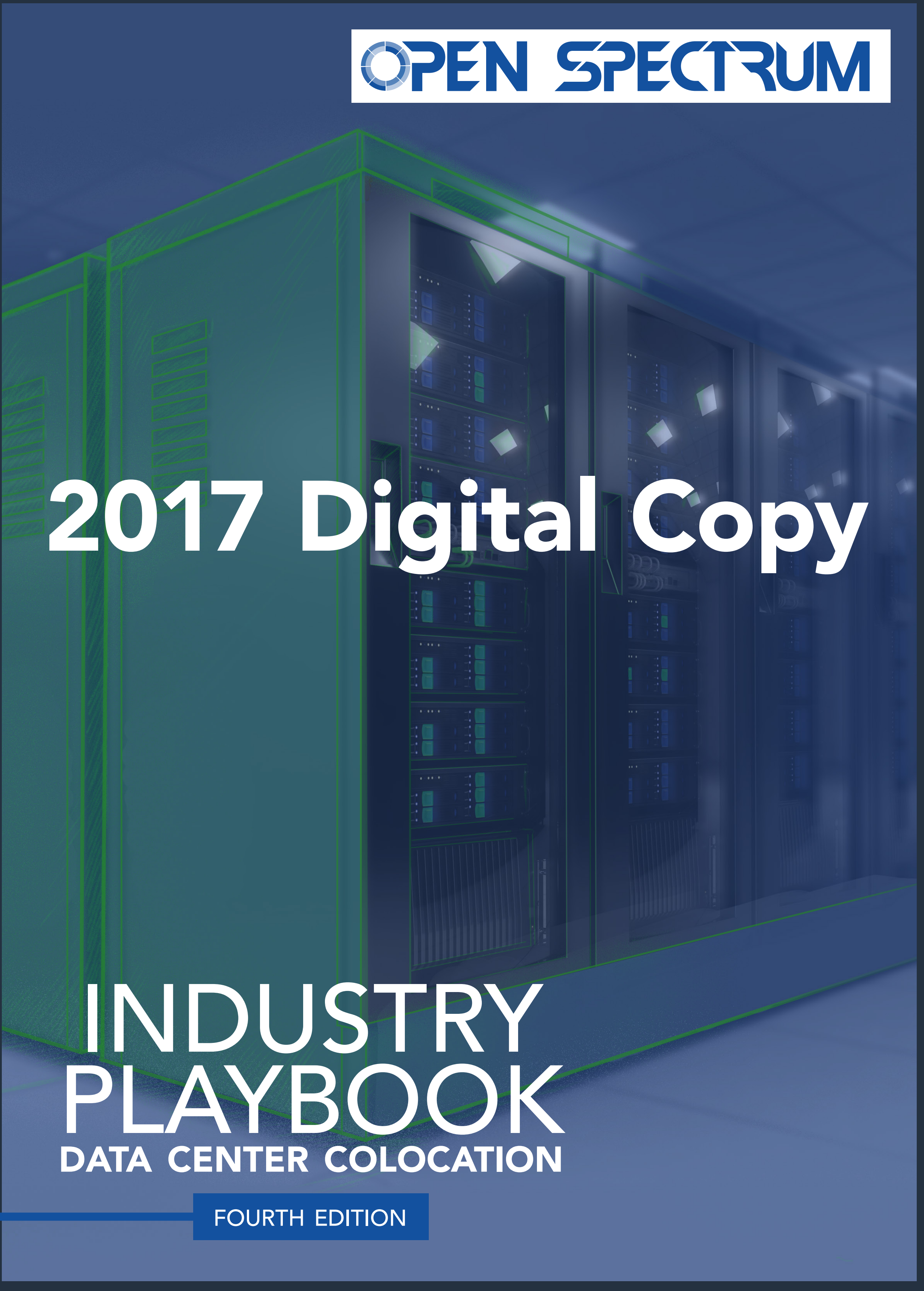


2017 Digital Copy

INDUSTRY PLAYBOOK

DATA CENTER COLOCATION

FOURTH EDITION

INDUSTRY PLAYBOOK

DATA CENTER COLOCATION

FOURTH EDITION

Sean Patrick Tario

OPEN SPECTRUM

RALEIGH, NORTH CAROLINA

Data Center Colocation Industry Playbook

Everything You Need to Know When Buying and Selling Data Center Colocation Services

Fourth Edition
Sean Patrick Tario

Data Center Colocation Industry Playbook: 4th Edition
Open Spectrum

Open Spectrum
308 S. Blount Street, Suite 1702
Raleigh, NC 27601

Author: Sean Patrick Tario
Content Contribution and Editors:
Micah Berninghausen, Data Center Industry Sales Engineer
Dan Vazquez, President, Vazquez Data Center Services
Mark C. Del Bianco, Data Center Industry Attorney and Counselor

For information on rights, translation, and bulk sales: *contact@openspectruminc.com*

ISBN: 153759107X
ISBN 13: 9781537591070

This book is dedicated to all those with an unquenchable thirst for knowledge and an understanding of the systems and industries that make the world work around us.
It is also dedicated to the servant leaders of the world who patiently and humbly guide us along our path.

ENDORSEMENTS

Having bought and sold many megawatts of power within data centers around the world, I can assure you the Open Spectrum bootcamps should be the de facto training for anyone who wishes to take themselves seriously in the industry.

Richard Donaldson, *Director Global Managed Services*
eBay Inc.

C7 recommends the Open Spectrum data center training for sales, technical, and customer support personnel. Truly top-notch training that is relevant to all facets of the data center. The training materials provided during the presentation are kept on everyone's desk as a go-to reference.

Dave Jenkins, *VP of Operations*
C7 Data Centers

Open Spectrum is on a mission to equip and educate, for the greater good of the entire colocation community and its constituents. The time and money spent on the two-day training was well worth the investment. Not only was it beneficial having common terms, calculations, and industry standards clarified, the ability to compare notes among industry colleagues was a pleasant surprise. I will definitely be encouraging staff members as well as members of the data center community to participate when given the opportunity.

Mike Rushing, *Business Development*
Westin Building Exchange

The volume of knowledge and experience brought to the table during these events is impressive, to say the least. Even if you only retain a fraction of the material covered, you cannot help but walk away a more sophisticated and savvy seller…or buyer, for that matter.

Farrah Kashef, *VP of Business Development*
Unitas Global

Sean and his team have obviously "been there and done that" in all facets of the IT and data center space and have a true talent for breaking down the overwhelming complexities of the landscape in a way for anyone to comprehend. It's truly a wealth of knowledge. My everyday focus is on equipment finance, and this bootcamp—as well as the book—has no doubt helped me to take a more well-rounded approach to servicing our client requirements and understanding what their true needs are.

Brian Fleming, *VP Equipment Finance*
Fort Capital Resources, LLC

After providing IT hardware and integration solutions to countless customers for over fifteen years, I thought I was already well informed on what goes on in the data center. This bootcamp increased my hands-on exposure to the language, variables, and ultimately the ecosystem that connects to our solutions. It has increased my usefulness to my customer, which is what it's all about.

Jeff Fasen, *VP Tech Solutions*
Redapt

This was by far the best training I have ever been through. I learned a ton of information and sincerely appreciated the hands-on support and overall design of the bootcamp. Having a third-party expert provide an overall view of the industry is far more valuable than learning through the eyes of any one company perspective.

David Hotaki, *Account Executive*
SunGard Availability Services, LP

Even with over a decade of experience in the telecommunications and data center industries, I found the Open Spectrum bootcamp to be a great learning experience for both novices and professionals in the industry. Highly recommend everyone—regardless of experience—attend. You will walk away learning something new and valuable, guaranteed.

Stuart Hanson, *VP of Leasing*
Sabey Data Centers

I came into this bootcamp not really knowing what to expect, and was extremely impressed by the relevance of the materials presented and the engaging format. Without question, this bootcamp was well worth my time and I am thankful to my team for giving me the opportunity to attend!

Jason Liebowitz, *Service Delivery Manager*
Colovore

TABLE OF CONTENTS

INTRODUCTION

PREFACE

At the cross roads of high tech, fluid dynamics, real estate and our energy grid sits the data center; an asset and industry, mostly hidden or misunderstood, that is at the core of the digital world we live in today. Data centers are the ever evolving and growing metropolitan city of this new age, designed to house the data that drives our economy forward.

If you've ever walked into one of these purpose built 100,000+ square foot facilities, with the security gates, iris scanners, biometric readers, badges, guards and heavy machinery all around you, it's hard not to have your draw drop and an overwhelming feeling that you're standing in something important and fascinating.

Some in the industry say that if you've walked through 2 or 3 of these facilities you've seen them all. My experience however has proven that this is far from reality. Each property is unique, and the use cases for a data center continues to grow in proportion to the volume of applications built and data sets aggregated every second. Data centers underwater, on ships, in space, from 1,000 sq ft to 1,000,000 sq ft and beyond, I love them all!

This book, as well as our live and online training, are all specifically designed to immerse you in the ecosystem of the data center. My hope is that you develop a unique and strategic advantage, wherever you may be in your career, through developing a better understanding of the ecosystem and language of the data center industry. The information provided, case studies reviewed and experiences shared within all of our materials, including this book, are designed to bring you up to speed as quickly as possible on the relevant information needed to buy and sell data center colocation services.

I guarantee that once you have mastered the materials covered, you will be more informed than 95% of those both buying and selling within the industry today. Make no mistake, to master these materials will require that you review and continue to review the material over the weeks, months and years to come. It's like learning any new language - your objective should be to make the terms, equations, and concepts second nature as you become fluent in the language of data center colocation.

To this end, I encourage you to visit www.openspectruminc.com and explore the other training materials our team continuously produces and updates. I also strongly recommend you consider attending one of our intensive two-day, live "Bootcamps" hosted in various cities around the world, where these materials are presented by either myself or one of our other experienced trainers.

Last but not least, thank you for taking the time to pick up and read this book! If you currently work in or around the data center industry, I hope this serves as an invaluable tool for you as you advance in your career. Please feel free to reach out to me directly at any time with any questions, comments or suggestions on new content that should be included in the next edition of the book, which is sure to come out shortly!

Key Terms and Abbreviations	
1,000 Kilowatts	1 Megawatt (MW)
1,000 Watts	1 Kilowatt (kW)
A/B	Primary/Redundant
Actual kW	kVA x Power Factor
ATS	Automatic Transfer Switch
Backhaul	Carrier needs to leverage another provider's fiber and network infrastructure to reach a client in a building.
Carrier Hotel	A data center with a significant number of carriers delivering services out of the facility. Likely one or more peering exchanges within the facility. Sometimes referred to as an Internet Exchange (IX).
CPI	Consumer Price Index
CRAC	Computer Room Air Conditioner
CRAH	Computer Room Air Handler
Dark Fiber	In standard fiber services, the carrier lights the fiber and defines the service the fiber provides. With dark fiber, the customer lights the fiber and therefore defines the service the fiber provides (CWDM, DWDM, 10gb, fiber channel, etc.).
DIA	Designated Interconnect Area (most of the time referred to as a MMR)
IaaS	Infrastructure as a Service
IX	Internet Exchange
kVA	kilo Volt Ampere; 1 kVA is equal to 1 kW when DC current is provided, as the current and voltage are not out of phase, and thus the Power Factor is 1.
kW	kilowatt
MMR	Meet Me Room
MPOE	Minimum Point of Entry; description for where the physical fiber strands from a telecom provider land in a building.
MRC	Monthly Recurring Cost
MRP	Managed Recovery Program
MRR	Monthly Recurring Revenue

Key Terms and Abbreviations	
MSA	Master Services Agreement
MW	Megawatt
NAP	Network Access Point
NRC	Non-Recurring Cost
NRR	Non-Recurring Revenue
On Net	Carrier has their own network infrastructure in the building to connect directly with clients.
Paid Peering	Same as peering, but with some form of compensation or settlement fee, generally at a reduced rate.
PDU	Power Distribution Unit
Peering	Voluntary interconnection of administratively separate Internet networks for the purpose of exchanging traffic between the customers of each network at no cost.
PoP	Point of Presence
Power Density	How much critical load a facility is designed to support described as a per cabinet or per square foot metric.
Power Factor	Measurement of efficiency for a particular device (0 to 1 range)
PUE	Power Usage Effectiveness
RPO	Recovery Point Objective
RPP	Remote Power Panel
RTO	Recovery Time Objective
SLA	Service Level Agreement
TCO	Total Cost of Ownership
Triple Net	Real Estate term referring to the inclusion of taxes, maintenance fees and insurance within a commercial lease.
UPS	Uninterruptible Power Supply

DATA CENTER INDUSTRY

CHAPTER ONE: DATA CENTER COMPANY TYPES

After studying this chapter, you will be able to:

- Define the principal legal structures of data center owner-operators.
- Interpret some of the ways that these legal structures influence business decisions, especially regarding service to clients and contract negotiations.
- Communicate to an infrastructure buyer the pros and cons of working with any given type of provider.

» OWNER TYPE: REAL ESTATE INVESTMENT TRUST (REIT)

These companies can either be public or private and are structured to act primarily as an investment agent specializing in real estate. It is interesting to note that despite the increased regulatory compliance requirements and the fees associated with managing a public company as compared to a privately held entity, almost all large national and global data center REITs are public today. From a purely financial and tax perspective, however, it is undeniable that there are a number of benefits for a company to structure itself as a REIT and obtain that designation from the SEC.

REITs must adhere to many regulations. They must be managed by a board of directors or trustees. They must not be a financial institution or an insurance company and must be jointly owned by one hundred persons or more. REITs must derive 95 percent or more of their income from dividends, interest, and property income. They must also pay dividends of at least 90 percent of income back to the investors.

