as a commercial differentiator by many in the data center hosting industry. The Tiers are listed in Table 35.3.

As of April 2013, the Uptime Institute had awarded 236 certifications for built data center environments around the world.

> The core objective of Uptime Institute Tiers is to guide a design topology that will deliver high levels of availability—as dictated by the owner's business case. Uptime Institute Tiers (Figure 35.6) evaluates data centers by their capability to allow maintenance and to withstand a fault.

In very simple terms, the higher the classification, the more likely it is to be offering a functioning service during a system, power, or maintenance outage. The majority of end user data centers tend to fall into a Tier 2 classification, or perhaps larger end users would choose to opt for a Tier 3 classification, whereas data center hosting providers or cloud providers would seek to provide their service from a Tier 3 or Tier 4 center.

The Uptime Institute provides a wealth of technical documentation, too much to review in this single chapter; but as we have discussed in our motor vehicle and aircraft analogy, it is worth reflecting the degrees of difference between the different tiers of classifications.

[7] http://uptimeinstitute.com/TierCertification/

TABLE 35.2 Example probability and risk assessment[a]

Threat	Probability (P)	Impact (I)	Risk = $P \times I$
Flooding—internal			
Flooding—external			
Fire—internal			
Fire—external			
Severe storms			
Wind storm			
Earthquake			
Tornado			
Hurricane			
Snow storm			
Ice storm			
Hail			
Drought			
Tsunami			
Mud slide			
Epidemic			
Pandemic			
Explosion			
Gas leak			
Structural failure, for example, bridge collapse			
IT—system software			
IT—applications			
IT—hardware			
IT—viruses			
IT—hacking, unauthorized intrusions			
IT—communications, connectivity			
IT—vendor failure			
IT—operational (human) error			
Utilities—water			
Utilities—sewage			
Utilities—electricity			
Utilities—gas			
Utilities—steam			
Utilities—communications			
Terrorism—biological			
Terrorism—chemical			
Terrorism—radiological			
Terrorism—nuclear			
Sabotage			
Bomb threat			
Criminal—theft			
Criminal—break-ins			
Criminal—vandalism			
Criminal—espionage			
Criminal—hostages			
Criminal—murder, rape, assault			
Criminal—bribery			
Work stoppage			
Work action, strike			
Civil disorder			
Human error			
Others			

[a] Courtesy of TechTarget.

One can see without the need for specific component detail that simply the electrical system design of redundancy between a Tier 1 (Fig. 35.7) and Tier 4 (Fig. 35.8) data centers is considerable.

TABLE 35.3 The Uptime Institute Tiers

Tier	Requirements
1	Single nonredundant distribution path serving the IT equipment Nonredundant capacity components Basic site infrastructure with expected availability of 99.671%
2	Meets or exceeds all Tier 1 requirements Redundant site infrastructure capacity components with expected availability of 99.741%
3	Meets or exceeds all Tier 1 and Tier 2 requirements Multiple independent distribution paths serving the IT equipment All IT equipment must be dual powered and fully compatible with the topology of a site's architecture Concurrently maintainable site infrastructure with expected availability of 99.982%
4	Meets or exceeds all Tier 1, Tier 2, and Tier 3 requirements All cooling equipment is independently dual powered, including chillers and heating, ventilating, and air-conditioning (HVAC) systems Fault-tolerant site infrastructure with electrical power storage and distribution facilities with expected availability of 99.995%

The primary definition of redundancy in a physical data is described in two variants: $N+1$ or $2N$.

The definition of the two has best been described in a blog from Data Center Mapping, courtesy of Sarah Pollock[8]:

> The simple way to look at $N+1$ is to think of it in terms of throwing a birthday party for your seven year old. Say you have ten guests and need ten cupcakes, but just in case you have an unexpected guest show up, you order eleven cupcakes. N represents the exact amount of cupcakes you need, and the extra cupcake represents the +1. Therefore you have $N+1$ cupcakes for the party. If you order 12 cupcakes, then you have $N+2$.
>
> Although an $N+1$ system contains redundant equipment, it is not, however, a fully redundant system and can still fail because the system is run on a common circuitry or feeds at one or more points rather than two completely separate feeds.
>
> Let's re-visit the birthday party analogy again.
>
> If you plan a birthday party with a $2N$ redundancy system in place, then you would have the ten cupcakes you need for the ten guests, plus an additional ten cupcakes. $2N$ is simply two times or double the amount of cupcakes you need.
>
> At a data center, a $2N$ system contains double the amount of equipment needed that run separately with no single points of failure. They are far more reliable than an $N+1$ system because they offer a fully redundant system that can be easily maintained on a regular basis without losing any power to subsequent systems.

[8]http://uptimeinstitute.com/TierCertification/

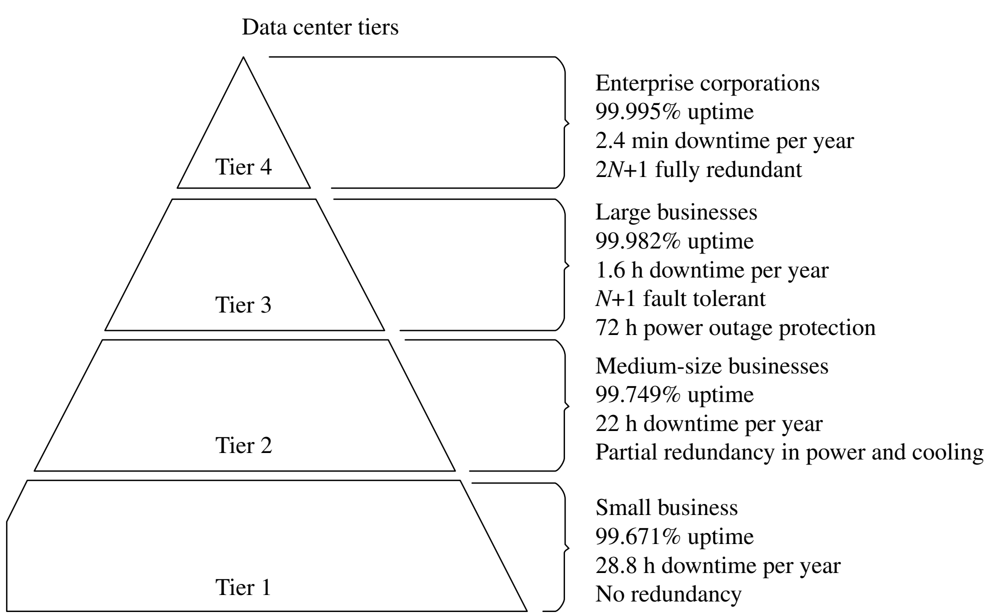


FIGURE 35.6 Data center tiers aligned to use cases (http://www.colocationamerica.com/data-center/tier-standards-overview.htm).

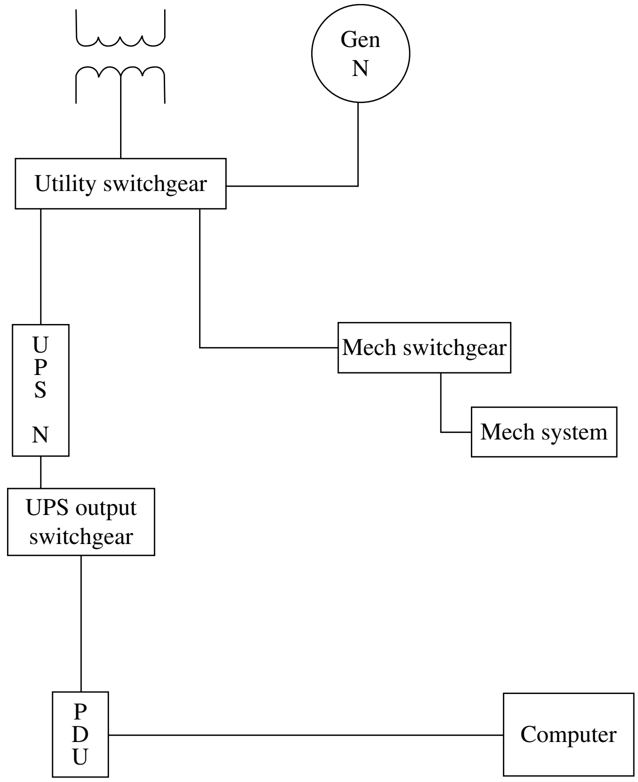


FIGURE 35.7 Uptime Institute Tier 1 data center indicative electrical system design.

It is both impossible to cover all aspects of physical data center redundancy design within a single chapter, but as you have seen, the wealth of material and best practice advice available is considerable and well proven around the world.

There is no need to leave the physical redundancy of your data center design to chance.

The approach to whether your organization's business needs demand a DR- or HA-based data center design is generally one of cost. Your choice of tier, or of which tier your partner data center hosting partner has built, will impact the cost per square feet of space at your disposal, but should not be impacted by a lack of best practice design guidance available from industry associations, professional bodies, or other data center users.