A REIT is exempt from federal corporate tax, but it must have at least 75 percent of its total assets invested in real estate. Finally, a REIT must derive at least 75 percent of its gross income from rents or mortgage interest.

Because REITs are not allowed to have more than 25 percent of their gross income generated from services besides rental and mortgage interests, few data center REITs offer hosting, network, or any services outside the standard wholesale lease agreement. Providers who do offer such additional services almost always do so through a wholly owned subsidiary or “partner” company. Companies structured as a REIT also tend to be very conservative and patient about acquiring new properties for growth. As alluded to above, their contracts must also be structured in the form of a lease agreement between the client and the REIT. In contrast, non-REITs offer their clients agreements that tend to be far simpler and shorter.

Examples of companies structured as data center REITs:

- Digital Realty
- CoreSite

- DuPont Fabros
- CyrusOne
- QTS Realty Trust

» OWNER TYPE: PRIVATE EQUITY - BACKED COMPANIES

Private equity (PE) firms have a substantial investment or controlling stake in data center companies. You can therefore assume that one of the primary objectives of a company backed by private equity will be to either sell or to go public within a three- to seven-year period. Data center companies are extremely capital intensive: PE firms generally invest tens to hundreds of millions of dollars into a company over a number of months and years, with a very clear disposition plan and strategy. This disposition plan will very clearly outline the executive team's commitment to forecast and expected revenue, cost, EBITDA, and head count targets over time. The plan will also contain a very specific exit strategy, as outlined above, with an expected return of two to ten times the investment(s) made.

Given the size of the investments and the goal of achieving substantial returns, you can guarantee massive changes will occur within these companies once private equity gets involved. Having watched the industry now for over a decade, it's somewhat humorous to see how the principals of these companies follow the same pattern: they immediately broadcast a message both internally and externally that "nothing is changing, the new investors are looking for a long-term hold," only to turn around and institute relatively immediate and massive restructuring, head-count doubling, and an inevitable public offering or acquisition within three to five years from the day the initial investment was made. Companies receiving PE capital injections will almost always emphasize their sales and marketing divisions as well.

It is important to note that drastic changes may in fact be a good thing—for both the employees and clients. New capital injections usually mean much-needed infrastructure upgrades, projects, and hires that have been "on hold" can be approved. With a clear exit strategy on the horizon, employees will likely be offered some form of stock option plan to incentivize existing talent to remain at the company and to attract new talent.

Examples of owners who are backed by private equity firms:

- Peak 10, backed by GI Partners and Seaport Capital
- Vantage, backed by Silver Lake Partners
- AIS and Atmosera, backed by Seaport Capital
- Data Pipe, backed by ABRY Partners

» OWNER TYPE: PUBLICLY OWNED COMPANIES

Data center providers now listed on public stock exchanges were almost all once backed by private equity money. When they become public, they will structure themselves as a REIT, or will soon begin the process of receiving approvals to structure themselves as a REIT. A public company must adhere to strict federal accounting rules and regulations and must disclose detailed financials and large contracts to the public. Though burdensome, this process can actually prove to be extremely valuable when negotiating with large-enterprise clients, as most CXOs and especially CFOs like to know the provider they will be working with is financially sound and will be around for decades to come.

Public companies must also manage their relationships with analysts very closely, as any real or perceived changes in management or services offered must be communicated to the investment community in such a way as to not cause a sudden investor scare or stock sell-off.

This is important to note: due to the complex nature of the IT services offered within a data center, not all analysts are well versed in how the ever-changing technological landscape of delivered services can impact a data center company. For example, many analysts and professionals in the industry showed concern that "cloud computing" was going to degrade the need for data-center services. They predicted a decline in the demand for data-center services as companies moved their data "into the cloud." This conclusion stands in stark contrast to the reality that "cloud computing" simply changes who manages the data and changes where the data is physically stored. In other words, "cloud computing" in no way changed the physical requirements needed for data storage. If anything, "cloud computing" just made it easier for small businesses to move their data from a server closet in their office into a secure data center run by a "cloud computing" service provider who maintained the servers for them. As a direct result of the knowledge gap between the reality of the cloud and what people believe, public companies tend to move very slowly and act in a far more conservative manner than private companies from a pricing, hiring, and site-acquisition perspective.

Examples of owners who are public companies:

- Digital Realty
- CoreSite
- DuPont Fabros
- CyrusOne
- QTS Realty Trust

» OWNER TYPE: TELECOMMUNICATIONS PROVIDERS

This group consists of a few telecommunications providers who are primarily in the business of owning the physical fiber in the ground and selling data and voice connectivity over that fiber. As a byproduct of these activities, many have over the past twenty-plus years built their own data centers in most primary and secondary markets across the country, to house their switch gear.

Most major telecom companies have been in the industry for a very long time, and the power and space requirements for even the most robust network switch and routing infrastructure have significantly decreased over time. This has left almost all providers with excess power and available space within the facilities they originally built to house only their own network equipment.

Consequently, many providers have opted to sell that available power and/or space; the footprint they needed for their own facilities simply shrank. However, in almost every case these legacy facilities have: low power density (1–2 kW per cabinet maximum allowable power usable), high PUE (1.8 to as high as 3), limited fiber coming into the building and on-net network options, limited security, limited remote hands-and-eyes services, and no portal online to manage tickets or accessibility to the facility.

Examples of telecommunications providers involved in delivering data center services:

- XO
- Level3
- Windstream
- CenturyLink
- Zayo

» OWNER TYPE: PRIVATE COMMERCIAL REAL ESTATE COMPANIES

Private data center companies are primarily run by commercial real estate executives, property owners, or general contractors who somehow managed to find themselves involved with a data center transaction at some point. Usually, upon realizing how lucrative and large these transactions can be, or upon realizing one of the properties they already own is being developed as a major Internet exchange by a tenant, a property owner decides to focus on the industry, or on further developing the available data center options at one of their properties.

The technical acumen across the company's board, the depth and breadth of services offered, infrastructure redundancy, and overall viability of the facility will vary greatly in this category. New investors have flooded the data center industry over the past decade.

Several firms have established themselves as relatively competent providers in the market (such as those listed below), primarily due to the experience and competence of the building facility managers. Though these companies approach deals as real estate lease transactions, because they are private companies they tend to be extremely flexible to work with and it's usually possible to access the owners of the company during negotiations early on in the process.

There are, unfortunately, more and more firms entering the data center business who know very little about the industry at large, how it has evolved over time, or how to deliver the retail colocation services most clients have come to expect from a data center provider. As it is almost always the case that managing a data center property is far more lucrative than managing a commercial office property, the primary desire to participate in the industry from this subset of data center owners is strictly profit-driven. Thus, you should be extremely cautious when working with newcomers to the industry who only have a background in real estate.

Examples of private commercial real estate companies:

- Sabey Data Centers
- Sentinel Data Centers
- 60 Hudson Street Owner, LLC (owner of 60 Hudson in NYC)
- The Benaroya Company

» OWNER TYPE: PRIVATE IT COMPANIES

Companies that fall into this category were originally started by IT engineers or executives who came to the conclusion either that it would either be cheaper for them to build and manage their own data centers, or that they could simply do a better job than the existing providers in the market at the time.

Originally these firms established themselves as competent providers in the market within the engineering circles of a local community, primarily due to the reality that the owner(s) of the company were industry peers and spoke the same language as their counterparts, who were acting as buyers. They tend to deliver a wider variety of solutions besides just power and space, giving clients the option to have a single provider manage their network, DR, and related systems. As private companies, they tend to also be extremely flexible to work with, offer service contracts instead of just lease agreements, and will provide access to the owners of the company during negotiations early on in the process.

Very few of these companies still exist today, however, due to the consolidation that has occurred in the industry; the rise of Softlayer, Rackspace, AWS, and Azure; and the reality that very few of these types of owners have the sales, marketing, or financial experience to scale their companies beyond a local client base.

Examples of private IT companies:

- Phoenix NAP (Phoenix, AZ)
- Alchemy Communications, Inc. (Los Angeles, CA)
- Hurricane Electric (Fremont, CA)

Chapter Summary

- Real estate investment trusts (REITs): Either public or private, specialize in real estate. Most REITs sell wholesale services due to income regulations.
- Private-equity-backed companies: Will often refresh their infrastructure and beef up their sales and marketing teams in the short term. May also experience massive restructurings.
- Publicly owned companies: Greater perceived financial stability and corporate staying power than private companies. Often structured as REITs, or pursue such structuring.
- Telecommunications providers: Often sell excess power and space at switching facilities. Facilities may have a low power density and a high PUE.
- Private commercial real estate companies (CREs): Often extremely flexible in negotiations. Some may have little industry-specific technical knowledge.
- Private IT companies: Typically offer a wide range of managed services, tend to be extremely flexible in negotiations.

Managed commercial multitenant retail facility with raised floor, image courtesy of Cologix

CHAPTER TWO: DATA CENTER SOLUTION TYPES

After studying this chapter, you will be able to:

- Define different solutions offered by data center operators.
- Quantify the value different solutions provide different clients, and determine why some might be more desirable than others at any given time.
- Communicate effectively to your clients which solutions would best meet their needs.

» BUILD-TO-SUIT DATA CENTERS

Some construction companies will serve as the general contractor, architect, and in some cases the landlord for a company wishing to build a single-tenant or multitenant facility. Such companies are usually large, commercial, general contractors with a "critical facilities" division that focuses exclusively on designing and building data centers. The electrical and cooling components included in the data center are therefore built to the client's specifications, or are based on the general contractor's recommendations.

Companies who build to suit are large corporations with massive data storage needs—Facebook, Yahoo, Microsoft, Google—or the managed wholesale or commercial data center owner/operators themselves, such as QTS, Internap, and Latisys.

Examples of build-to-suit providers:

- Swinerton Builders
- DPR Construction
- Turner Construction

» SHELL BUILDING DATA CENTERS

A company is generally looking at this as a pure real estate investment play. It involves selling or leasing a previously occupied data center or commercial/industrial building that will need upgrading to become relevant and usable. Whoever purchases or leases the property will be responsible for maintaining and managing the cooling, electrical, and related infrastructure required to run the facility. Sometimes a shell can be simply a server room within a larger commercial office building.

Clients looking for shell buildings are often corporations with fairly large data storage needs, intellectual capital in house, and enough cash in the bank to make the investments needed to retrofit, operate, and maintain a data center. More often than not, the company already employs an

IT operations manager who has experience with managing a data center and who pushes the CXO to go this route instead of working with a managed commercial or wholesale data center owner/operator.

Nearly anyone selling a commercial warehouse, a former microchip manufacturing plant, or a related IT manufacturing facility with the ability to pull significant power and fiber into the building can be considered as an example of someone offering a shell. Commercial real-estate brokerage firms such as Colliers, CBRE, and Jones Lang LaSalle now have “critical infrastructure” teams dedicated entirely to selling and sourcing these specific types of buildings for their clients.

» PREFABRICATED DATA CENTERS

Sometimes a company delivers and assembles a partially predefined block of electrical and cooling infrastructure that can be stacked next to or on top of another. Most prefabrication companies also provide built-in building management systems to allow clients to manage an entire facility themselves. However, some offer to manage the environment after it is built for an additional expense. Clients for prefab centers are similar to those who would be interested in the build-to-suit or shell options, have predefined or predictable compute metrics, and are growing to or beyond 1MW within a twelve-to-twenty-four-month period.

Examples of shell providers:

- Compass Datacenters
- HPE
- BladeRoom Group

» MODULAR DATA CENTERS

Modular data centers often consist of a delivered freight container that includes all required electrical, cooling, and network infrastructure already installed. A client simply needs to plug in power and water sources, and to link it to some form of Internet connectivity for the container to go live. These containers can sit within a warehouse, on the roof of a building, or in a garage or parking lot—they simply stand on their own.

Clients for modular centers are those with fixed infrastructure and compute requirements, but who have unique placement requirements and/or a limited time frame to set up compute resources in locations without similar infrastructure already in place.

Examples of modular providers:

- IO Data Centers
- SGI ICE

» MANAGED WHOLESALE DATA CENTERS

Companies who provide these are in the business of building, managing, and leasing large data center suites or entire buildings. The data facility is prebuilt and likely has predefined cooling, electrical, security, and network components in place—though the option may be provided to have N, N+1, 2N, or 2N+1 power delivered.

Clients for managed centers are those who do not want to be in the business of building or managing their own data centers, who are comfortable with structuring long-term lease contracts, and who will require 750 kW of power or more within a twelve-to-twenty-four month time frame from deployment. Due to the size of their environments, these clients generally prefer to be more hands-on within the facility, and will want to have some say in how the physical cabinets, cages, electrical distribution, network topology, and power density are deployed and managed within the facility. They also generally have at least one person coming and going from the data center daily to swap disk drives or to perform related tasks within the caged environment provided. Contracts are generally five-to-ten-year lease agreements.

Examples of managed wholesale providers:

- Digital Realty
- DuPont Fabros
- T5 Data Centers

» MANAGED COMMERCIAL RETAIL DATA CENTER

These companies are in the business of building, managing, and selling small chunks of power and space to numerous tenants within one data center. The client leases space in a facility that is already built and that has predefined power density, cooling, electrical, security, and network components in place. Little flexibility is offered besides customizing the circuit types and the physical environment segregation, such as private cabinets, cages, or suites. Providers who offer this type of service are either leasing large chunks of space and power from a landlord such as Digital Realty, or the provider itself is the landlord and is capable of offering a managed wholesale product within the building as well.