35.4 VIRTUALIZATION BRINGS OUT-OF-THE-BOX DR SURVIVABILITY

According to research by the Taneja Group,[9] 55% of users still use tape as their primary backup/DR medium, an approach designed to keep data safe and restorable in the event of facility of system failure. The speed at which a tape retrieved from an off-site storage facility and restored was the approximate guide to how quickly a company could respond to a data center or system disaster.

This process for recovery meant IT staff managing data backup software, schedules, tape libraries, and off-site archiving facilities or service providers. Complex processes must be coordinated to separately recover and reconfigure servers and data sets, often in multiple locations, and as a result, recovery times are often too long and unpredictable.

For many organizations, this is simply not fast enough, and investment in data center-to-data center (or site-to-site) replication between storage systems provided a faster, if not more expensive, recovery times.

Data replication between data centers—either synchronous or asynchronous—shifts the cost of slower recovery speed into new costs of hardware and an additional cost burden onto the wide area network, where the high cost of bandwidth remains the prime barrier to deploying widespread replication for DR.

But the watchword for many organizations, and the reason many have inadequate DR plans today, is duplication.

Not only do organizations have to duplicate data, but they duplicate servers, networks, security appliances, and a range of supporting technologies. Making an exact copy of everything doesn't just double the price of an infrastructure; it can make it several times more expensive to build systems, and the cost of designing, architecting, and building replica DR systems can have a dramatic increase.

And for 99.999% of its life, what happens to it? Nothing!

This entire recovery infrastructure sits idle waiting for the moment to come when something at the primary data center fails and it springs into life for the hours, days, or weeks necessary for primary systems to recover.

Firstly, this approach is expensive.

It is the same in essence as buying a second home and keeping it heated, furnished, and maintained—just in case your primary residence burns down. But, in today's more complex world of IT systems, it is not simply a very expensive form of insurance policy; it is becoming more difficult to judge whether when things do go wrong that your systems will be recovered in the way that you need or intended them to.

Second, as many companies did not bother to check the quality of the data on their tape backups, a growing number of organizations are unable or unwilling to test their DR plans and systems.

Why? Oddly, they do not test, for the very reason they built them in the first place, risk; risk that the very testing of the DR plan, etc.

Risk that the very testing of the DR plan or system will in itself cause an outage. Because of the growing complexity of systems that have been built within the modern data center.

Taking down sub-systems or entire physical environments to test that the DR systems and processes are working

[9] http://www.colocationamerica.com/data-center/tier-standards-overview.htm

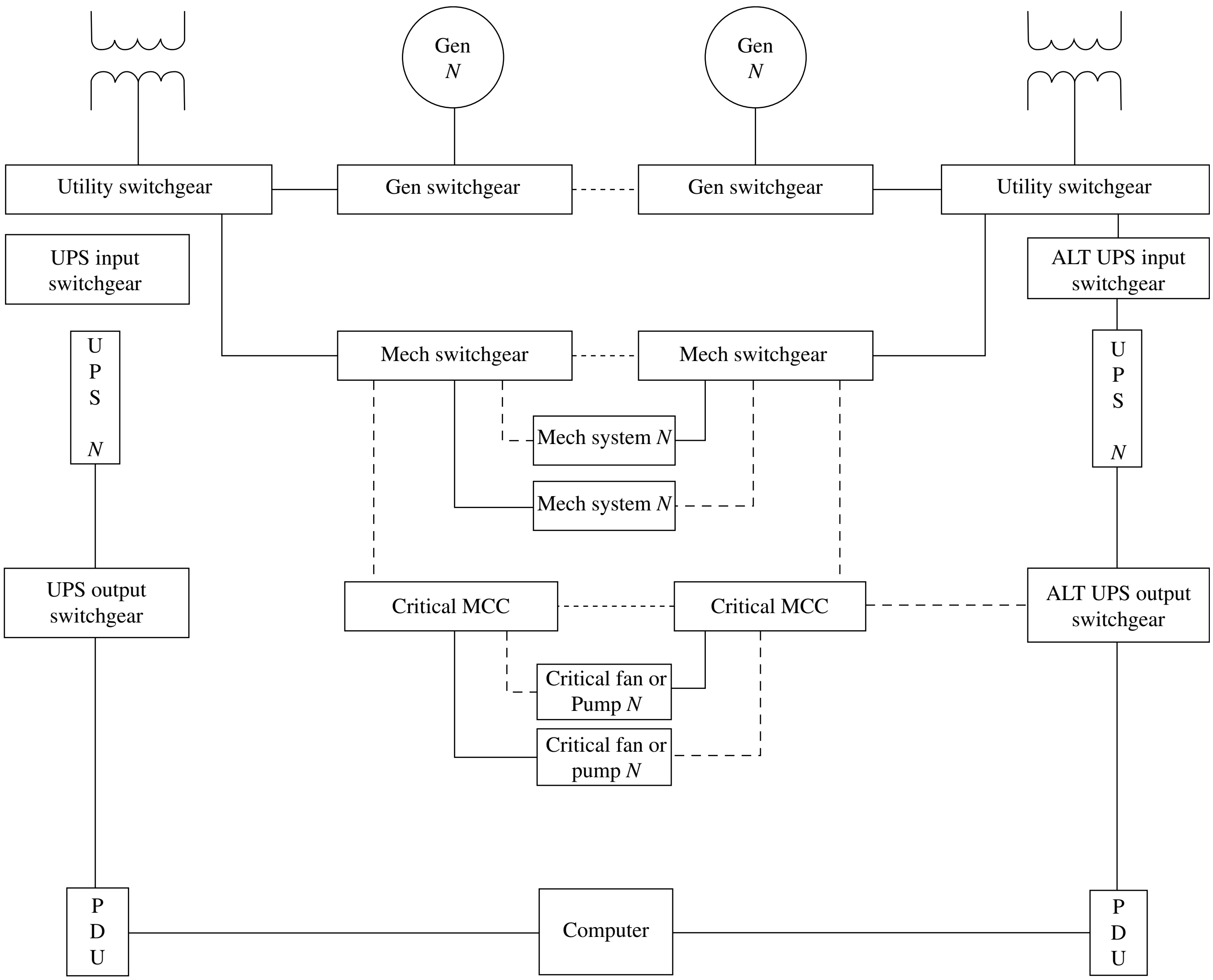


FIGURE 35.8 Uptime Institute Tier 4 indicative electrical system design.

as planned is something many IT departments and their board of directors are now shying away from. A fire alarm test is disruptive enough for many employees; imagine a full and proper test of a DR plan.

In today's complex working environments, where IT server and storage systems run every more complex applications and deliver services to desktops, laptops, tablets, and mobile devices and where systems are reliant on other core components and systems, the fully tested data center DR plan, and I mean fully tested where every system is taken down and every user is proven to be able to continue to operate in the manner envisaged in the DR plan, is possibly unachievable for most organizations.

The risk is simply too high.

But the advent of enterprise infrastructures based on virtual technologies and advancements in the use of management tools that allow the automation and orchestration of very complex previously manually delivered processes are making a DR test something far less risky and far more achievable.

In fact, the layering of virtualization on top of server, storage, and networking hardware means that replicating services across different local or remote infrastructures has become a natural design discipline.

One could argue that a virtual server is in a natural DR or HA state all the time.

Moving a VM between blades in a chassis, between chassis in a data center, or between data centers is becoming simpler by the day.

Companies such as VMware routinely discuss DR of servers not as a separate discipline but as part of their core solutions and product offering[10]:

[10] http://www.vmware.com/business-continuity/disaster-recovery.html

> Many organizations today do not have adequate disaster recovery (DR) protection for their applications. In most cases, disaster recovery is perceived as too expensive, complex and unreliable for any but the most mission-critical applications.
>
> Disaster recovery is a form of insurance to protect your IT assets when a disaster strikes. And just like good insurance, the best disaster recovery should provide great protection, with minimum hassle, at the lowest possible cost. VMware provides the most reliable, cost-effective, and simple disaster protection for all virtualized applications.
>
> Using VMware, organizations can effectively meet core requirements for disaster recovery:
>
> - Rapid recovery with automation
> - Reliable recovery, non-disruptive testing automation and simplified testing of recovery plans
> - Affordable recovery without requiring a duplicate, idle data center

VMware and almost all of the enterprise and open-source virtualization engines now directly bring an out-of-the-box approach to improving your data center recovery posture, with them not simply addressing the technical challenges of recovering individual services but also trying to address the whole recovery plan, recovery testing, and illuminating many of the manual processes that would need to be initiated in an outage situation:

> - Centralized recovery plans: With vCenter Site Recovery Manager, setting up a centralized and automated recovery plan is simple and can be done in a matter of minutes through an interface that is tightly integrated with vCenter Server.
> - Automated failover and site migrations: vCenter Site Recovery Manager automates the entire site recovery and migration process. Upon initiating a disaster failover, business services are automatically recovered with limited or no manual intervention.
> - Non-disruptive testing: With vCenter Site Recovery Manager, failover testing can be performed as frequently as required and is non-disruptive to production systems. Organizations are able to quickly identify any problems with recovery plans to enable fast resolution.
> - Broad choice of replication options to best align costs with business requirements Use built-in vSphere Replication for affordable replication, and storage replication for large, business-critical environments. vCenter Site Recovery Manager supports a broad range of storage-based replication products from VMware's storage partners.

And this is not just limited to the inherent DR capabilities built into virtualization software on servers.

The servers themselves are becoming smarter and more aligned to providing off-the-shelf inherent DR capabilities.

A growing number of organizations are now building agile infrastructures that can be repurposed on the fly, from being test development resources to DR resources in minutes if not seconds, and not simply supporting those applications that can reside on a VM. We can now move "physical" hardware between data centers in minutes.

The advent of stateless computing blades, such as the Cisco Unified Computing System (UCS), now allows virtual hardware to be moved at will. This is a new and exciting development in the world of computing and DR, as we can now lift and reinstate whole application sets with minimal effort and complexity.

In Cisco UCS, the characteristics of a blade server are held not in the physical hardware, but in programmable software, called a service profile. The Cisco service profile (Fig. 35.9) allows key definitions to be changed in software, allowing a spare blade server in a chassis to be reprogrammed to take on the persona of another blade that may have failed. With service profiles, the newly programmed blade is an exact replica of the failed blade, a clone, from its configuration to the network address, world wide name, and all other attributes.

In essence, the service profile can be moved between similar blade types in a single chassis and between chassis and therefore between data centers. This means some dramatic changes to the way DR infrastructure can be architected and operated. It also means, and more technical reading may well be necessary to grasp the benefits of service profiles, whole application stacks can be recovered in different data centers in the time it takes to make a cup of coffee.[11]

Storage vendors such as NetApp and EMC are bringing the cost and complexity of data replication down in similar ways.

NetApp's SnapMirror® technology claims to be able to drive massive technology and operational efficiencies into data replication while reducing the overhead these technologies have historically had on the network between data centers. NetApp claims a 60% total cost of ownership saving on traditional storage mirroring technologies, and as the storage industry has advanced its use of data compression techniques, the amount of storage you need for your primary and replicated data continues to reduce year on year.[12]

The network has also virtualized, and we are not moving into an era of programmable networks, Software-Defined Networks (SDN).

These new programmable virtual networks create a new powerful option to help improve DR responses. Because a virtual network can span more than one data center, the

[11] http://www.tanejagroup.com/

[12] http://www.vmware.com/solutions/datacenter/business-continuity/disaster-recovery.html

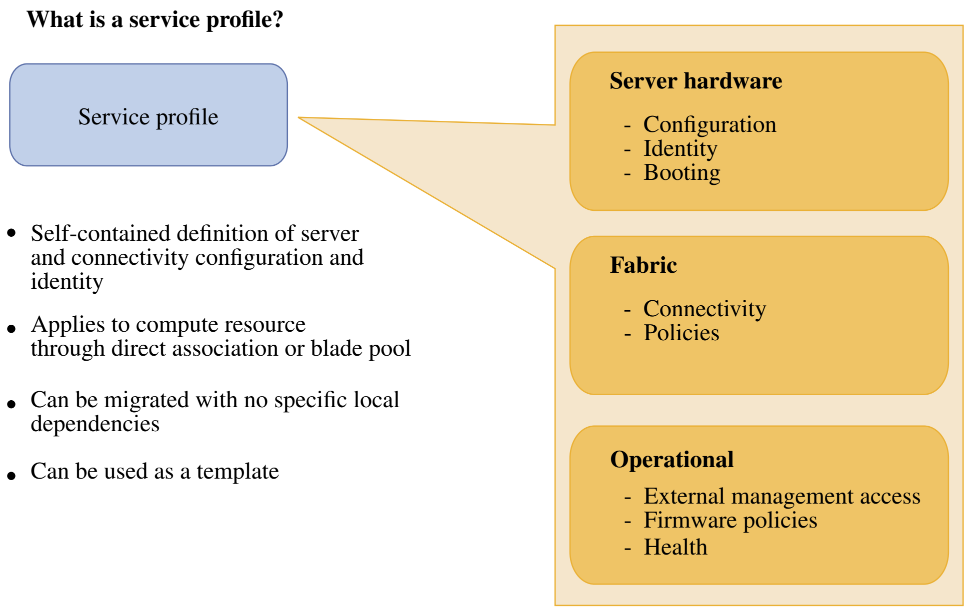


FIGURE 35.9 Cisco UCS service profile definition. Courtesy of Cisco System, Inc.

network can literally be in two places at once. The following views on how SDN can help reduce cost and complexity of DR and HA environments were provided by SDNCentral, a website dedicated to the SDN community.[13]

The DR capabilities of virtual networks span a number of key challenges, which have been historically difficult or expensive to overcome with nonvirtual networks.

- WAN Traffic Optimization. Virtual network architectures mean traffic is handled at the virtual routing layer, and this both increases the flexibility of network design while reducing the amount of networking hardware needed to create complex environments, and the virtual network can naturally optimize traffic to a DR or HA site.
- Virtual Network Interface Cards (vNICs). Instead of allocating one or two ports on a device, administrators are able to dedicate numerous virtual ports to a service that may be required. Furthermore, this can help with moving live VMs from one site to another and facilitating the HA or DR portions of an environment.
- DR Testing and Development. Virtualization has already helped in this area. However, from a DR and HA perspective, SDN can help even further. Imagine having the ability to recreate an entire networking environment to mirror an existing infrastructure. The difference? Everything is network virtualization and completely isolated. SDN can help create connections between applications, services, VMs, and numerous other workloads. Effectively, administrators are able to test their environment, DR plan, or HA methodology completely from a secured and isolated configuration. In many situations, you can even emulate the end user environment to create a truly powerful testing platform.
- Utilizing Load Balancers. With network virtualization, new types of load balancers are helping not only with traffic control but also with DR and HA. Tools like Global Server Load Balancing (GSLB) not only port users to the appropriate data center based on their location and IP address—it can assist with a DR plan as well. By setting up a GSLB environment, users can be pushed to a recover data center completely transparently should an emergency occur. This virtual cross-WAN heartbeat would check for availability of a data center and push users to a more available one if there is a situation.

[13] http://www.cisco.com/en/US/prod/collateral/ps10265/ps10281/white_paper_c11-590518.pdf

There are clear benefits to working with network virtualization. Organizations are able to be more agile, control their networking footprint better, and scale their infrastructure in-line with the business needs. More network manufacturers are creating options for virtual networks. The SDN market is expanding in large part because of the push around virtualization. IT goals are set around efficiencies and consolidation—and utilizing network virtualization technologies is helping organizations accomplish those goals. By reducing physical footprints and increasing the agility of a networking infrastructure by introducing SDN components, companies can continue to build robust, DR-ready environments.

For many organizations, the inherent benefits of virtual technologies bring a clearer and less complex approach to the issue of DR and HA.

On the basis that the chances that your data center will unlikely burn down and on the assumption that the mechanical and environment systems within your Tier 1, 2, 3, or 4 data center respond appropriately to a failure, the loss of IT enterprise hardware is the most likely cause of a physical malfunction resulting in outage or loss.

In building consolidated virtual architectures, managed by automated and orchestrated management tools, which allow complex manual processes, of which many are needed in a failure scenario to be ready to be initiated at the touch of a button, the percentage chances of having an IT infrastructure that never fails are now extremely high.

35.5 DR AND CLOUD

Is outsourcing and cloud the simple solution to the problem of DR and business continuity?

Why own and operate any of your own DR infrastructures when a growing number of organizations from data center hosting companies to managed service providers and cloud providers will do it for you?

The growth in the disaster recovery as a service (DRaaS) market has outstripped most other IT service markets in the past few years.

According to research by Markets and Markets,[14] the global DRaaS and cloud-based business continuity is forecasted to grow from $640.8 million in 2013 to $5.77 billion by 2018, at a CAGR of 55.2%.

Many organizations already outsource a number of major aspects of their organizational DR plan—the most widely used outsource being the provisioning of "hot desks" by third-party companies. If your office is flooded, a percentage of your staff are normally allocated places in shared facilities operated by specialist companies, many of whom also offer relocated data center space. This market has itself been impacted by technology, with more and more organizations having remote working as a normal part of daily life for a growing number of staff. Work is a thing people do, not somewhere they go, is now the mantra of organizations that are using technology to provide both better work–life balances to their employees and also cut down on the amount of physically office real estate they need to own.