Clients who seek this service are those who do not want to be in the business of building or managing their own data centers, lack the domain expertise or infrastructure to manage the environment within their offices, and who have relatively minimal power requirements. Most clients require a lot of hand-holding through the decision-making process, the setup, and through ongoing management of the environment as well. Clients tend to rely heavily on the provider for "remote hands and eyes" and advice for their initial network design and power distribution. Contracts are generally one-to-three-year service contracts, or, in some cases, lease agreements that are designed to look more like simpler service contracts.

Examples of managed commercial retail providers:

- Equinix
- Telx
- Cologix

> SECTION NOTE: *It is increasingly the case that any one company can provide two or more of the types of solutions presented above. For example, some traditional managed commercial service providers are now offering managed wholesale and build-to-suit solutions for customers, while some managed wholesale service providers are now offering managed commercial solutions. The lines are blurring!*

Chapter Summary

- Built-to-suit data centers: Owned by large corporations with large data storage needs. Often built by a commercial contractor.
- Shell buildings: Older data centers or industrial buildings sold with intent to refurbish.
- Prefabricated, predefined data centers: Can be built out in stages through the use of blocks of infrastructure. Clients generally have predefined and predictable compute metrics and expect to grow beyond 1MW in power needs within one to two years.
- Modular data centers: Freight containers that house all the infrastructure needs of a data center. The containers can sit just about anywhere, and are useful for clients with fixed compute requirements in unusual locations.
- Managed wholesale companies: Owners lease large suites or buildings. They don't manage the client's daily needs. Clients typically have large infrastructure needs but don't want to build their own data center.
- Managed commercial retail companies: Owners build out and manage small sections of power and space for multiple small tenants. Many of these companies lease space from larger wholesale providers. Clients rely heavily on the retail provider for management and advice.

CHAPTER THREE: BILLING AND PRICING MODELS

After studying this chapter, you will be able to:

- Determine usable wattage for single and three-phase circuits.
- Calculate MRC for the industry's primary billing models.
- Evaluate basic costs.

» ESSENTIAL DATA CENTER COLOCATION EQUATIONS

Every server, switch, router, load balancer, firewall, and storage device you deploy in a data center will consume power. For this reason, the bulk of the monthly recurring charges (MRC) will be tied to power in some form.

Knowing exactly how much power your hardware will consume over a period of time—both when idle and in full production—is therefore key to ensuring you do not oversubscribe and overpay for services. The sophisticated data center buyer tracks exactly how much power is consumed per device over a period of at least six months, and is able to architect and design a data center environment to optimize the power density provided and the usable load on each circuit deployed within each cabinet. With this in mind, anyone buying or selling data center services *must know* the equations below, as they are the specific and primary products each data center provider is selling. For those specifically selling data center colocation services, knowing these equations is also absolutely essential if you hope to be viewed as anything more than "just a ticket-taking sales guy/girl" by your clients!

	Delivered Watts*	Usable Watts (80%)*
20 Amps x 120 Volts =	2,400	1,920
30 Amps x 120 Volts =	3,600	2,880
20 Amps x 208 Volts =	4,160	3,328
30 Amps x 208 Volts =	6,240	4,992
20 Amps x 208 Volts – 3 Phase =	7,197	5,758
30 Amps x 208 Volts – 3 Phase =	10,795	8636
50 Amps x 208 Volts – 3 Phase =	17,992	14,394
60 Amps x 208 Volts – 3 Phase =	21,590	17,272

* 3-Phase Circuits, Delivered Watts = Amps x Volts x 1.73

Figure 1-Essential Power Equations

Taking the step to memorize the table and equations on the previous page will give you the ability to quickly determine the needed power circuit configuration for a deployment based on the actual power load required. In conjunction with the specific $/kW rate of a given region or from a particular provider, these formulas will also help you quickly calculate the total monthly recurring bill based on power load required, without having to dive into a CRM system or having to call a sales engineer to run the calculations for you.

» CIRCUIT-BASED BILLING

In this model, the client is charged a fixed dollar-per-circuit-per-month rate, regardless of the power consumed. This in addition to rent for the physical space—cabinet, cage, or suite.

> ***Total Monthly Bill*** = (# of circuits delivered x $ per circuit) + (sq. ft. of cage delivered and/or # of cabinets used) ($ per sq. ft. and/or $ per cabinet)

Example

Power Need	2 x 30Amp 208V primary circuits 2 x 30Amp 208V redundant circuits, 2 x 20Amp 120V primary circuit 2 x 20Amp 120V redundant circuits Client will only be drawing @ 10 kW across all circuits delivered.
Space Need	2 cabinets. No cage.
Pricing	30Amp 208V Primary = $1,500/month
	30Amp 208V Redundant = $750/month
	20Amp 120V Primary = $600/month
	20Amp 120V Redundant = $300/month
	Cabinet = $500/month
Solution Cost Equation	($1,500 x 2) + ($750 x 2) + ($600 x 2) + ($300 x 2) + ($500 x 2) = $7,300/month

» PER kW, ALL-IN BILLING

In this model, clients are charged a fixed $ per kW rate that covers the cost of all power and space contracted. The number of circuits delivered within an environment is only relevant when looking at the install/setup fees charged to the client. Usage is tracked each month and, should a client exceed their contracted kW usage, an overage fee may be applied or the client will be forced to renegotiate the contract for a higher usage commit.

Total Monthly Bill = kW usage commit x $ per kW all-in rate

Note: when calculating a kW commit, this number should be rounded to a whole number. For example, 153.33=153 or 153.55=154

Example

Power Need	2 x 30Amp 208V primary circuits 2 x 30Amp 208V redundant circuits 2 x 20Amp 120V primary circuits 2 x 20Amp 120V redundant circuits Client will only be drawing @ 10 kW across all circuits delivered.
Space Need	2 cabinets. No cage.
Pricing	1 kW usable commit = $300/kW all-in
Solution Cost Equation	($300 x 10 kW) = $3,000/month

» METERED POWER BILLING

There are four components to metered power billing calculations:

1. kW commit per month
2. Actual monthly power usage
3. Cost of power from the utility (cents per kW hour)
4. PUE or some variation of "cooling and infrastructure uplift"

In this model, the client is charged a fixed dollar-per-kW commit that generally serves as rent to the data center service provider. In addition, the client is charged for the actual power consumed at the local utility rate multiplied by the contracted PUE.

Calculating hours per month = (365 days in a year/12 months) = 30.42 days per month x 24 hours = 730 hours per month

Cost of 1 kW of usage per month = 730 hours x cents per kW per hour rate = $ per kW per month

Total Monthly Bill = (kW commit x $/kW Rate) + (Actual kW Usage x 730 hours x Utility Rate x PUE)

Note: when calculating a kW commit, this number should be rounded to a whole number. For example, 153.33=153 or 153.55=154

Example

Power Need	2 x 30Amp 208V primary circuits 2 x 30Amp 208V redundant circuits, 2 x 20Amp 120V primary circuit 2 x 20Amp 120V redundant circuits Client will only be drawing @ 10 kW across all circuits delivered.
Space Need	2 cabinets. No cage.
Pricing	$195/kW rent
	$0.10/kW hour from utility
	PUE = 1.6
Solution Cost Equation	($195 x 10 kW) + (10 kW x 730 hours x $0.10 x 1.6 PUE) = $3,118/month

It is extremely important to note that within the metered power billing model your bill will almost always be ***variable*** and will change from month to month based on ***actual*** power usage. The 10 kW in red above will thus change each month based on actual power usage/consumption.

» EVOLUTION OF DATA CENTER COLOCATION PRICING

The three pricing models outlined above have most definitely evolved over the last decade. Originally, large wholesale customers were required to sign what amounted to a triple-net commercial lease. This effectively required the client to cover a fixed percentage of all incurred expenses of the facility, including taxes, insurance, and maintenance. If a client was to lease the entire facility and all available power, this would have meant they were paying for nearly 100 percent of that facility's operating costs. If a client was to lease a large suite that constituted at 25 percent of the available power and space within the facility, this would be at or near 25 percent of the facility's operating costs.

Over time, wholesale customers started demanding—and service providers soon began offering—a simpler means of calculating a monthly bill. This came in the form of the metered power billing model we have today, which is more along the lines of a modified gross commercial lease. The difference between a triple-net and a modified gross commercial lease is that a modified gross lease only requires the client to cover a fixed percentage of the facility's somewhat variable operating costs, versus a percentage of the total costs involved with owning and managing the facility. As such, within the modified gross model, the owner of the facility is entirely responsible for taxes and insurance and must recoup these fees indirectly from the rent portion of their clients' bills. In looking at the different components that make up the metered power model and equation, we see that this operating cost is covered within the actual power consumed and contracted PUE.

As a means of simplifying the billing process even further, some service providers started offering an all-in billing models to customers. Not surprisingly, this arrangement matches the terms of an all-in commercial lease, which charges a client a fixed dollar per square foot rate—nothing more. Any and all fixed and variable operations, maintenance, taxes, and insurance fees are covered by the owner of the facility, and therefore indirectly covered by the all-in rate extended to the client.

The commercial retail billing model has evolved drastically over the last five years, much to the dismay of some data center owner-operators. Historically, the only billing model offered to commercial retail customers was the circuit-based billing model. Under this model, it was easy for providers to bill for the specific circuits, cabinets and/or cage space a client contracted for on a month-to-month basis—and the majority of customers consumed less than 40 percent of the usable power on the power circuits they were paying for each month! The modified gross and metered power billing models share the ***actual*** operational costs of the facility with the client, so this model bills the client for the maximum ***theoretical*** operational costs of the facility, and specifically power, regardless of actual consumption.

Data center owners made—and still make—significant margins on every circuit they sell within their facility under the commercial retail billing model, in addition to the margins already built into the cabinet and cage pricing extended to customers. To put this in perspective, the average monthly income generated within data center facilities during the early 2000s was upward of fifty dollars per square foot, which is nearly ten to fifty times that of a pure play commercial warehouse or office building. This reality helps explain in part why so many investors and brokers from the commercial real estate sector have been flooding into the data center industry. Customers who fully understood this dynamic early on began optimizing every usable kW available within their contracted and deployed circuits to ensure they were not effectively wasting money.

Private cage on raised floor, image courtesy of Cologix

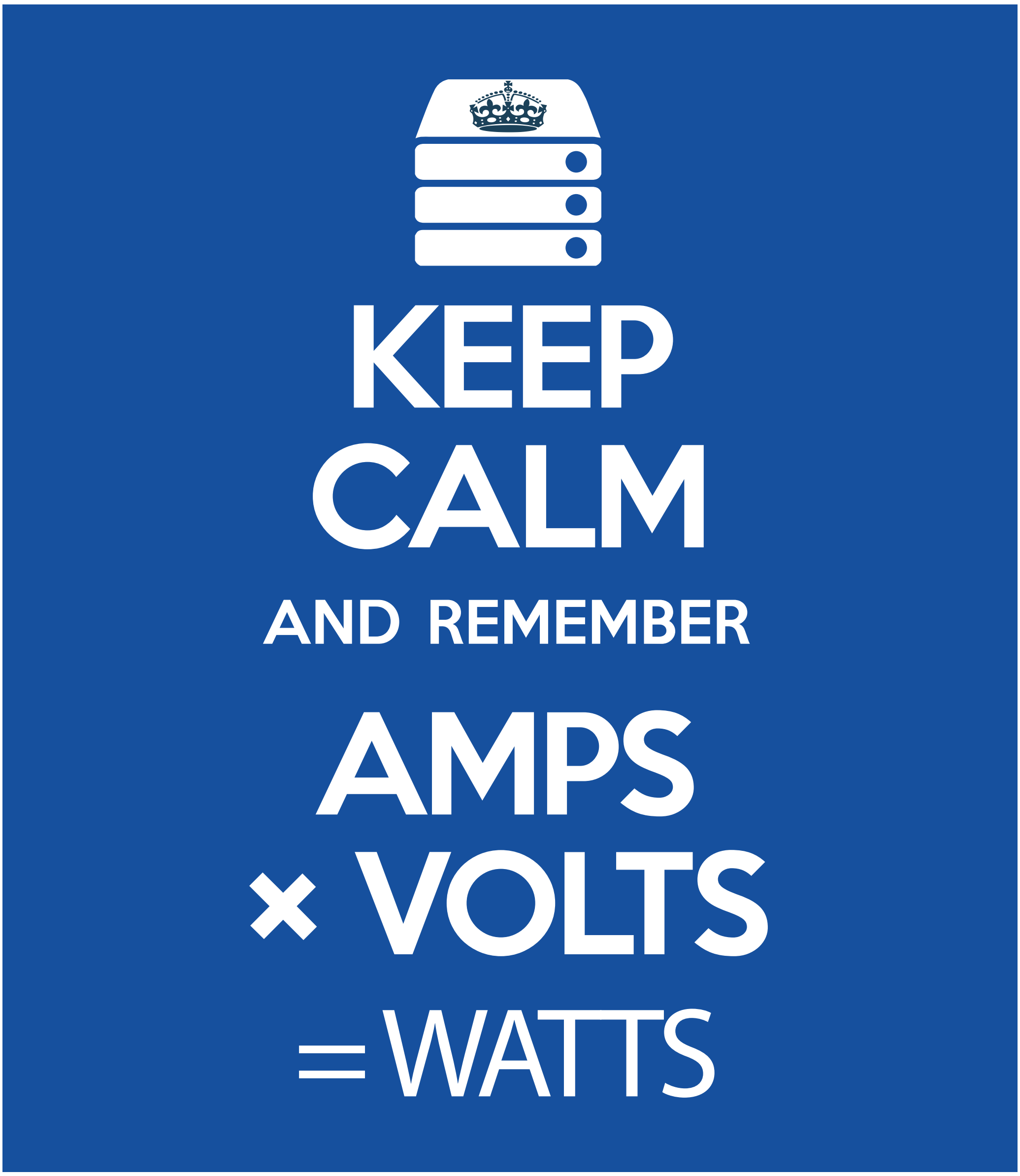
Such clients were, and still are, few and far between! Due to both market demand and service providers attempting to find creative ways to incentivize new customers—as seen above with wholesale clients—the all-in billing model soon became available for commercial retail clients as well. In the early 2000s, for example, RagingWire out of Sacramento gained relatively quick recognition and success within the industry by offering this all-in model to those looking for DR sites outside the inevitable earthquake zones up and down the West Coast. Although the all-in billing model does not guarantee power optimization within a service contract, it was a step in the right direction since clients were not forced to pay for the entire usable or delivered load of power. Rather, they could commit to a fraction of the available power delivered, based on what they thought a more reasonable projected load would be.