The emergence of "cloud" technologies, hosted, agile, consumption-based IT services, primarily covering IT infrastructure with IaaS and PaaS and SaaS, now offers a growing range of individual and integrated DR services.

Backup as a Service was one of the first cloud-based services used by many business and even government customers.

Whether you're using simple file-based software tools or more complex image-based appliances, these backup services move your data into secure cloud storage where it can be retrieved almost immediately it is needed. You own none of the infrastructure and incur no operational overhead in running all of the backup systems that keep your data safe and secure. When a major or minor disaster occurs (a user deleting a file or corrupted data), the affected files can be brought back from your cloud service provider.

DRaaS adds several more layers of service to create a fully shrink-wrapped approach to consumption-based DR. Instead of just storing backup files off-site where they can be restored to your premises, DRaaS offerings are underpinned by cloud-based computing services, allowing you to bring up a "new" environment in the cloud as a series of VMs along with your data. DRaaS offers compute and data all in one service offering and generally paid for out of operational, not capital, expenditure. Many DRaaS services offer only VM-based services; but, in this growing market, no matter what systems you have (x86, RISC, or mainframe), there are providers of DRaaS services for most mainstream compute platforms and operating systems.

The market for cloud-based DRaaS services is maturing, as does the maturity of the wider cloud market in general, and the lure of not having to buy, build, or operate your own internal DR infrastructure is highly appealing to many organizations, especially smaller businesses that may not have the resources, skills, or money, to afford even the most rudimentary DR infrastructure.

For many, it will become the ultimate insurance policy against data center of enterprise infrastructure failure, for a payment of the fraction of the cost of doing it yourself.

Consuming DRaaS from cloud providers brings another advantage. DRaaS providers work at scale, and this scale forces them to invest in greater levels of DR and HA planning themselves. Each customer should benefit from accessing their service from a more robust environment they themselves could individually afford.

And cloud service providers use the kind of HA Tier 3 and 4 data centers described in an earlier part of this chapter; so in theory, you have access to systems on demand, run by service providers who invest heavily and operational excellence and security and whose data centers generally meet the highest build and operational standards.

But, like all service contracts, choosing the right provider, with the right service levels, who has invested in best practice processes and systems and who is going to be there when you need them most is critical.

Moving from using backup as a service to full DRaaS is a big leap, and your business continuity plans should assess the risks of your cloud provider suffering from all of the potential outage scenarios you thought you might have in the first place.

The cloud DR strategy is going to be very relevant to many organizations, but outsourcing a problem comes with its own challenges.

But, over the coming years, the growing maturity of cloud-based services can only mean that DRaaS in the cloud

[14] http://www.netapp.com/us/system/pdf-reader.aspx?m=snapmirror.pdf&cc=us

may well become the primary DR strategy for the majority of small-to-medium business and even a growing number of very large commercial organizations.

REFERENCES

[1] Anonymous. Disaster recovery, Wikipedia. Available at http://en.wikipedia.org/wiki/Disaster_recovery. Accessed on August 1, 2014.

[2] Anonymous. High availability, Wikipedia. Available at http://en.wikipedia.org/wiki/High_availability. Accessed on August 1, 2014.

[3] Holden C. German cars "among worst for engine failures." Campbell: Auto Express. January 2013. Available at http://www.autoexpress.co.uk/car-news/consumer-news/62383/german-cars-among-worst-engine-failures. Accessed on September 20, 2014.

[4] Handling engine malfunctions. Blagnac Cedex: Airbus Customers Services. December 2006. Available at http://www.skybrary.aero/bookshelf/books/193.pdf. Accessed on September 20, 2014.

[5] Causes of fatal accidents by decade. PlaneCrashInfo.com. Available at http://www.planecrashinfo.com/cause.htm. Accessed on August 1, 2014.

[6] Bigelow S. The causes and costs of data center system downtime: Advisory Board Q&A. TechTarget. Available at http://searchdatacenter.techtarget.com/feature/The-causes-and-costs-of-data-center-system-downtime-Advisory-Board-QA. Accessed on September 20, 2014.

[7] GeekAbout website. 40 Most disastrous cable messes. Available at http://www.geekabout.com/2008-02-19-479/40-most-disastrous-cable-messes.html; http://www.geekabout.com/2008-02-19-479/40-most-disastrous-cable-messes.html. Accessed on September 20, 2014.

[8] Business continuity planning. Wikipedia. Available at http://en.wikipedia.org/wiki/Business_continuity_planning. Accessed on August 2, 2014.

[9] Sponsor by Emerson Network Power. Calculating the cost of data center outages. Traverse City: Ponemon Institute LLC. February 2011. Available at http://www.emersonnetworkpower.com/documentation/en-us/brands/liebert/documents/white%20papers/data-center-costs_24659-r02-11.pdf. Accessed on September 20, 2014.

[10] Miller R. The year in downtime: Top 10 outages of 2012. Cincinnati: Data Center Knowledge. December 2012. Available at http://www.datacenterknowledge.com/archives/2012/12/17/the-year-in-downtime-top-10-outages-of-2012/. Accessed on September 20, 2014.

FURTHER READING

Brocade/Hitachi. Beyond disaster recovery—build datacenter resilience. Hitachi. 2007. Available at http://www.brocade.com/downloads/documents/solution_briefs/HDS_Brocade_Brochure_Web.pdf. Accessed on May 23, 2014.

Introduction to business continuity planning. Sans Institute. 2014. Available at https://www.sans.org/reading-room/whitepapers/recovery/introduction-business-continuity-planning-559. Accessed on May 23, 2014.

Smith J. Business. 2010. Continuity planning & disaster recovery planning. Purdue University. Available at https://www.google.com/#q=smith+j+business+continuity+planning+%26+disaster+recovery+planning+purdue+university. Accessed on August 16, 2014.

36

LESSONS LEARNED FROM NATURAL DISASTERS AND PREPAREDNESS OF DATA CENTERS

Hwaiyu Geng[1] and Masatoshi Kajimoto[2]

[1]*Amica Association, Palo Alto, CA, USA*

[2]*ISACA, Tokyo, Japan*

36.1 INTRODUCTION

"Keys to reducing the deadly effect of natural disasters are: anticipation, education and information. Unfortunately not enough priority have been given to these" [1]. The goal of this chapter is to widen the awareness, prevention, and preparedness of data center stakeholders toward natural disaster.

Human errors such as improperly executed procedures or maintenance have been shown to have brought down over 70% of all data centers. In addition to natural disasters, terrorist attack to the physical infrastructure could be devastating. Thus, it is imperative to have a plan for business continuity (BC) and disaster recovery (DR). The more we learn from past experience and augment BC and DR, the more we will be prepared for the inevitable.

36.2 DESIGN FOR BUSINESS CONTINUITY AND DISASTER RECOVERY

An ounce of prevention is better than a pound of cure. Data centers infrastructure should be built robustly with consideration of BC and DR requirements that are beyond jurisdictional building codes and standards. The International Building Code (IBC) or California Building Code generally addresses life safety of occupants. The codes provide little regard to property or functional losses and BC. To sustain data center operations after a natural or man-made disaster, design of a data center must consider system redundancy. Building structural and nonstructural components must be hardened with considerations of appropriate BC.

There are many aspects of redundancy, thus robustness, that have been addressed in this handbook. To share some highlights, the U.S. Federal Emergency Management Agency has develop the following design guidelines to harden building structure and nonstructure components for a seismic event:

- "Installing Seismic Restraints for Mechanical Equipment," FEMA, December 2002
- "Installing Seismic Restraints for Electrical Equipment," FEMA, January 2004
- "Installing Seismic Restraints for Duct and Pipe," FEMA, January 2004

Probable Maximum Loss (PML), a term that defines the value of maximum loss of property expected from a disaster, could be used to justify hardening costs. PML should also include loss of sales and market share. Costs of designing and installing seismic restraints are minimal compared to PML, in particular if it is a marginal costs for a new construction project. The American Society of Testing Materials (ASTM) has issued guidelines on how to estimate seismic loss. The document was updated in 2007 with additional terms beyond PML: Scenario Expected Loss (SEL), Scenario Upper Loss (SUL), and Probable Loss (PL).

Electrical power outage is commonly caused by natural disasters. Alternative power source such as proven fuel cell technology could be considered in lieu of power grid. For Internet Service Provider (ISP), ensure multiple Internet connections that include combination of fiber cable, T1 lines, satellite, and DSL to provide baseline service if one or two connections fail.