It is extremely important to point out that the only way an all-in billing model or metered power billing model can even be offered to a client is if there is a mechanism in place to monitor the actual load consumed within a given client's environment. For example, a client may commit to using 3 kW of the 4.992 kW delivered on a 30 amp, 208-volt circuit, but how does the service provider know for sure the client is at or below 3 kW? Although it has almost always been easy to read the entire load consumed within an entire RPP, this is most definitely not the case on a circuit-by-circuit basis. Many providers utilize branch circuit monitoring to monitor the power delivered through the power whips. Some even combine the A and B feeds to identify if they are over the 80 percent threshold. Those colocation facilities that do not provide branch circuit monitoring require a clamp on a circuit throughout the month to pull an average load, or preinstall an in-rack PDU or RPP that is capable of measuring and reporting this data back to the service provider on a regular basis. Either way, the ability to monitor actual load on any one circuit or small group of circuits within an environment was quite expensive and not considered standard practice until very recently.

Even if a commercial retail client wanted to be billed this way by a facility, unless the systems were already in place to allow proactive monitoring of power load on a circuit-by-circuit basis, monitoring loads may simply have been too expensive for the service provider to consider as a viable option. In addition, given the margins service providers were making on the unused power clients were paying for within the circuit based billing model, the service provider was less than incentivized to change its billing model.

Over the past five years, as an almost endless stream of new facilities came online in every major market around the world, technological progress and competition eventually forced the hand of service providers to start transitioning. The "lower cost of ownership," "flexibility," and "why pay for the power you're not using?" arguments repeated over and over again by new facilities and providers to clients have become so common that today the circuit-based billing model has become perceived as a legacy byproduct of old technology. Providers who do not have the technology in place to monitor power usage on a circuit-by-circuit basis are thus forced to either sell around this limiting factor or to try and reduce their per-circuit pricing to a level that can compete with the alternative and more flexible billing models offered in the local market.

This dynamic, with some service providers offering metered power to commercial retail colocation customers, has equated to a significant drop in the average dollar per square foot providers are achieving within both new and legacy facilities across the country. Only a small handful of facilities around the world have escaped this pricing drop. We will discuss the reasons why in a later chapter, but the short answer has to do with the fact that some data centers are simply far more valuable than others.


KEEP
CALM
AND REMEMBER
AMPS
× VOLTS
= WATTS

» PRICING MODELS CASE STUDIES

Case Study One

"Gaming platform WANParty is looking for a colocation provider that will support them in Santa Clara, California. Their equity partners are fanatics for marginal savings and have dictated extremely stringent requirements for how op-ex dollars will be spent relative to infrastructure growth. The client will be migrating out of its office and installing high-density compute racks and a low-density network rack. The prospects on the table need to demonstrate that they are getting the very best perceived 'value' in the environment as a whole, which unfortunately for WANParty means even the slightest savings per month will win the deal due to superior TCO."

Power, Space, and Terms Requested

- 7 compute racks at 9 kW per rack
- 1 network rack at not a watt less than 3 kW rack
- All kW numbers represent necessary usable load and client's historical records do not indicate load will spike beyond the load requested above
- Assume all power and racks will be "live" after the migration
- Assume 3-year term and approvals have been given by all providers to waive setup fees
- Assume cross-connect fees are the same from all providers

Solution Pricing Proposed

Provider A will be offering a circuit-based billing model

- Cabinet = $700 MRC
- Primary 30Amp 208V = $1,200 MRC
- Redundant 30Amp 208V = $600 MRC
- Primary 20Amp 208V = $795 MRC
- Redundant 20Amp 208V = $397.50 MRC

Provider B will be offering a per-kW, all-in billing model

- Per kW all-in rate = $275

Provider C will be offering a metered power billing model

- $145 per kW per month rent
- PUE = fixed 1.5
- $0.12 per kWh

Solution Equations And Questions

Provider A: circuit-based model spend	= MRC$
	= TCO$
Provider B: per-kW, all-in spend	= MRC$
	= TCO$
Provider C: metered power model spend	= MRC$
	= TCO$

What is the ideal pricing model for the client?

Why is this the ideal pricing model?

How might the power density of each facility above affect the pricing offered by each provider?

Case Study Two

"ScalePoP Inc. has been around for over three years. They have run the numbers on their cost within AWS over the next year, should they continue to scale as quickly as they are now. They have come to the conclusion that establishing a physical, managed environment within a data center will be a far better option than paying a premium to remain within the public cloud. Their technical cofounder/CTO is about to make a large hardware purchase and has little data to base his decision on as to actual power usage on this hardware once the production environment is migrated. The client's HQ is in Chicago, and they are hoping to keep their environment relatively close to home so they can stay hands-on for a few years until they have their systems, network, and related data center operations documented well enough to consider expanding or migrating the environment elsewhere."

Power, Space, and Terms Requested

- 2 racks with 1 primary and 1 redundant 30Amp 208V 3-Phase circuit per cabinet
- Assume all power and racks will be "live" shortly after install
- Assume 3-year term and approvals have been given by all providers to waive setup fees
- Assume cross-connect fees are the same from all providers
- Remember to round the total kW commit to the nearest whole number

Solution Pricing Proposed

Provider A will be offering a circuit-based billing model

- Cabinet = $500 MRC.
- Primary 30Amp 208V 3-Phase = $2,750 MRC
- Redundant 30Amp 208V 3-Phase = $1,375 MRC

Provider B will be offering a per-kW, all-in billing model

- Per kW, all-in rate = $600

Provider C will be offering a per-kW, all-in billing model

- Per kW, all-in rate = $400

Provider D will be offering a metered power billing model

- $275 per kW per month rent
- $0.11 per kWh
- PUE = fixed 1.6

Solution Equations And Questions

Provider A: circuit-based billing model	= MRC$
	= TCO$
Provider B: per-kW, all-in billing model	= MRC$
	= TCO$
Provider C: per-kW, all-in billing model	= MRC$
	= TCO$
Provider D: metered power model spend	= MRC$
	= TCO$

What is the ideal pricing model for the client?

Why is this the ideal pricing model?

How might the power density of the facilities above affect the pricing offered by each provider?

Case Study Three

"LargeCo Inc. has been around forever and is looking for a disaster recovery facility to setup an active/active environment mirroring their existing 450 kW deployment in Ashburn, Virginia. Their senior director of data center operations knows exactly what he needs and exactly what he should be paying based on their existing environment and general regional pricing, based on information he's pulled from friends in the industry."

Power, Space, and Terms Requested

- 1 primary and 1 redundant 30Amp 208V circuit per cabinet
- 4.5 kW average usage per cabinet per month
- 100-cabinet private suite
- Assume 5-year term and approvals have been given by all providers to waive setup fees
- Assume cross-connect fees are the same from all providers

Solution Pricing Proposed

Provider A will be offering a per-kW, all-in billing model

- Per kW, all-in rate = $210

Provider B will be offering a metered power billing model

- $135 per kW per month rent
- $0.06 per kWh
- PUE = fixed 1.6

Provider C will be offering a metered power billing model

- $135 per kW per month rent
- $0.06 per kWh
- PUE = fixed 1.4

Solution Equations And Questions

Provider A: per-kW, all-in spend	= MRC$
	= TCO$
Provider B: metered power model spend	= MRC$
	= TCO$
Provider C: metered power model spend	= MRC$
	= TCO$

What is the ideal pricing model for the client, and who presents the best option?

Why is this the ideal pricing model?

How might the willingness of each provider to extend flexible ramp terms affect the total cost of ownership (TCO) for the client from each provider?

Chapter Summary

- Calculating wattage: Single phase circuits
  - Watts = Amps * Volts
  - Usable Watts = Watts * 80%
- Calculating wattage: Three-phase circuits
  - Watts = Amps * Volts * 1.73
  - Usable Watts = Watts * 80%
- MRC: Circuit-based billing
  - Total Monthly Bill = (# of circuits delivered x $ per circuit) + (sq. ft. of cage delivered and/or # of cabinets used) ($ per sq. ft. and/or $ per cabinet)
- MRC: Per-kW, all-in billing
  - Total Monthly Bill = kW usage commit x $ per kW all-in rate
- MRC: Metered power billing
  - Calculating hours per month = (365 days in a year/12 months) = 30.42 days per month x 24 hours = 730 hours per month
  - Cost of 1 kW of usage per month = 730 hours x cents per kW per hour rate = $ per kW per month
  - Total Monthly Bill = (kW commit x $/kW rate) + (actual kW usage x 730 hours x utility rate x PUE)

CHAPTER FOUR: CONTRACTS

After studying this chapter, you should be able to:

- Identify an infrastructure buyer's key concerns and negotiation leverage points.
- Identify an infrastructure seller's key concerns and negotiation leverage points.
- Discuss concerns with buyers or sellers with empathy and understanding.

One of the most important pieces of the equation when purchasing or selling data center services is the legal contract signed by the client and the provider. Unfortunately, these contracts are far too often overlooked by the client or reviewed by attorneys who are unfamiliar with the industry. Clients and service providers have both been taken advantage of as a result, costing hundreds of millions of dollars in legal fees and damages.

» ISSUES BUYERS CARE ABOUT

Automatic Renewal Clauses

These clauses automatically sign a client up for an additional—and often lengthy—term of service without the client having to proactively negotiate for or approve a contract extension. In some states it is illegal for a provider to auto-renew a client without notification and confirmation of receipt. Auto-renewal language is increasingly being removed from data center service provider contracts because, over the past few years, a majority of clients have refused to agree to such clauses.

Termination Clause For Chronic Outages

When multiple outages or other SLA violations occur within a defined period of time a client should have the ability to walk away free and clear from a contract, without paying early termination fees.

If, however, a client wishes to terminate a contract for convenience (i.e. without cause), they will generally be required to pay a percentage of the remaining expected and contracted fees due per the contract's stated duration. This percentage varies from zero to 100 percent, depending on the remaining term in the contract. Although I've seen a few cases where a client successfully waived all early termination fees, it is an infrequent occurrence. This is primarily due to the reality that a number of important financial variables are tied to the value of each provider's contract.

Enforceability Of The "Uptime" Guarantee

Service outages within data centers cost clients deployed in the facility money, so providers offer compensation for not delivering on their "uptime" commitments. A Service Level Agreement (SLA) will provide guidance as to how much compensation a client will be awarded as a function of the length of the outage and a percentage of the contracted monthly recurring bill. Far too frequently, the compensation offered by providers who fail to live up to their "guarantee" is just a credit equal to a small fraction of the monthly recurring amount a client is paying. Given the inability of an SLA to be enforced, there is little incentive for a provider to live up to its "uptime guarantee."

Maintenance Windows

If a data center meets concurrently maintainable standards or better—Tier or Class III or higher—maintenance windows should not be excluded from the definition of "downtime" within the SLA. Again, far too often providers try to exempt themselves or provide loopholes to keep them from being held responsible for living up to their uptime "guarantee."

Insurance Requirements

Almost all providers of data center colocation services require their clients to maintain some level of liability insurance. Though this has not always been the case, more and more providers are now requiring their customers to provide proof that they have such insurance in place. This proof requirement is typically driven by risk standards established by the provider's insurance company. Requiring clients to have insurance is reasonable, as it ensures there is a clear path to some form of recourse and compensation should a client, due to gross negligence or willful intention, perform an act that harms either the facility itself or the other clients' property or data within the facility.

Indemnification

This type of clause speaks to who is responsible, and to what extent, should certain actions adversely affecting either party occur within the data center. It is important to ensure that this language defines which party is being indemnified—and for what—as specifically as possible. It is also important to ensure the obligations and recourse are reasonably mutual between the parties, and not heavily in favor of the data center's owner.

Prenegotiated Price Increases

Almost all wholesale agreements contain annual price increases, even fixed-term contracts. It is also almost guaranteed that a commercial retail service provider will attempt to increase a client's monthly recurring rates by 3 to 15 percent at the end of a one- to three-year term. It is therefore extremely important for clients to negotiate ***ahead of time*** exactly what the rate increase will be after the end of the initial term. This will prevent a provider from drastically increasing rates unexpectedly. Although the sales executive or manager working with you on the initial contract may swear up and down that such a rate increase will be "reasonable" and never be more than X percent, or will never occur at all, the reality is that he or she will likely not be with the organization or in the same role in two or five years, when your contract is up for renewal. A new management team may also be put in place by then, with a different standard for what is and is not "reasonable." Hence, it is always best to get this in writing ***within*** the service contract itself—not simply over e-mail.