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

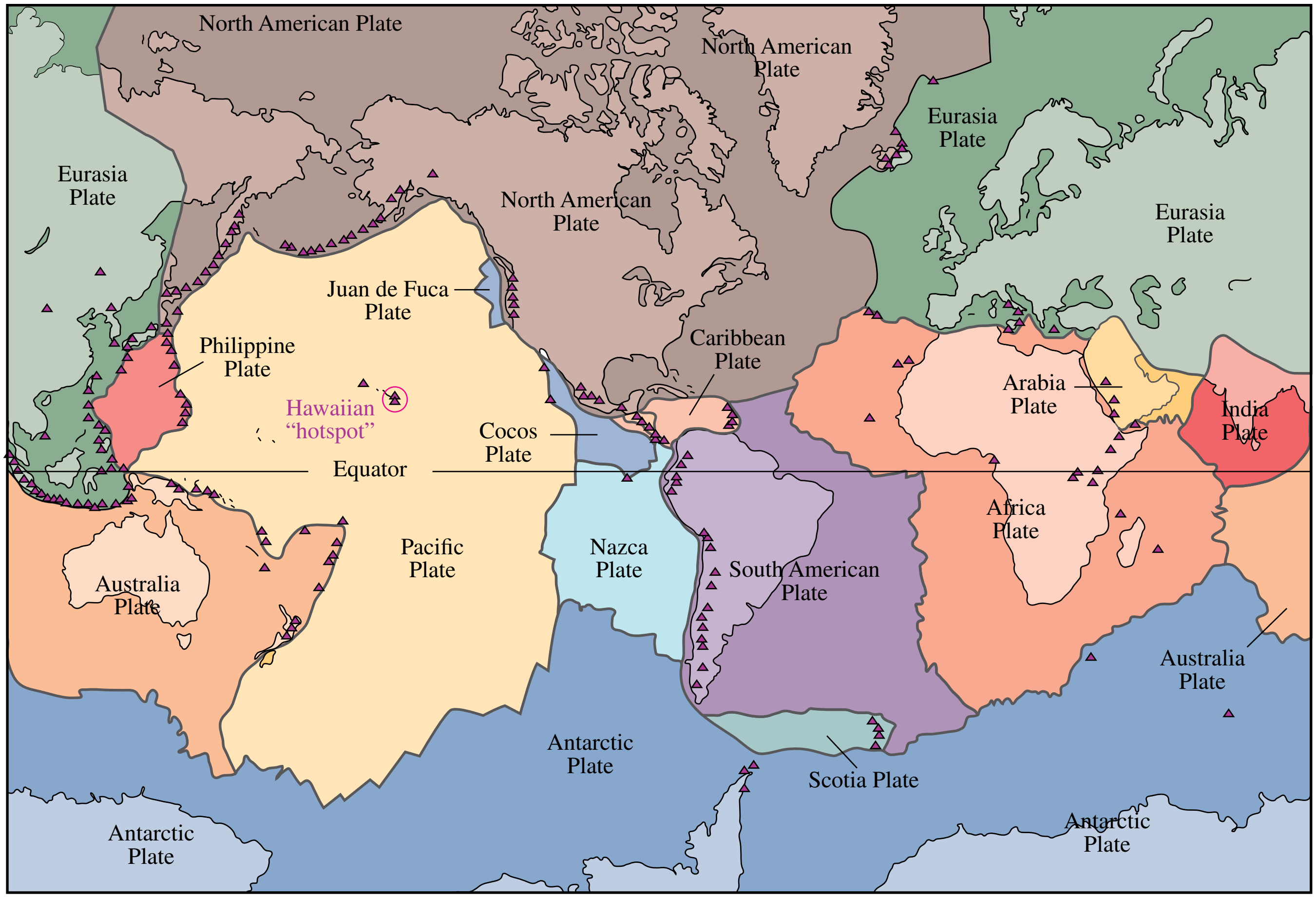


FIGURE 36.1 "Tectonic Plates and Active Volcanoes of the World." From The U.S. Geological Survey. http://pubs.usgs.gov/gip/117/gip117_ebook.pdf

36.3 NATURAL DISASTERS

There are many natural disasters including earthquakes, tsunamis, volcanic eruptions, hurricanes, tornados, wildfire, heat waves, etc. Although many natural disasters could impact operations of data centers, this chapter will concentrate in lessons learned from the Great East Japan Earthquake and the Eastern U.S. Superstorm Sandy. An earthquake or man-made disaster is unpredictable. Hurricane or storm is predictable that allows advanced time for preparedness.

Earthquakes are closely related to tsunamis and volcanic activities. Approximately 100% of Japan is in seismic zone, 60% for China, and 40% for the U.S. with "Ring of Fire" [2] from Australia, Asia, to Americas as well as other regions of the world (Fig. 36.1). An earthquake is caused by displacement of ground plates due to a "divergent, convergent, or transform fault plate boundary".[1] Earthquakes could trigger tsunamis,[2] landslides, or volcanic activities.

"A tsunami is a series of waves as a result of an undersea earthquake. A tide gauge measures the water level every minute, effectively measuring the height of each wave as it passes the gauge. After the predicted normal water level, including tides, is removed, the result shows the deviation from normal water levels due to a tsunami" (Fig. 36.2).

36.4 THE 2011 GREAT EAST JAPAN EARTHQUAKE

On March 11, 2011, a subduction zone earthquake of magnitude 9.0 occurred off the Pacific coast of Tohoku, often referred to as the Great East Japan Earthquake in Japan (Fig. 36.2). This was the largest earthquake ever to have hit Japan since instrumental recordings began in 1900 and

[1] http://pubs.usgs.gov/gip/117/gip117_ebook.pdf

[2] Progression of the Tohoku tsunami across the Pacific Ocean with tide gauge water level measurements showing deviation from predicted tide levels. (Source: http://tidesandcurrents.noaa.gov/).