Prenegotiated Costs For Additional Services

It is extremely important to ascertain ahead of time what the rates will be—both immediately and during the remainder of the term on your contract—for remote hands/eyes, cross-connects, power circuits, network upgrades, etc. Negotiating for great rates on initial services is good, but not if you find that the rates for the new or upgraded services you'll need as you expand equate to 150 or 200 percent of what you start off paying. Save yourself guesswork and frustration by identifying ***at the outset*** all of the services you might need to request from your service provider during the term of your contract. Ensure that your agreement expresses ***in writing*** exactly what those services will cost you should you order them later.

Coterminous Contract Terms For Services

It is extremely helpful when negotiating for new services to ensure that any new contracts have an end date that is the same as your existing services agreement. If you intend to renew, rope all your services into a new agreement with the same contracted end date. Do not expect your sales executive to bring this to your attention, as sales executives are generally compensated on a percentage of the total contract value and are thus incentivized to ensure your new contract is for as long of a term as possible.

» LEVERAGE POINTS FOR BUYERS

Size Of Initial Deployment

The larger the size of your initial deployment with a service provider, the more leverage it can give you. Large orders equate to a larger contract value over the short term—and long term—for the provider.

Contracted Growth

The more you can commit in writing to growing over time, the more leverage you will have with a provider. Growth, like a large initial order, will equate to a larger contract value over the long term for the provider. This will also allow the facility manager to plan for changes in the use of power and cooling across the data center floor more accurately. However, in today's environment—with virtualization and manufacturers making systems more efficient—what is being seen is actually a decrease in power consumption and physical footprint. If a first right of refusal is available, it may be a worthwhile alternative.

Term Length

With some exceptions, which stem mostly from providers hoping to raise a client's rates significantly after the end of an initial term, a longer contract term length will give you more leverage with a provider. Long-term contracts are very valuable to the provider, because the total contract value is usually significant. Longer term lengths also mean that neither you nor the provider needs to spend time or resources renegotiating contract extensions or expansions.

Brand Recognition

A highly recognized brand may be given preferential pricing and treatment. As such, a brand's logo on data center advertising and financial materials lends credibility to a provider. New clients will know someone larger than they has already put the provider through an intensive vetting process. Sales executives are also almost always incentivized to pick up new companies, as more clients mean a lower cost of net new revenue down the road, when those clients start expanding and growing.

Newness Of The Facility And Space Availability

Given the extensive capital required to build, operate, maintain, market, and sell a data center, providers with new facilities are generally frantic to cover their monthly overhead as soon as possible. Providers in this situation are generally willing to give the most beneficial pricing and terms to the initial subset of clients who contract to deploy within the facility.

Competition

Simple supply-and-demand economics play out here, as they do in nearly every other industry. The more options you have available to you, the more leverage you will have when negotiating for services.

Tier Or Class Level Of Data Center

All data centers are not created equal. Lower tiers or classes of data centers are cheaper to build and maintain. The downside is that class level correlates directly to more frequent chances of interruption in power or cooling. Buyers should be aware of the cost to them of downtime and should research a data center's classification level prior to contract.

Multiple Services

If your total spend with a service provider includes multiple services such as network hosting, cross-connects, etc., you will have more leverage when negotiating. Not only will the total value of the contract increase for the provider, but more services mean that the contracted dollar per square foot and dollar per kW will also increase for them. This is highly relevant, as the sales leadership of the provider company generally measures success by looking at the average dollar per square foot and/or average dollar per kW within each facility.

Downtime Allowed		
Uptime SLA Offered	Month Only	Year Only
100%	0	0
99.999% = "Five Nines"	26 seconds	5 minutes, 29 seconds
99.99% = "Four Nines"	4 minutes, 23 seconds	52 minutes, 34 seconds
99.9% = "Three Nines"	43 minutes, 48 seconds	525 minutes, 36 seconds

Figure 2-Uptime percentage translated to time in a calendar month

The table above will help you understand what an offered SLA will amount to in actual downtime, should the SLA be offered over the span of a month or year. The period of time most providers use when calculating their uptime SLA is a month, so the acceptable downtime is in the column titled "Month Only." It is important that this is clearly defined within the SLA contract, as there is a significant and clear difference between a "Five Nines" SLA calculated each month versus each year!

Any outage, no matter how long, can end up keeping a client out of commission much longer than simply the time the actual power or network is down. This is because it can take minutes, hours, or even days for a company to bring the entirety of its system back to a production state. There are documented cases of small power outages affecting large environments and keeping a company's systems off-line for weeks and even months, and that precipitated a complete redesign and rebuilding of the affected applications and systems themselves after multiple failed attempts to recover were unsuccessful.

The "teeth" in an SLA, should a violation occur, are also relevant. It is sometimes hard to take current industry standards seriously when only a fraction of the monthly recurring bill from a provider will be returned to the client as financial remedy due to an outage. For this reason, the SLA should ***never*** be expected to provide proper or full compensation for the effects of an outage; a proper Managed Recovery Program (MRP) should be implemented if downtime would cause serious financial "pain" to a company.

It may sound surprising, but in most cases human error is the root cause of an outage, not a failure of the mechanical systems. Only a minority of facilities around the world have been online for longer than five years without having some form of power outage affecting even a limited number of clients within the facility, regardless of the redundancy of the architecture deployed. In fact, some believe that, the more redundancy that is built into a system, the more prone to failure it may actually become. Advocates of this theory argue that the increased complexity and number of moving parts within the system increase the potential for failure due to either human or mechanical causes.

» ISSUES SERVICE PROVIDERS CARE ABOUT

Contracted $/kW or 4/sq. Ft.

Nearly every commercial data center around the world has targets tied to the dollar per kW or dollar per square foot achieved within the facility. As such, it is almost a certain that every contract signed with a client within a facility, regardless of how it is being billed, has at some point been reduced to and evaluated by one or both of these metrics by the provider.

Forecasted Revenue Targets

In most cases, a data center owner is highly incentivized to hit forecasted monthly, quarterly, and annual revenue targets to demonstrate to its creditors, investors, and/or the public markets that it is still a sound investment. Access to additional debt, capital, and increased corporate valuations are assuredly tied to this achievement.

Limiting Liability

Should an outage occur, a provider must ensure that the contracted SLA reimbursements will not bankrupt the business. In part, this is why the only "teeth" offered to back up most SLAs are nominal credits that at best amount only to a fraction of the monthly bill each client pays. You will almost never see a data center provider take responsibility for projected economic business losses or third-party impact due to an outage. Also, liability almost always remains limited to the specific facility used by the client and does not spread to the parent company or across all the facilities under ownership or management from the parent company. The total liability a provider will take ownership for due to an outage or gross negligence occurring within one facility will thus almost never extend beyond the financials and management of that local facility.

Client Insurance

See details above, under "Insurance Requirements" within the "Issues Buyers Care About" section of this chapter.

Access To Cabinets, Cages, And/Or Suites

Client attorneys tend to try and argue that a service provider should never be given access to their client's environment within a data center, or at the very least should require an approval from the client before entering. The reality, however, is that service providers ***must have access*** in case of an emergency. The logic behind this firm stance from service providers is sound and reasonable. If for some reason there is a fire within a client's environment, wouldn't you want and expect the service provider to act, with prudence, to extinguish the fire and prevent its spread to other clients' hardware hosted within the facility? There is no regulatory or compliance requirement we have come across to date that mandates a data center service provider cannot enter a client's environment within the context of such a scenario. If you work in an industry where such strict requirements could be a real consideration, then placing your IT infrastructure within a commercial colocation facility likely isn't the ideal solution for your needs.

Termination Of Services

When a client stops paying for contracted services or routinely violates other usage or nondisclosure terms of an agreement, a service provider will always fight to retain its right to stop delivering power and related services to the client and to do everything in its power to receive compensation for the lost contracted revenue. Although there are some states and some forms of bankruptcy filings that make the process of terminating services with a client complicated for the service provider, it is a common myth that a service provider is obligated to continue delivering power and services to a client should they stop payment and/or file for bankruptcy.

Happy Clients

Although a service provider's main goal is always to try to make as much money as possible and to be liable for as little as possible, we have found that almost all providers do genuinely want to develop long-term, productive partnerships and relationships with their clients. The cost of acquiring net new business is always far greater than the cost of receiving renewal revenue from an existing client. Miscommunication between a client and a service provider is almost always at the root of major points of contention between the two parties. Unless there are outside political or economic forces driving a hard-line stance on a topic, a service provider will accommodate a reasonable request for modifications to be made to a contract or services agreement.

» LEVERAGE POINTS FOR SERVICE PROVIDERS

Size, Term And Growth

The size, contracted growth, and term of a deal will most assuredly dictate how willing a provider is to negotiate the terms of the agreement and how flexible they want to be with said terms. The hard reality is that nearly every point that is negotiated away from a standard service template will require management and legal resources internally. Providers will always ask themselves if the long-term margin on a deal is worth the short-term time and money outlay needed to secure the deal. Hence, single-cabinet deals with nominal power requirements and no contracted growth tend to be accompanied by a "take it or leave it" attitude from most service providers. Larger cage or suite deals, on the other hand, will likely receive full attention from the service provider, who will be extremely responsive and as aggressive on pricing and flexible on terms as possible.

Availability Within The Facility

Established facilities in large markets with various connectivity options and key clientele—such as a financial exchange, online advertising exchange, or IaaS provider—are almost always at or near 100 percent capacity and are therefore able to command a premium in the market relative to the other data center options available. It is also the case that once the revenue a provider is receiving on a property comes close to or matches the monthly spend, rates will go up for new customers and net new services.

Proximity And Quantity Of Key Network Providers

The general rule is that the more network providers available within a facility, the higher the demand—and likely the price—for the facility. This cost may be offset somewhat by the fact that a client is likely to get a far better price for network services when there is competition among providers. For those pushing significant bandwidth on a regular basis, a local peering exchange is almost always a tool leveraged to drive down their total cost for network services and to develop faster routes for clients to access to their content.

Proximity And Connection To Local IX(s)

If a data center can serve as an extension of the local Internet Exchange (60 Hudson in New York City, PAIX in Silicon Valley, One Wilshire in Los Angeles, 365 Main in San Francisco, 350 Cermak in Chicago, etc.), it can in essence quickly and easily interconnect clients within the building to any of the network providers within the local IX. Data centers in close proximity to, or in direct connection with, the IX enjoy some leverage.

Competition

Although an increasingly rare occurrence today, a provider will have leverage if there is limited or no competition within a local market providing similar or the same services.

Cost Of Power

With power serving as one of the primary operational costs that both providers and clients pay for data center services, those providers with facilities located in regions with a lower cost of power have a competitive advantage from a pure total-cost-of-ownership perspective.

Energy Efficiency

Less power needed to run a data center translates to lower costs associated with the ongoing operations of the facility. Providers who invest in the technology and resources to optimize their energy efficiency programs generally earn higher margins and offer a lower cost of ownership for clients.

Multiple Locations

If a client expects to expand its footprint within a given region, continent, or even around the world, that will be a key leverage point for service providers who do in fact have multiple locations available.

Managed Networks

Most small commercial retail colocation customers would prefer not to need to negotiate separately for limited network services in addition to power and space services. Therefore, those providers who can quickly quote and deliver some form of managed network solution have a leg up on the competition.

Managed Infrastructure And Related Services

Although clients may never need their service provider for anything more than delivering stable power within a secure colocation facility, providers who can offer additional options like managed DR, systems, database, or application services have a psychological advantage with clients. This is not so much because the majority of commercial retail colocation clients actually ***need or will use*** these services, but because they like to know that these services are available should they be needed down the road.

Security

A corporation's data is far more likely to be stolen or compromised via an attack over a network than by someone physically walking in and stealing the hardware it sits on. One's perception almost always drives one's reality, however, and data centers with established physical security, secure access procedures, and security personnel are able to give clients the peace of mind that their data is at least physically safe.

Remote Hands And Eyes

Almost all commercial retail data center operators have a team available to perform basic tasks for a client within its cabinet, cage, or suite. How competent a team is at quickly responding to and resolving problems and communicating back to the client is what truly differentiates providers within the industry. Those providers who can prove they deliver on what they promise have more leverage when negotiating with clients.

Facility Uptime Track Record

Though outages within data centers are unfortunately commonplace, some facilities are more prone to downtime than others. Obviously, facilities that have had limited or no customer-impacting outages are more favorable. It is unfortunately shocking how few clients ask their providers when the last customer-impacting event occurred at a facility before signing up for a service agreement.

Chapter Summary

Buyer concerns and leverage points

Concerns:	Leverage points:
• Automatic Renewal Clauses	• Size of initial deployment
• Termination clause for chronic outages	• Contracted growth
• Enforceability of uptime guarantee	• Term length
• Maintenance windows	• Brand recognition
• Insurance requirements	• Newness of the facility and available space
• Indemnification	• Competition
• Prenegotiated price increases	• Multiple services
• Prenegotiated costs for additional services	
• Coterminous contract terms for service	

Seller concerns and leverage points

Concerns:	Leverage points:
• Contracted $/kW or 4/sq. ft.	• Size, term, and growth
• Forecasted revenue targets	• Availability within facility
• Limiting liability	• Proximity of key network providers
• Client insurance	• Proximity to and connection with local IX
• Access to cabinet and/or suite	• Competition
• Termination of services	• Cost of power
• Happy clients	• Energy efficiency
	• Multiple locations
	• Managed network
	• Other managed services
	• Security
	• Remote hands and eyes
	• Facility uptime record

CHAPTER FIVE: AUDIT AND COMPLIANCE OVERVIEW

After studying this chapter, you should be able to:

- Describe the key auditing organizations and their compliance requirements.
- Communicate to your client the significance of a facility's certifications.
- Describe where to turn to in the book after you've forgotten all about key auditing organizations and their compliance requirements.

» AICPA

The American Institute of Certified Professional Accountants (AICPA) is the world's largest member association for the accounting profession, with more than 394,000 members in 128 countries.