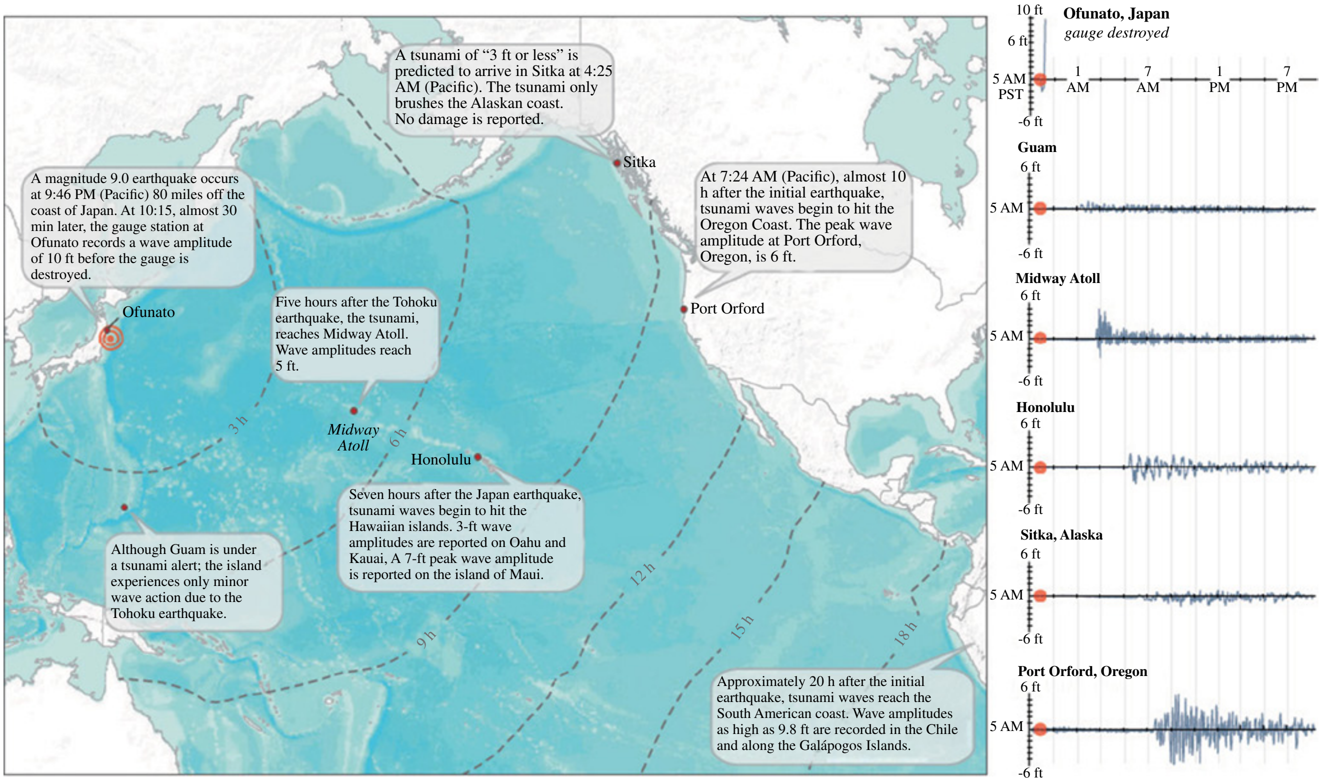


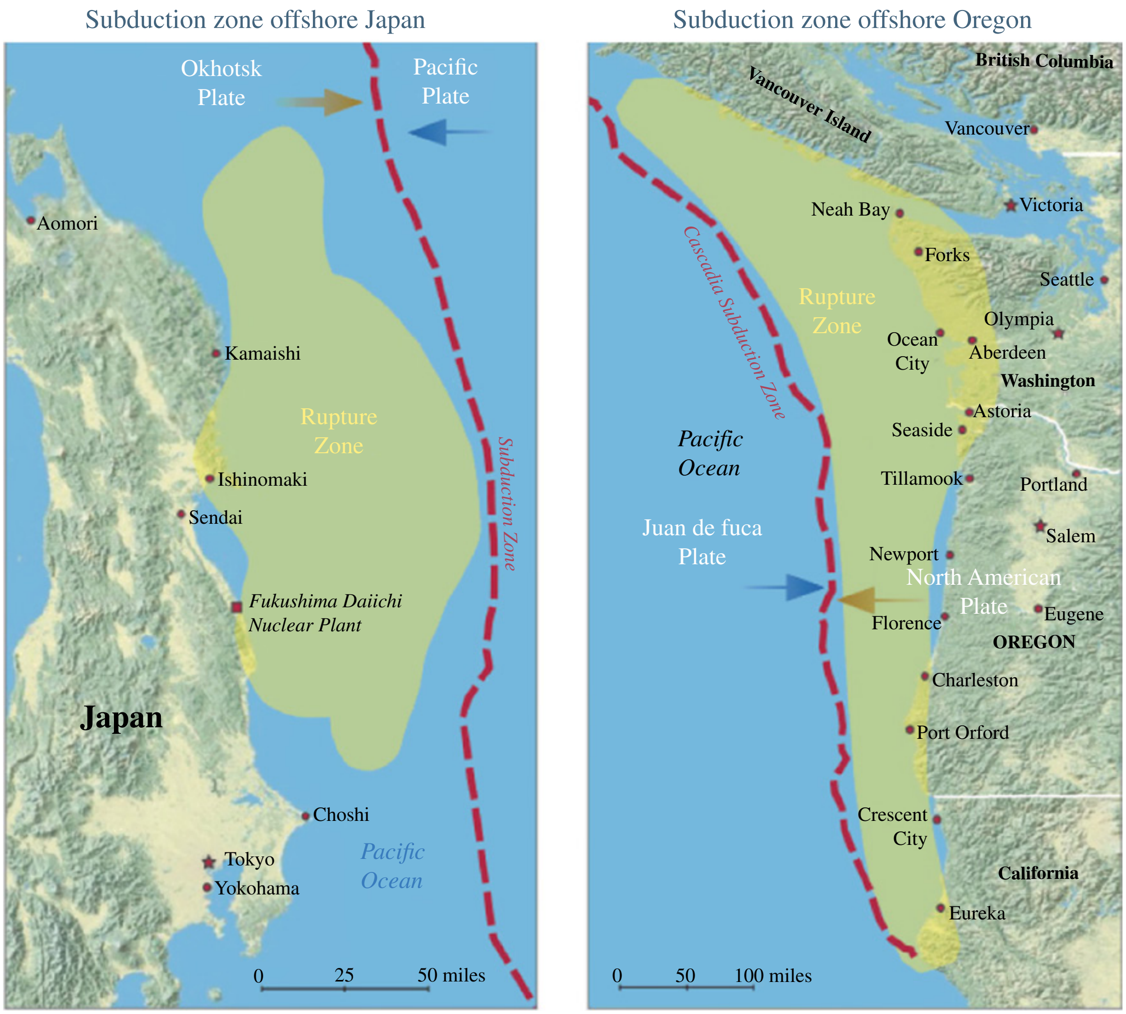
FIGURE 36.2 Progression of the Tohoku tsunami across the Pacific Ocean with tide gauge water level measurements showing deviation from predicted tide levels. From http://tidesandcurrents.noaa.gov/

the fourth largest ever recorded according to the U.S. Geological Survey. After 30 min, a devastating tsunami, at speeds up to 500 miles/h (313 km/h), struck the great east coastal line (406 miles or 650 km) of Japan and flooded 59,000 acres (24,000 ha) of agricultural land. The event left 18,884 dead, 2,636 missing, and thousands more injured. It leveled 127,290 houses. This megadisaster included an earthquake, a tsunami, a nuclear power plant shutdown, and a disruption of global supply chains. It was the costliest natural disaster in world history. The economic cost of damages to buildings, roads, ports, and others was estimated to be US$235 billion. The loss of life and properties could have been far greater, were it not for the fact that Japan had an advanced disaster risk management system, built up from lessons learned in nearly 2000 years of her history (Fig. 36.3).

36.4.1 Lessons Learned from Japan Earthquake and Tsunami [3]

First, let us look at what happened:

1. First 3 days
   - Many enterprises lost key personnel (decision makers, those who were responsible for disaster response, etc.) as a result of the tsunami.
   - The telephone/communication networks were congested as a result of the disaster.
   - The electricity supply was stopped.
   - Many organizations suffered simultaneously, meaning that not only was an enterprise suffering and/or unable to function but so were its vendors in many cases.
   - Information technology (IT) resources were heavily damaged or lost.
   - Transportation routes were heavily damaged.
   - Earthquake aftershocks continued to occur.
2. First 3 months
   - Rotational (but not well-planned) blackouts were enforced.
   - Electricity shortages continued in the disaster area.
   - Legal "Saving Electricity" measures were enforced in the Kanto area (Tokyo and surrounding prefectures).
3. Four months and beyond
   - Electricity shortage in the western side of Japan became severe. Because of the shutdown of nuclear power plants, many factories, data centers, and others from the Kanto area migrated to the western side of Japan.



(left) Area between coastal line and dot line is the exact footprint of the Tohoku rupture zone.
(right) Area between dot line and coastal line indicate a region where earthquakes can occur in the Pacific Northwest.

FIGURE 36.3 Lessons learned: Oregon's tectonic setting: a mirror image of Japan's. From Ref. [2].

Next, let us look at the impacts related to IT-related businesses:

1. First
   - Chains of command were lost. Almost nobody could decide appropriate measures for IT infrastructure recoveries.
   - Communication channels were lost. It was nearly impossible to get correct information such as "Who is still living?," "Who is in charge?," "What happened?," and "What is the current status?"
   - Stock of fuel for emergency power supply was very limited (1 or 2 days).
   - Server rooms were strictly protected by electronic security systems, so without enough electricity, these security systems became obstacles for emergency responses as the aftershocks kept coming. At some organizations, they kept the door of their server room open.
2. Second
   - Replacement facilities or equipments (servers, PCs, etc.) were not supplied quickly from the vendors because so many organizations had suffered from the disaster.
   - Backup centers also suffered in the disaster. Therefore, quick recovery was almost impossible at many organizations.
   - Because of rotational blackouts, companies could not access their servers from remote offices.
   - Data recovery was a very heavy task. If a company's backup rotation was once per week, they lost almost 1 week's worth of data. In some cases, both electronic and paper-based backups of data were lost.

- In the areas evacuated as a result of the nuclear power plant accidents, nobody could enter their own offices.
- Many IT-related devices were washed away by the tsunami. Some fell into the wrong hands.
- Fuel supply was interrupted in many areas because transportation routes were not repaired quickly.
- Many organizations moved their data centers and factories to the western side of Japan.
- The emergency power supply could not operate for long periods of time. They were designed for short-term operation.
- Monthly data processing was impossible in many organizations, resulting in much delay.

3. Third
   - At the western side of Japan, many organizations confronted electricity shortages and could not start full operations.

Now that we have a full picture of the devastation that occurred, let us look at the lessons we've learned:

1. Bad situations can continue for a long time. Quick recovery is sometimes impossible. Be prepared for this.
2. Prepare as many people as you possibly can who can respond to disasters. Having a fixed definition of roles and responsibilities may be hazardous.
3. Data encryption is indispensable.
4. A cloud computing-like environment can be very helpful in situations like this.
5. Uncertainty-based risk management is necessary.
   - In Japanese history, many huge earthquakes and tsunamis were recorded. We must study our history more carefully and note that those things can happen to us.
   - Recently on the same seismic zone (the Circum Pacific Earthquake Belt Zone), many heavy earthquakes and tsunamis have occurred. The "Sumatra Disasters," from 2004 to 2010, caused major earthquakes and tsunamis, including a magnitude 9.1 earthquake in 2004. We must learn from these disasters. And we must take account of the fact that a similar-size disaster can occur anytime on the same seismic zone.
   - Although we cannot predict exactly when, where, and how, we can prepare for the uncertainties.
6. Preparation of many risk scenarios may be useless. Too many risk response manuals will serve as a "tranquilizer" for the organization. Instead, implement a risk management framework that can serve you well in preparing and responding to a disaster.

Disasters can occur anytime and anywhere. Sit down with your colleagues and make a plan now.

36.5 THE 2012 EASTERN U.S. COAST SUPERSTORM SANDY

On October 29, 2012, Sandy, a category 2 Superstorm, made landfall in New Jersey and pummeled the U.S. east coast from Florida to Maine with effect felt across 24 eastern states. Sandy forced winds extend 175 miles out its eye, making it much larger than most storms of its type (Fig. 36.4). It drove a catastrophic "storm surge" into the New Jersey, New York, and Connecticut regions with 80 miles/h (129 km/h) sustained winds, and the heavy rain battered the densely populated states of New York and New Jersey. The storm surge (high winds pushing on the ocean's surface) reached 13.88 ft at Battery Park, New York, surpassing the old record of 10.02 ft by Hurricane Donna in 1960. It is the second largest Atlantic storm on record. The storm produced severe flooding along the Atlantic Coast, contributed to fuel shortage across the New York metropolitan area, and 2 ft of snow in areas of West Virginia, Virginia, Maryland, and North Carolina. The storm killed 117 people in the United States and 69 more in Canada and Caribbean, caused US$50 billion in damages, and left 8.5 million residents without electrical power. Sandy ranks as the second costliest tropical cyclone on record, after Hurricane Katrina of 2005. Detailed chronological events can be found at Cable News Network's (CNN) "Hurricane Sandy Fast Facts" [4]. Here are some highlights relating to data center's preparedness:

- Authorities suspended train, subway, commuter rail, and bus services. Close to 11 million commuters were without service. Airlines canceled flights.
- Three reactors experienced trips, or shutdowns, during the storm, according to a Nuclear Regulatory Commission statement.
- A total of 7.9 million businesses and households were without electric power in 15 states and the District of Columbia. (This lasted until November 7 with 600,000 people without power.)
- Areas hit by Sandy experienced gas shortage due to loss of electrical power at gas stations.
- The U.S. Energy Information Administration reported that approximately 67% of gas stations in metropolitan New York did not have gas for sale.
- New York City public schools announced via their official feed that schools would begin to open on November 5.
- A strong low-pressure system with powerful northeasterly winds coming from the ocean ahead of a storm hits the areas already damaged by Sandy.

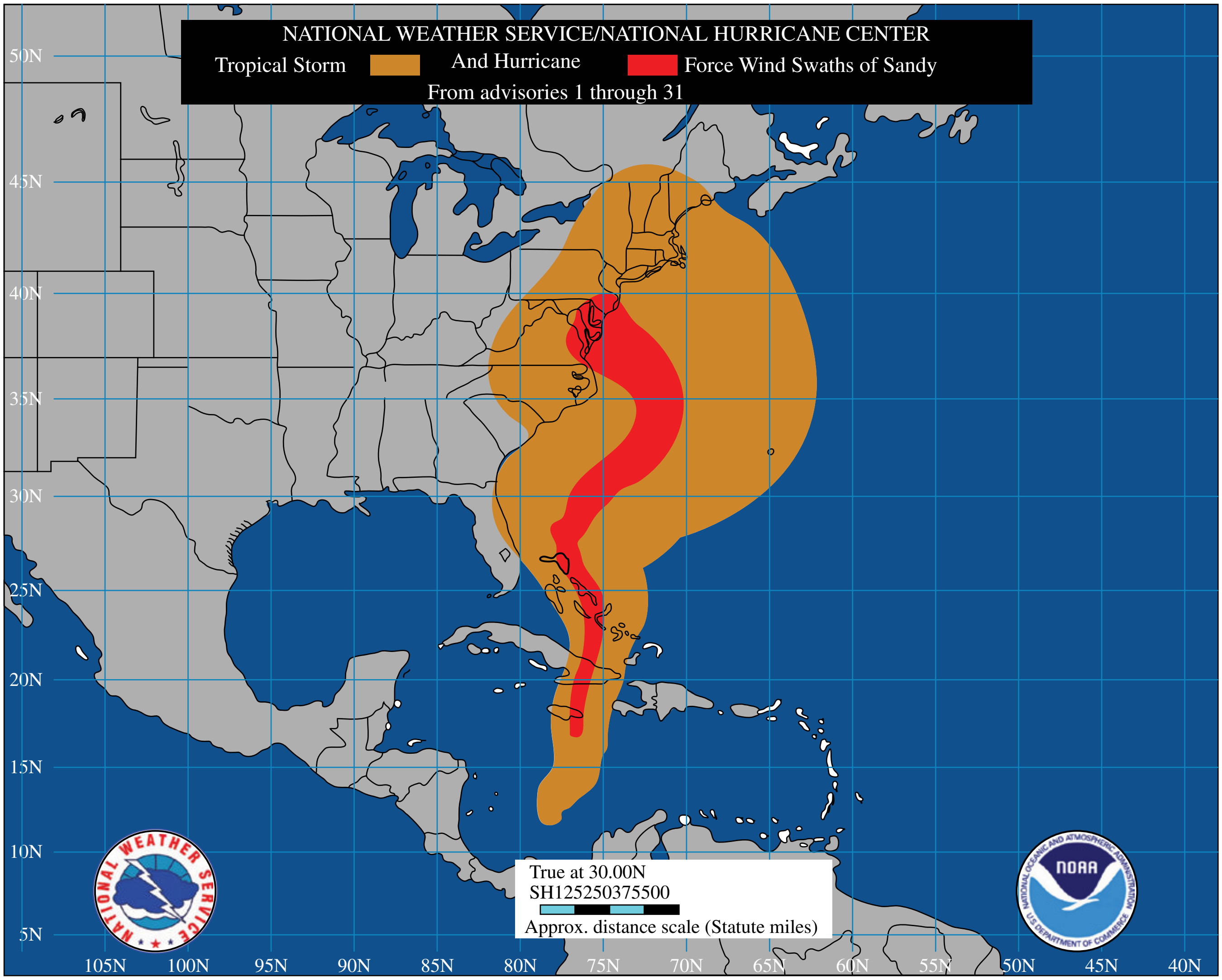


FIGURE 36.4 Sandy made landfall near Atlantic City, NJ, on October 29, 2012, as a post-tropical cyclone after traveling up the southeastern U.S. Coast as a Category 1 Hurricane. (Source: Hurricane Sandy FEMA After-Action-Report, July 1, 2013).

36.5.1 Lessons Learned from Eastern U.S. Coast Superstorm Sandy

Modern weather forecasts provided warning of Sandy well before the storm hit. The devastation in areas such as Manhattan during Sandy involved flooded water (Fig. 36.5) that was of exceptional magnitude.

Most companies had done a great job having data center out of basement or even above first floor. However, due to the need of rack level cooling, there was a tendency of designing lower level as gray space to house supporting infrastructure including electrical switchgear, mechanical equipment (chillers and pumps), and UPS. This was still a good design so long as the gray space was not in the basement and was above ground.

For New York City, fire and building codes require fuel tank to be at the bottom of a building. As a result, basements are equipped with high-capacity sump pumps to protect basement that contains diesel fuel, switchgear, and mechanical room. To avoid flood damage to fuel delivery pumps, encase pump in a waterproof box or use submersible fuel pump in the fuel tank with watertight power feed and tall vents. To hold water back, build a watertight fuel tank room so water couldn't get in. When floodwater fills up the basement, the fuel tank could lift off the foundation and get disconnected. Provision should be made for a fuel truck to physically connect the fuel oil risers in a building to run the emergency generators.

The following are some lessons learned from Sandy that could be considered in preparing your BC/DR planning.

FIGURE 36.5 A subway station in Hoboken, NJ, as well as many in Manhattan, was flooded. Courtesy AP/Port Authority of New York and New Jersey.

BC/DR Planning

- Firms that heeded a storm warning, communicated early and often, were better prepared and fared well.
- Invoke the BC/DR as soon as you hear a storm warning, and follow the procedure previously established.
- Planning period for 48, 96, or 144 h duration.
- Provide customer with ample information in time for them to make better decision on their DR plan.
- Work with your customers or business partners during regular DR drill. Work out a DR solution for the worst-case scenario.
- Review noncritical tasks and agree on when to bring them online during an emergency recover process.
- Reevaluate overall BC/DR preparedness after an event, including the performance of service providers, what worked well, what did not, etc.

Communications

- Harden communication system (equip engine generator at home) with key staff and decision makers who are critical to incident management process.
- Prearrange full-service office near key staff and decision makers who can't reach data center.
- Communicate status to decision makers, staff, customers by email; post status on a website with dashboard.
- Communication tools include email, direct phone call, website posting, and instant message.
- Use social media and GPS to communicate and locate employees.
- Federal Communications Committee reported that 25% of cell phone towers lost power, rendering mobile phone useless. Diversity of telecoms providers includes 3G/4G and satellite phone.
- Provision to charge mobile phones.
- Staff phone line 24/7.
- Redundancy of voice and data infrastructures.
- Employees who were unable or unsafe to leave home due to storm damage and lack of transportation or unwilling to leave their families could work through telecommuting systems and can do what need to be done remotely.