» SAS 70

> *For nearly eighteen years, the Statement on Auditing Standards 70 (SAS 70) was the authoritative guidance document that allowed service organizations to disclose their control activities and processes to their customers and their customers' auditors in a uniform reporting format. The issuance of a service auditor's report prepared in accordance with SAS 70 signifies that a service organization had its control objectives and control activities examined by an independent accounting and auditing firm. The service auditor's report, which includes the service auditor's opinion, is issued to the service organization at the conclusion of a SAS 70 examination. (http://sas70.com/sas70_overview.html)*

It is important to note that SAS 70 audits simply reported whether a provider was following the internal processes and procedures they had set for themselves. That means that there were effectively no standards or best practices that a provider was measured against besides its own. Many companies also mistakenly viewed and spoke to the SAS 70 audit as a form of certification process, which was not the case. Those clients who were required to follow Sarbanes-Oxley would often mistakenly check a box on their report upon seeing that their provider had been through a SAS 70 audit, completely unaware that the opinion of the auditor cited within the report may have identified numerous critical areas where improvement was recommended due to failure to meet or exceed the providers own self-assessed standards.

» SSAE 16

The Auditing Standards Board of the AICPA finalized its "Reporting on Controls at a Service Organization" in January 2010, and issued a resulting Statement on Standards for Attestation Engagements 16 (SSAE 16), which effectively replaced SAS 70 as the authoritative guidance for reporting on service organizations. The SSAE 16 was formally issued in April 2010 and became effective on June 15, 2011. It was drafted with the intention and purpose of updating the US service organization reporting standards so that it mirrors and complies with the new international service organization reporting standard, ISAE 3402. Almost all data center providers that had previously undergone a SAS 70 audit converted to the new standard in 2011, and now have issued SSAE 16 reports instead.

» SOC REPORTS

Service Organization Control Reports® are internal control reports with guidelines established by the AICPA on the services provided by a service organization. They provide valuable information to clients and potential clients, who can use them to assess and address the risks associated with an outsourced service.

SOC 1, Types 1 and 2

The SOC Type 1 report focuses on a description of a service organization's systems and on the suitability of the design of its controls to achieve the related control objectives included in the description, as of a specified date.

A Type 2 report contains the same opinions as a Type 1 report, with the addition of an opinion on the operating effectiveness of the controls used to achieve the related control objectives included in the description throughout a specified period. A Type 2 report also includes a detailed description of the service auditor's tests of controls and results.

The use of these reports is restricted to the management of the service organization, user entities, and user auditors.

SOC 2

This report on controls at a service organization will be relevant to security, availability, processing integrity, and confidentiality or privacy. It uses the AICPA's Trust Services Principles and Criteria of: security, availability, process integrity, confidentiality, and privacy. It's similar to an SOC 1 in that a Type 1 or Type 2 report is available. It also includes a description of the service auditor's tests of controls and results. Use of this report may be restricted.

SOC 3

This is a Trust Services report for service organizations. It covers the same subject matter as SOC 2; it does not include a description of the service auditor's tests of controls and results. Also, the description of the system is less detailed than the description in a SOC 2 report. A seal can be posted on a service organization's website to verify that it has performed this audit. The Canadian Institute of Chartered Accountants (CICA) administers a seal program for these engagements, if the CPA is licensed for the seal by the CICA. The use and distribution of this report is not restricted.

» HIPAA

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) establishes national standards for protecting sensitive health care patient data. Any company that deals with protected health care information must ensure that all the HIPAA-required physical, network, process, and security measures are in place and followed.

» PCI DSS

The Payment Card Industry Data Security Standard (PCI DSS) represents a common set of industry tools and measurements to help ensure the safe handling of sensitive information. Initially created by aligning Visa's Account Information Security (AIS)/Cardholder Information Security (CISP) programs with MasterCard's Site Data Protection (SDP) program, the standard provides an actionable framework for developing a robust account data security process—including preventing, detecting, and reacting to security incidents.

» FISMA

The Federal Information Security Management Act (FISMA) is a federal law enacted in 2002 as Title III of the E-Government Act of 2002. The law assigns specific responsibilities to federal agencies, the National Institute of Standards and Technology (NIST), and the Office of Management and Budget (OMB) in order to strengthen information security systems. In particular, FISMA requires the head of each agency to implement policies and procedures to cost-effectively reduce information technology security risks to an acceptable level. According to FISMA, the term "information security" is defined as protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide integrity, confidentiality, and availability.

» NIST

The National Institute of Standards and Technology (NIST) is a nonregulatory agency of the US Department of Commerce. Its general stated mission is to "promote US innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life" (REF 57-A).

Specifically, NIST is responsible for creating and updating two complementary publications that provide guidance and recommendations for ensuring data security and privacy protection in federal information systems and organizations—a role assigned to NIST under FISMA. The publications are so well-known by now that they are generally referred to just by their numbers.

The first *Security and Privacy Controls for Federal Information Systems and Organizations* (Special Publication 800–53), is an encyclopedic catalog, organized by function, of available methods or "controls" that can be established to safeguard an information system—no matter how small or large.

Supplemental guides are updated and released on an almost annual basis, which serve as companions to SP 800–53. The most recent is *Assessing Security and Privacy Controls in Federal Information Systems and Organizations: Building Effective Assessment Plans* (SP 800–53A). If SP 800–53 is all about planning for appropriate controls to safeguard an information system; SP 800–53A is a methodology for determining how well you did. REF 57-B

Fisma, Nist, And The Data Center Industry

It is important to note that SSAE16 SOC 1, Type 2 has nothing to do with FISMA. First, for a SOC 1, Type 2 the customer identifies risks and mitigation steps; FISMA follows NIST 800–53. Please note that these are two different standards. FISMA is a standard defined by NIST. These audits ensure that all controls are applied and tested on an ongoing basis within the data center, and give the facility the ability to proclaim that they comply with federal regulations and are thus FISMA/NIST compliant.

» FedRAMP

Federal Risk and Authorization Management Program (FedRAMP) is a government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services. This approach uses a "do once, use many times" framework that saves cost, time, and staff required by eliminating redundant agency security assessments.

FedRAMP is mandatory for federal agency deployments and service models at the low- and moderate-risk impact levels hosted with third-party commercial service providers. Privately hosted deployments intended for single organizations and implemented fully within federal facilities are the only exception. Additionally, executive departments and agencies must submit a quarterly report listing all existing hosted services that do not meet FedRAMP requirements, along with an appropriate rationale for why not and proposed resolutions for achieving compliance.

As FedRAMP is primarily concerned with the processes and procedures around how a company is managing data, data center facilities themselves have no need to be FedRAMP authorized. The IaaS and managed hosting providers ***within*** the data center, however, would be required to work through the FedRAMP authorization process, and the data center provider will likely be approached to answer questions about who has access to the hardware the data is stored within and what processes and procedures have been put in place to safeguard access. Many companies have simply aligned their SSAE16 with FedRAMP. To accomplish this, they ensure their risk assessments contain items from FedRAMP. It is less work and less expense to have a single auditor complete the audit.

» ISO

The International Organization for Standardization (ISO) is an independent, nongovernmental membership organization, and the world's largest developer of voluntary International Standards. The ISO story began in 1946, when delegates from twenty-five countries met at the Institute of Civil Engineers in London and decided to create a new international organization "to facilitate the international coordination and unification of industrial standards." In February 1947, the new organization officially began operations as the ISO.

Since then, they have published over 19,500 International Standards covering almost all aspects of technology and manufacturing. Today the ISO maintains members from 166 countries and 3,368 technical bodies to take care of standard development. More than 150 people work full time for the ISO's Central Secretariat in Geneva, Switzerland.

Due to the governing bodies and standards already in place for data center facilities and service providers within the United States, very few companies here bother requesting information regarding ISO certifications—few are even familiar with the various different ISO certifications offered and available. As ISO is an international organization, however, it is common to see requests for, and references to, specific ISO certifications within RFPs coming from companies headquartered elsewhere but looking to acquire new facilities here. Some of the relevant ISO certifications you may come across in our industry are listed below.

ISO 9001: Quality Management Standard

ISO 9001 is the internationally recognized standard for the quality management of businesses. The ISO 9001 standard applies to the business processes that create and control the products and services an organization supplies. It prescribes systematic control of activities to ensure that the needs and expectations of customers are met.

ISO 22301: Business Continuity Management Standard

ISO 22301 provides a framework for managing business continuity within an organization and replaced ISO BS 2599-2, serving as an “upgrade” from the previous standard. When implemented properly, business continuity management decreases the possibility of a disruptive incident, but if such an incident does occur, it helps an organization be ready to respond in an appropriate way. Adherence to this ISO thus drastically decreases the potential damage from such an incident.

ISO 27001: Information Security Management Standard

ISO 27001 is a specification for the management of information security. It is applicable to all sectors of industry, government, and commerce—and not confined to information held on computers. It addresses the security of information spanning: security policy, organizational security, asset classification and control, personnel security, physical and environmental security, communications and operations management, access control, system development and maintenance, business continuity management, and compliance.

With all the above certifications and compliance standards, it is important to understand where the responsibilities of the provider and customer start/end, respectively. For example, if a standard colocation facility is getting PCI DSS certification, the colocation facility would only be looking at sections nine and twelve—the physical security and policies—since it does not have logical access to the customers’ systems or internal access to their networks. This is something that needs to be discussed with an auditor prior to initiating a contract for an audit, because it will significantly impact the price and time for the audit.

» ASHRAE

The American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE) was founded in 1894 to serve as a source of technical standards and guidelines. Since that time, it has grown into an international society that offers educational information, courses, seminars, career guidance, and publications. The organization also promotes a code of ethics for HVAC professionals and serves as a liaison with the general public. One of the organization’s most important functions is to promote research and development in efficient, environmentally friendly technologies.

As cooling is such a critical component within the data center, it is no wonder that ASHRAE is very involved with the development of modern data facilities. The ASHRAE Technical Committee 9.9 created an initial set of thermal guidelines for data centers in 2004, followed by a second edition in 2008 and a third in 2011. Each new iteration of the guidelines outlined classes of servers and other data center equipment that could function at better than previously acceptable levels of operating temperature and humidity. When these standards are properly implemented within the data center, it can reduce reliance on traditional mechanical cooling methodologies to save on cooling costs and extend the reliability of its data center hardware. It is important to note, however, that ASHRAE only provides recommendations and suggestions. Data center facilities are not bound to follow any of their guidelines.

Most customers request that ASHRAE standards are followed, and in fact request a clause to that effect in their contracts. Utilizing standards provides customers with a sense of confidence that the operation of a facility is mature. Facilities need to be careful of is how long it may take to get back up to standards if they fall outside the window or, at a minimum, have some verbiage to protect themselves. If the customer is adding systems, thereby adding heat to the floor, it may take some time for the floor to adjust to the additional heat load. Also, some customers may request a humidity level that only allows fifteen points from the low to the high—40 percent to 55 percent allowable, for example. Again, adding or removing systems will impact the humidity on an entire floor, and it will take time for the cooling equipment to adjust.

Environmental Conditions				
ASHRAE Thermal Guidelines for Data Processing Environments				
Table 2.1: Equipment Environment Specifications				
Class	Allowable Dry Bulb (°C)	Recommended Dry Bulb (°C)	Allowable % Relative Humidity	Recommended % Relative Humidity
1	15 to 32	20 to 25	20–80	40–55
2	10 to 35	20 to 80	20–80	40–55
3	5 to 35	NA	8–80	N/A
4	10 to 40	NA	8–80	N/A
Max. altitude = 10,000 feet with dry bulb de-rating applied in some cases.				

Figure 3-ASHRAE thermal guidelines for data processing environments, courtesy of ASHRAE

Chapter Summary

Introduction to various auditing organizations.

- AICPA: World's largest organization of accountants.
- SAS 70: Signifies that an independent accounting firm audited an organization.
- SSAE 16: Replacement for SAS 70 as of 2011.
- SOC reports: prepared internally to help users assess risk.
- HIPAA: National standards for protecting patient data related to health care.
- Payment card industry data standard: joint security protocol developed by Visa and MasterCard.
- FISMA: Federal standards for federal agency data security practices.
- NIST: Creates and promotes nonregulatory guidelines for data security in federal systems and organizations.
- FedRAMP: The US federal government's authorization program for managing off-site data security.
- ISO: Nongovernmental organization that publishes standards for everything. Really, EVERYTHING.
- ASHRAE: Provides suggestions on cooling implementation in data centers. Yes, there is a whole organization devoted to this.

Shell managed wholesale raised floor, image courtesy of QTS

DATA CENTER DESIGN

CHAPTER SIX: POWER

After studying this chapter, you should:

- Be able to accurately describe how kVA differs from kW, describe how single-phase and three-phase circuits differ, and describe how these and other elements work together to deliver power to racks.
- Be able to calculate PUE and describe how it applies to infrastructure buyers.
- Describe in layman's terms what clients mean when they state their technical needs, as well as what solutions will or will not serve them appropriately.

» kVA AND kW

Both kVA and kW are primary units of measurement for power in the data center industry. They are sometimes used interchangeably, although there are subtle differences between the two that must be taken into account. The first term, kVA, specifically refers to kilo volt amperes and is called "apparent power." The second, kW, specifically refers to kilowatt and is called "real power." A third term, kVAR, is also used when speaking about these terms and refers to the difference between the kVA and kW; it is called the "power factor."

In direct current (DC) circuits, kVA is equal to kW because the current and voltages do not get out of phase during transmission. However, "real power" and "apparent power" differ in alternating current (AC) circuits—which is what the majority of commercial retail colocation facilities deliver to their clients on a data center floor.

The power factor leads or lags, depending on the way that the load shifts the phase of the current compared to the phase of the voltage. In AC circuits, voltage and current may get out of phase, leading to a difference in kW and kVA that will be based on the power factor.

The relationship between kVA, kW, and the power factor is described mathematically as:

- kW = kVA x power factor
- kVA = kW/power factor
- power factor = kW/kVA

The illustration below provides a practical, albeit comical, representation of the difference between kVA and kW:

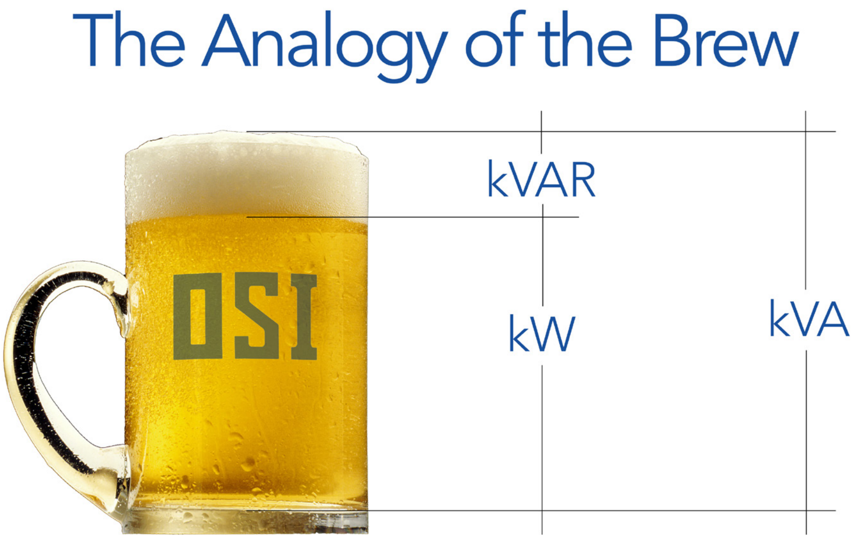


Figure 4-The relationship of kW to kVAR to kVA

» 120 VOLT, 208 VOLT, AND THREE-PHASE POWER

To understand how power is delivered in the data center you need to first understand the difference between single-phase and three-phase power distribution. Most residential homes are wired with single-phase power; this uses one AC voltage delivered over two hot wires, one neutral wire, and a ground. The voltage across the two hot wires measures 240 volts for large appliances such as an oven or dryer, and across any hot-to-neutral measures 120 volts, which accommodates nearly everything else in the home.

Most commercial businesses are wired with three-phase power that consists of three AC voltages separated from one another by 120 electrical degrees, or by a third of a cycle. These systems deliver power over three hot wires, where the voltage across any two hot wires measures 208 volts. Another way to look at three-phase power is as a combination of three single-phase circuits that deliver power in such a way that the load is the same at any instant, as outlined in the waveform picture on the next page.

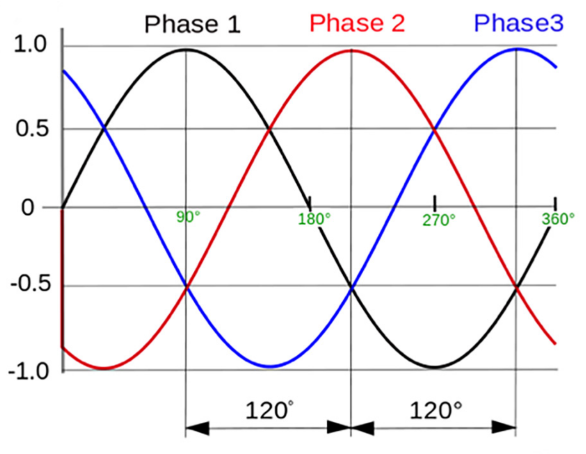


Figure 5-Three phase waveforms diagram

Because the load is constant, three-phase power is ideal for large motors and heavy loads—such as servers—as it eliminates the need for starting capacitors. Three-phase power distribution is also ideal because 120 volt, 208 volt, and three-phase circuits can all be pulled from the same four-wire system. In addition, it allows for fewer wires and thus less copper and lower voltages for the same power transmission as single-phase, making it a less expensive and safer option for the data center.

- 120 volt = single phase load + neutral/ground
- 208 volt = 2 x single phase loads + neutral/ground
- 208 volt 3-phase = 3 x single phase loads + neutral/ground

Those actively calculating load across three-phase circuits know that the equation requires a multiplier of 1.73. For example, the delivered load of a 30-amp, 208-volt, three-phase circuit is calculated: 30 x 208 x 1.73 = 10,795 watts. But where does this 1.73 multiplier even come from? The answer is that it's due to the voltage between the phases being 1.732 times the phase to neutral voltage. Its value can be derived from what is called "vector analysis," which boils down to: 2cos30 = 1.732. This simple equation, however, should make sense of things for you:

- 120 volt (single phase) x 1.73 = 208 volt (2-phase)
- 208 volt (2-phase) x 1.73 = 360 volt (3-phase)

» GROUNDING

Anyone who has been on a data center floor long enough will eventually hear the electricians or operations staff mention the "grounding wire(s)" running above or below the floor. Grounding is connecting electrical equipment and wiring systems to the earth by a wire or other conductor. The proper grounding of the electrical systems and circuits within the data center are absolutely essential—for both the lives of the humans and the hardware hard at work within the building.

Protection against lightning by lightning rods or other means is also a form of grounding.

As noted above, each circuit delivered to the data center floor and within your home also includes a neutral and/or ground within the circuit. If you've ever personally installed your water heater or watched someone work with a circuit-breaker panel, you likely learned quickly what role grounding plays in safely delivering power! The primary purpose of grounding is to reduce the risk of serious electric shock from current leaking into uninsulated metal parts of an appliance, power tool, or other electrical device. In a properly grounded system, any leaking current is carried away harmlessly.

Grounding is also used within the data center to prevent accumulation of hazardous static electricity charges, which can cause hard drives and other related components to fail for seemingly no reason. As a result, most cabinets that servers sit in are also tied to a separate grounding wire; this helps to absorb much of the excess current that may be leaking around the data center floor and to send it safely deep into the ground.

Almost all servers and network equipment deployed within the data center have the ability to receive power from two separate feeds. The reason for this is to provide power redundancy to the device should one of the power feeds drop. Data center providers also specifically design and deliver redundant paths of power distribution from the utility to the server to ensure uptime for the hardware on the data center floor. Unfortunately, the naming conventions used by data center providers to identify the delivery of two circuits of the same phase and amperage to achieve redundancy across the two feeds are often confusing and/or inaccurate. The most common names are listed below.

» PRIMARY AND REDUNDANT, A SIDE AND B SIDE, ACTIVE AND PASSIVE

The names above all suggest that one of the circuits is 100 percent live while the other is at 0 percent and on standby in case the primary circuit fails. This is simply ***not*** the case. If a single server has both of its power cords plugged into live power outlets, the load will be shared across both power plugs equally.

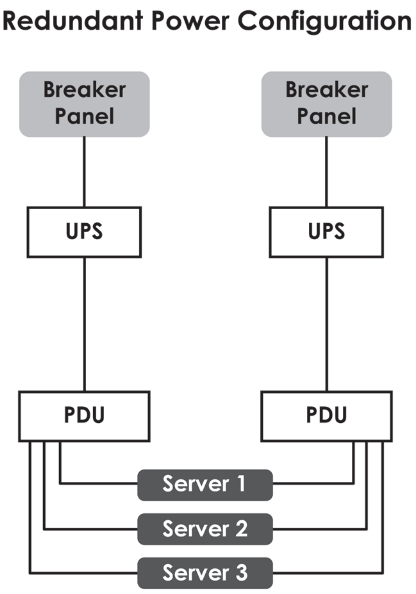


Figure 6-Redundant power configuration diagram

» EXAMPLE ONE: PRIMARY AND REDUNDANT

A single cabinet has a primary 20-amp, 120-volt circuit and redundant 20-amp, 120-volt circuit installed. The total load consumed by the servers within the cabinet amounts to 1.5 kW and each server has two power cords, one attached to a primary circuit and one attached to a redundant circuit. Within such a setup, the primary circuit will be pulling 750 watts; the redundant circuit pulls 750 watts as well. Should either of the two circuits fail and stop delivering power, the other will automatically start pulling 1.5 kW, the total load required from the servers within the cabinet.

» EXAMPLE TWO: A SIDE AND B SIDE

A single server is attached to two separate, live power outlets attached to different power circuits—A and B—and draw a total of 200 watts. Within such a setup, outlet A will be pulling 100 watts while outlet B will be pulling 100 watts. Should either of the two circuits fail and stop delivering power, the other will automatically start pulling 200 watts, the total load required from the server.

» EXAMPLE THREE: ACTIVE AND PASSIVE

A fully loaded blade chassis within a cabinet has one power cord attached to an active 30-amp, 208-volt, 3-phase circuit and one attached to a passive 30-amp, 208-volt, 3-phase circuit, drawing a total of 2 kW. Within such a setup, the active circuit and passive circuit will each be delivering 1 kW to the chassis. Should either of the two circuits fail and stop delivering power, the other will automatically start pulling 2 kW, the total load required from the server.

These three examples drive home the point that, regardless what a provider may call these circuits, they are each active and capable of delivering the entire load to the server should one fail. Assuming, of course, that the total load consumed by the servers attached to both power circuits does not exceed the load of any one circuit.

For example, if there are 2 x 20-amp, 120-volt circuits installed within a cabinet, the total delivered power within the cabinet is 2 x 20 x 120 = 4.8 kW. If the load consumed by the servers within a cabinet is 2.5 kW, however, and if one of the circuits fails, the entire cabinet will lose power; the 20-amp, 120-volt circuit that remains is only capable of delivering 2.4 kW. The circuit will overload and the breaker within the circuit will trip/pop, discontinuing the delivery of power through the circuit.

What makes this even more confusing is the reality that, within a circuit-based billing model, despite both circuits delivering power at an equal load, most providers will charge customers a reduced rate for the "redundant/b side/passive" circuit. Typically this is at around 50 percent the cost of the "primary/A side/active" circuit. This reduced cost leads many to believe that there is less power being delivered across the second circuit and that it is simply there as insurance to carry the full load of the other circuit should it fail, which is not the case.

There are a few important facts to walk away with here:

- Power is delivered across both circuits equally.
- Assuming a server is plugged into two different circuits, if power is lost from one circuit, the load will transfer seamlessly to the remaining live circuit.
- The reduced rate provided for the secondary circuit is contingent upon proper power management across both live circuits such that the power consumed by both does not exceed the delivered power of either.
- Many service providers will offer different uptime guarantees—"five nines" versus "three nines," for example—within their SLA should a client choose to leverage a secondary circuit instead of primary only.
- The majority of downtime accrued by clients within production-grade, commercial retail colocation facilities is due to an improper understanding and implementation of the principles outlined above at the hardware receptacle level. Circuit redundancy is thus rendered null/void when the aggregate load across dual circuits exceeds what an individual breaker can support during a failure of either the primary or redundant delivered load.

» COMMON PLUG TYPES

IEC 60320 C13/14

This plug type is what almost all standard server equipment deployed within the data center requires. The picture above shows the receptacle for such a plug type; there will generally be twelve to twenty-four such receptacles within any given in-rack horizontal or vertical PDU strip. C13 refers to the male side of the outlet, which would plug into the receptacle shown here. C14 refers to the female side of the outlet shown here.

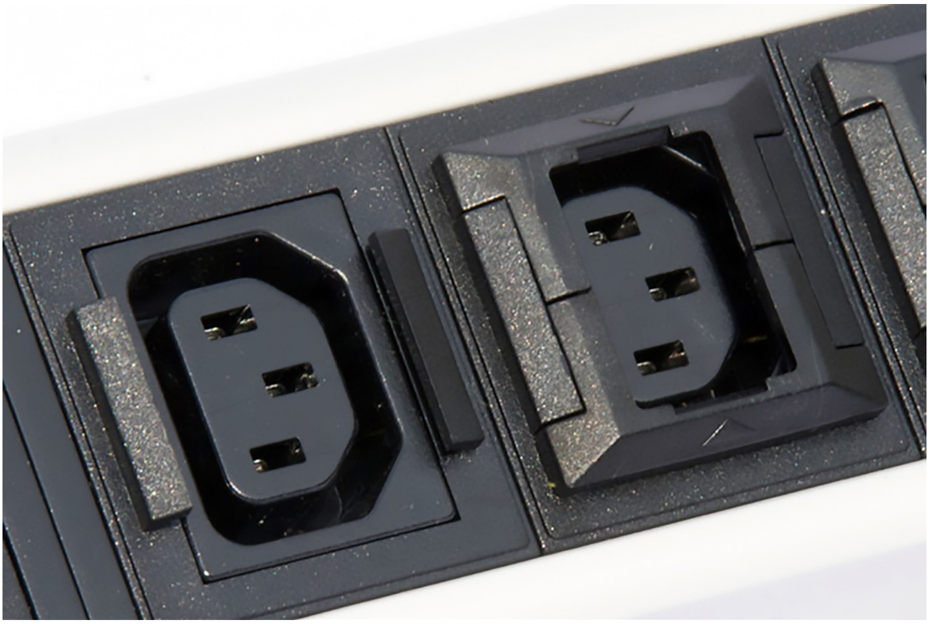


Figure 7-IEC 60320 C14 plug receptacle

There is a limit to the amps these power cords can support. Some of them only support 10 amps, others a bit more. Users should be aware of the amount of amps required to power a device.