Emergency Power/Backup Generators

- 12 h on-site fuel storage minimum.
- Regular testing of switchover.
- Consider multiple fuel delivery contracts with diesel fuel suppliers from diverse locations.
- Check your backup generators for maintenance requirements and limitations to support 48, 96, or 144 h plans.

Logistics

- Obtain roll-up generators, extension cords, and needed spare parts delivered before storm.
- Acquire survival resources including food, drinking water, flashlights, sleeping bags, clothing, medication, and personal medical needs.
- Understand BC Plan of your suppliers and third party for supply chain disruption.
- Stock gasoline for personal vehicles.
- Prearrange to ensure purchases during the crisis with limited credit line and a delay internal procurement system.

Preventive Maintenance

- Conduct BC/DR training drills quarterly, semiannually, or annually.
- Regular BC/DR testing of systems and infrastructure with incident respond team member.
- Fully test the mission critical infrastructure that supports data center before a storm.
- Regularly operate facilities on generators for 6–8 h a time.
- Regularly conduct pull-the-plug test. The load goes on batteries, UPS, and generators run for 10 h.
- Before a disaster event, top off fuel tank and test run generators.

Human Resources

- Preplan incident support teams that can deploy staff from nonaffected datacenters at a short notice so staff can join before road or airport is closed (terrorist attack may be an exception).

Information Technology

- Move IT loads to other sites prior to an event.
- Regular backup storage and mirroring process may need to be changed due to vulnerable power loss.
- Move email system, documentation, and storage to an external cloud provider to ensure uninterrupted communication.
- Moving communication systems to Cloud-based hosting may prove very valuable.
- Prioritize critical tasks to streamline recovery.
- Oftentimes, IT securities are more lax during and after a disaster using weak links in security and business. It is important to remain vigilant and apply security best practices during BC/DR process.

There are many BC/DR practices relating to data center and IT areas that could be considered to ensure resilience. These include deploying data in redundant facilities, add-on modular data center containers, co-location, cloud, etc.

36.6 CONCLUSIONS

To streamline data center BC and DR, risk assessments and crisis management relating to natural or man-made disasters should be well thought out before an event, with a detailed mitigating plan. Good DR plans should be concise and succinct. Staff may not have time to read a long and complicated plan during an event. Investments must be considered to ensure the integrity of data center building structural and nonstructural components as well as IT equipment and infrastructure.

Disaster preparedness including incident response team, human resources, policy and procedure, communication protocol, training, and drill must be systematically reviewed and practiced at local and company-wide levels. Engage your customers or business partners with an emergency recovery and operations plan. Coordinate with various groups such as diesel oil suppliers, utility companies, and governments to ensure needed supplies and electrical power during rollbacks.

Building a culture of BC/DR preparedness for a sustainable data center is everyone's business.

Practice makes perfect. Involve personnel at data center operations, IT, and customers with planning and systematical training and drill. Pay special attention to communication and coordination.

Harden key staff's home with emergency power setup. Take care of your employees so they have heat and electricity at home.

After-the-event studies should include potential vulnerabilities of key service providers, cloud services on emails or documentation, insurance policies for loss of use, flood coverage, etc. Lessons learned after each event must be discussed to see what worked well and what didn't. Incorporate them to continuously improve BC/DR technology, policy, and procedure and reach a new level of preparedness. Do not repeat the same mistakes when repairing damages after disasters.

It is better to be prepared than to be sorry—take extra steps when there is a natural disaster warning. By doing our homework, data center downtime could be significantly reduced and recovery time shortened.

REFERENCES

[1] Natural disaster preparedness and education for sustainable development. UNESCO Bangkok; 2007.

[2] The 2011 Japan earthquake and tsunami: lessons for the Oregon Coast. The Oregon Department of Geology and Mineral Industries; Winter 2012.

[3] Kajimoto M. One year later: lessons learned from the Japanese tsunami. ISACA; March 2012.

[4] Cable news network. Available at http://www.cnn.com/2013/07/13/world/americas/hurricane-sandy-fast-facts/. Accessed on May 23, 2014.

FURTHER READING

Blake E, Kimberlain T, Berg R, Cangialosi J, Beven J. Tropical cyclone report Hurricane Sandy. Washington, DC: National Hurricane Center, National Oceanic and Atmosphere Administration; February 12, 2013.

Bosco ML. Hurricane Sandy prompts new way of thinking about data center resiliency. FacilitiesNet; May 2013. Available at http://www.facilitiesnet.com/datacenters/article/Hurricane-Sandy-Prompts-New-Way-Of-Thinking-About-Data-Center-Resiliency--14015. Accessed on May 22, 2014.

Brotherton H. Data center recovery best practices: before, during, and after disaster recovery execution [Master thesis]. West Lafayette: Purdue University; 2011. Available at http://works.bepress.com/heatherbrotherton/. Accessed on May 22, 2014.

Disaster recovery: 10 lessons from Hurricane Sandy. Deloitte; November 2012. Available at http://deloitte.wsj.com/cio/2012/11/29/disaster-recovery-planning-10-lessons-learned-from-hurricane-sandy/. Accessed on May 22, 2014.

IBM. Faster disaster recovery in IBM business process manager. August 2013. Available at http://www.ibm.com/developerworks/bpm/bpmjournal/1308_zhang/1308_zhang-pdf. Accessed on May 22, 2014.

Industry impact and lessons learned from Hurricane Sandy. Tellefsen and Company; January 2013. Available at http://www.

futuresindustry.org/downloads/Industry-Impact-and-Lessons-Learned-From-Hurricane-Sandy_Summary-Report.pdf. Accessed on May 22, 2014.

Japan Metrological Agency. Lessons learned from the tsunami disaster caused by the 2011 Great East Japan Earthquake and Improvements in JMA's Tsunami Warning System. October 2013. Available at http://www.seisvol.kishou.go.jp/eq/eng/tsunami/LessonsLearned_Improvements_brochure.pdf. Accessed on May 22, 2014.

Kidman A. Top 10 data center management lessons from Hurricane Sandy. Lifehacker; April 2013. Available at http://www.lifehacker.com.au/2013/04/top-10-data-centre-management-lessons-from-hurricane-sandy/. Accessed on May 22, 2014.

Learning from the Great East Japan Earthquake and Tsunami Policy Perspectives—Summary Statement. Japan-UNESCO-UNU Symposium; February 2012. Available at http://ioc-tsunami.org/index.php?option=com_oe&task=viewEventRecord&eventID=1035. Accessed on May 22, 2014.

Ready, Prepare, Plan, Stay Informed. FEMA.gov. Available at http://www.ready.gov/business/implementation/continuity. Accessed on May 22, 2014.

Rimler B. Lessons learned from Hurricane Sandy. Schneider Electric Data Center; November 29, 2012. Available at http://blog.schneider-electric.com/datacenter/2012/11/29/lessons-learned-from-hurricane-sandy/. Accessed May 22, 2014; the U.S. Geological Survey; 2010. Available at http://pubs.usgs.gove/gip/117/. Accessed on May 22, 2014.

Suppasri A, Shuto N, Imamura F, Koshimura S, Mas E, Yalciner A. Lessons Learned from the 2011 Great East Japan tsunami: performance of tsunami countermeasure, coastal building, and tsunami evacuation in Japan. Pure Appl Geophys 2013;170(6–8): 993–1018.

Sverdlik Y. Japan data centers face rolling blackouts. London, UK: Datacenter Dynamics; March 15, 2011.

Tilling R, Heliker C, Swanson D. Eruptions of Hawaiian Volcanoes—past, present, and future. Reston: U.S. Geological Survey; 2010. Available at http://pubs.usgs.gov/gip/117/gip117_ebook.pdf. Accessed on May 22, 2014.

Tuner WP, Seidman D. Sandy. 2013 Symposium Uptime Institute; May 2013. Available at http://symposium.uptimeinstitute.com/images/stories/Symposium_2013/2013_Slides/Day3/0516_sandylessonslearned.pdf. Accessed on May 22, 2014.

World Bank. The Great East Japan earthquake learning from mega-disasters. Washington, DC: World Bank; 2012.

Yamanaka A, Kishimoto Z. The realities of disaster recovery: how the Japan Data Center Council is successfully operating in the aftermath of the earthquake. Shinjuku, Tokyo, Japan: JDCC, Palo Alto, CA; AltaTerra Research; June 30, 2011.

INDEX

Data Center Handbook, First Edition. Hwaiyu Geng.

Companion website: http://www.wiley.com/go/datacenterhandbook

WILEY END USER LICENSE AGREEMENT