L5-20 and L5-30

The L5 is specifically designated for 120-volt power. The "L" stands for locking and the "5" for 120 volts. Most data center companies utilize locking receptacles rather than nonlocking receptacles. The second number—20 or 30—refers to the specific amperage of the circuit.

L5-20 = 20-amp, 120-volt circuit

L5-30 = 30-amp, 120-volt circuit

Figure 8-L5 plug

L6-20 and L6-30

The L6 is specifically designated for 208-volt power delivered with the number preceding, which refers to the specific amperage of the circuit. The "L" again stands for locking and the "6" stands for 208 volts. The plugs themselves look almost identical to the L5 shown above, and can be easily confused. However, most plugs have a description of the plug written on the individual outlets to avoid confusion. The "P" and "R" following the 20 or 30 follows the same logic outlined above. A "P" plug is a male plug and an "R" plug is a female plug.

L6-20 = 20-amp, 208-volt circuit

L6-30 = 30-amp, 208-volt circuit

L21-20 and L21-30

The L21 is specifically designated for 208-volt, three-phase power delivered with the number preceding, which refers to the specific amperage of the circuit. As in all other circuits, the "P" and "R" following the 20 or 30 follows the same logic outlined above. Note that the five prongs displayed are tied to the three hot wires, one neutral, and one ground, delivered within all 208-volt, three-phase circuits.

L21-20 = 20 amp, 208-volt, 3-phase circuit
L21-30 = 30 amp, 208-volt, 3-phase circuit

Figure 9-L21 plug

» THE 80 PERCENT RULE

Most commercial retail colocation providers will contractually limit their clients from exceeding 80 percent of the delivered power available within a contracted circuit. By doing so, the provider retains the right—after sufficient warning—to turn off the power delivered on a specific circuit should it consistently exceed this 80 percent figure.

Those providers who do not make this a specific contractual stipulation will still more than likely set alarms and alerts that go off when circuits exceed this 80 percent figure. They will also notify the "offending" client when this becomes a chronic issue, and likely recommend they redesign their power configuration within the cabinet or add additional power circuits within the cabinet to carry some of the excess load.

Power density and forecasted cooling load calculations also take into consideration the net of the delivered load multiplied by 80 percent (= amps x volts x 80%). Despite the pervasive reliance the data center industry places on this 80 percent figure, very few understand where it came from or why it's used today.

Myth One

The 80 percent rule is an Institute of Electrical and Electronics Engineers (IEEE) standard and a best practice used in data centers around the world.

Myth Two

The circuit breaker will quickly overheat and pop/fail should the load on a circuit exceed 80 percent of the delivered load for an extended period of time.

Myth Three

Data center providers are simply looking to make more money by charging clients for the 20 percent capacity on the circuit that they are unable to use.

The Facts

The 80 percent requirement comes from the National Electric Code (NEC), which is published by the National Fire Protection Association. In essence, the 80 percent rule is an enforceable electrical code within the United States that is intended to protect those using dangerous levels of power from causing damage or harm to themselves or to the environment around them.

As a result, due to the complexity and relative magnitude of risk a data center presents within a community, the facility will almost always require at least one annual visit from the local fire marshal. Should the marshal see that there is a chronic issue with clients abiding by the 80 percent requirement or that the provider is not taking the issue seriously, he or she has the authority to pull the occupancy permit for the facility and require that the data center cease operations. One of the other reasons data center owners will include language enforcing the 80 percent threshold within their contracts is due to insurance requirements—or insurance rebates they can achieve for doing so.

This 80 percent figure varies around the world; the NEC code is specifically relevant only within the United States. For example, some commercial data center providers in Europe work off a 70 percent load threshold standard, while others allow customers to run as high as 90 percent. As a result, when procuring services outside the United States, you should always ask the data center provider what the threshold is on the circuits delivered within the facility. Should the rack design within an environment intend on using an 80 percent usable load but the provider works with a 70 percent limit, the entire intended architecture for the hardware and power may not work. Last but not least, it is also worth noting that you ***can*** technically exceed the 80 percent threshold on a circuit without causing any problems or issues with the circuit itself. In fact, some circuits will run at or slightly above 100 percent of the delivered capacity without failing/popping…though not for very long!

» AC AND DC POWER

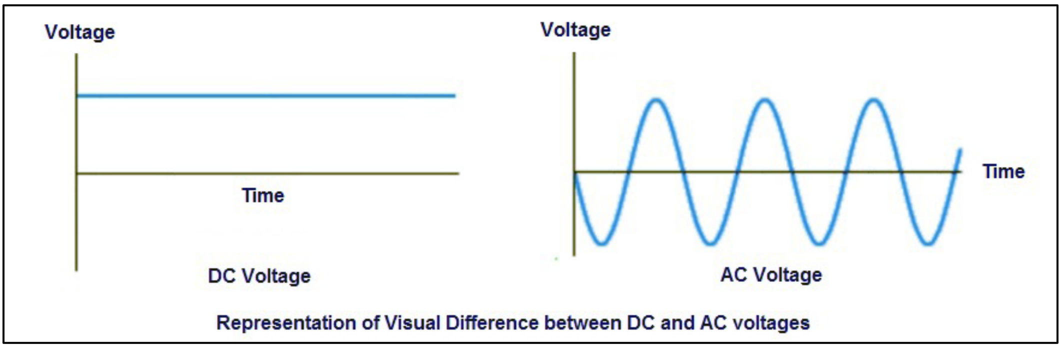


Figure 10-AC and DC voltage polarity diagram

Alternating Current (AC) and Direct Current (DC) are two different forms of voltage, or current, used for the transmission of electrical power. The main difference between the two has to do with the direction and flow of electrons within each system. An AC system transmits electrons in two opposite directions, alternating sixty times per second, which is visually represented as a sine wave. A DC system transmits electrons in one direction, which is visually represented as a constant straight line.

Utility companies traditionally transmit power using AC because it allows for efficient power transfer over higher voltages and longer distances, requiring less current and therefore less heat. To put this in perspective, as Watts = Amps x Volts, 1 million watts can be delivered far easier and with far less infrastructure via 1 amp at 1 million volts than 1 volt at 1 million amps. Before the power is distributed to individual clients, transformers lower the voltages so that 120-volt or 240-volt power is delivered from the wall sockets within your home or office.

It's interesting to note that most devices with batteries or fuel cells receive and transmit DC power. The big block attached to the wall adapter power outlet for your laptop is thus responsible for converting the AC power pulled from the wall socket to DC power before it reaches your device. The same AC-to-DC conversion is applied, just on a larger scale, within any data center mechanical room housing battery banks.

Every server, switch, and router in the data center requires DC power to operate as well. We don't hear very much about DC conversion in the data center, however, because most servers perform the AC-to-DC conversion within the device itself. To avoid the power loss that occurs any time power is converted from AC to DC or DC to AC, some servers are equipped with DC power supplies and can take the DC feed directly. This is specifically the purpose of a “DC plant” within a data center and, to this end, more and more manufacturers are offering DC power supplies as an option over the traditional, preinstalled AC power supplies.

» POWER USAGE EFFECTIVENESS

One of the primary goals of any data center colocation service provider is to meet the cooling needs of the equipment housed within the facility. A data center, therefore, in many respects, can be viewed simply as a giant refrigeration warehouse with the bulk of a provider’s ongoing maintenance costs tied to the power consumption needed to cool the infrastructure deployed within the facility.

A facility’s stated Power Usage Effectiveness (PUE) shows how much overhead is associated with the delivery of power to the end customer’s infrastructure within the facility. The equation takes the total amount of power consumed by the entire facility—which, in most cases, can be pulled directly from the monthly bill from the utility provider—and divides this by the actual power consumed by the IT load on the data center floor. This measurement includes all the internal operational components leveraged to run the data center, such as CRAC and CRAH units, chillers, UPSs, PDUs, security features, lighting, and monitoring equipment.

For example:

- A PUE of 1.2 would represent a 20 percent overhead, or 20 Watts of additional power required to deliver 100 Watts of usable power to the customer = 120 total Watts.
- A PUE of 1.7 would represent a 70 percent overhead, or 70 Watts of additional power required to deliver 100 Watts of usable power to the customer = 170 total Watts.

Thus, the more successful a provider is at driving down the total power required to deliver the needed cooling and related support infrastructure within the building, the lower their cost basis will be. Lower power requirements will also equate to more power for the provider to deliver and sell to its clients. In turn, lower power requirements can also increase the revenue potential for the building as well as the power density that can be delivered within each cabinet—if the building is designed to manage the additional power and power density delivered to the data center floor. The conversion of cost to revenue through a successful reduction of PUE is not a guarantee.

In general, it is safe to assume that older data centers and those built in hot climates such as Arizona, Nevada, and Texas will have higher PUEs, while newer data centers and those built in cool climates, such as those in Washington, Oregon, New York, and New Jersey, will have lower PUEs.

» PUE IN PRACTICE

Unfortunately, PUE has become far more of a sales and marketing tool than a useful engineering calculation based on hard facts or actual costs incurred by a facility.

For example:

- Some providers will state a theoretical PUE that is based on computer modeling that assumes the entire data center is completely full and every usable kW of capacity is consumed within the facility. Another assumption common for this kind of modeling, especially for commercial data center providers, is the actual power used within each cabinet across the entire data center floor. The initial actual PUE of the facility, therefore, may vary drastically from the theoretical PUE, if and when the entire facility is in production.

- Some providers, when calculating PUE, may only take into consideration the cost of cooling. They won't add in the additional power consumed by other lighting and non-IT-related infrastructure leveraged, or the power loss that occurs from the different transformers, or UPS systems leveraged between the IT load on the floor and the utility feed to the facility.

- Almost all providers who bill using a metered power method include this PUE figure in their calculations when billing every month. To this extent, some may simply include PUE as a "cooling and infrastructure uplift," though the calculation remains the same. Of these providers, only a few will do the actual calculations each month, making the figure variable over the term of your contract. In contracts, most now set a fixed PUE during the term of the contract as a way to compete on price with other facilities. To this extent, the PUE figure is not based on any real hard metrics, and is simply designed to lower the total contracted rate for the client, and to make the facility a more attractive option on paper.

Therefore, it is essential to not simply trust the sales or marketing materials from any provider presented in regards to PUE. Ask the hard questions to dig into the actual calculations and costs used when calculating this figure. Even a variable PUE figure offered should be inspected so you know exactly how it is being calculated each month. Knowing this information will both help you defend the PUE figure(s) your firm is presenting and help you clarify just how effective and efficient a provider is truly being at delivering power to the IT equipment on the data center floor. If you find the calculations presented do not align with the actual hard data you've been presented with, consider it an immediate red flag. It indicates that most of the other figures and data points you've been provided, or are providing, should also be inspected closely.

PUE	Per kW/hour	Cost per 1 kW	Cost per 10 kW	Cost per 100 kW
1.2	$0.05	$43.80	$438.00	$4,380.00
1.3	$0.05	$47.45	$474.50	$4,745.00
1.4	$0.05	$51.10	$511.00	$5,110.00
1.5	$0.05	$54.75	$547.50	$5,475.00
1.6	$0.05	$58.40	$584.00	$5,840.00
1.7	$0.05	$62.05	$620.50	$6,205.00
1.8	$0.05	$65.70	$657.00	$6,570.00

Figure 11-Effects of PUE on cost per kW

Chapter Summary

- kVA vs kW: For our purposes, kVA is "real power" and kW is "apparent power." In AC circuits, not all the wattage makes it to the target device. The wattage that does is the kVA.
- kVAR: The difference in wattage between kW and kVA.
- 120 volt versus 208 volt: Remember how watts = volts x amps? Well, when you add volts, you increase wattage. 208 is used for larger appliances. It's also more dangerous, and extra watts cost extra money.
- Single-phase versus three-phase power: Only applies to AC circuits. In single-phase, the electric load constantly reverses polarity—think flow of water in a pipe. Three-phase combines three single-phase circuits that all reverse flows at different times, so the target device always experiences the same direction of flow.
- Grounding: Wiring that connects load-bearing circuits to the earth. This wiring dissipates potentially dangerous build-ups of electricity.
- Primary and redundant circuitry: Central to uptime. Two independent power supplies that deliver power to the data center at the same time.
- Common plug types: There are common plug types. Flip back to an earlier page and know them.
- The 80 percent rule: NEC requirement meant to protect humans from dangerous levels of power (electrical power—not political, social, relational, financial, etc.)
- PUE = Total facility wattage/usable wattage delivered to clients

CHAPTER SEVEN: POWER DISTRIBUTION SYSTEMS

After studying this chapter, you should be able to:

- Describe the key pieces of infrastructure that keep a data center supplied with redundant power.
- Describe how these systems work together to ensure maximum uptime.
- Evaluate a specific data center and describe how that facility's systems determine its reliability.

» UTILITY SUBSTATION

Each utility provider across the country maintains and manages substations that deliver power to the respective residential and commercial clients they serve. Each substation feeds back to a major power distribution center from the utility or larger substation. Some are also dedicated to specific government buildings, hospitals, and related critical infrastructure services, and thus a data center fed by the same substation that also feeds these "critical" buildings and services has the added benefit of receiving special priority should a major outage occur.

Figure 12-Utility sub-station

Some data centers can be fed from two separate substations, adding further redundancy to their power distribution capabilities, such that if either substation were to have an outage, the facility would still be supported by the other. Should a provider claim that they are, in fact, supported by dual substations, it is important to find out exactly how the facility is leveraging the power received. This is because a facility may in fact need the power from both substations to maintain the critical load across the facility. An outage at one of the substations would thus take down the entire facility, or most of the facility, as a result. Hence, the apparent benefits of the dual substation feed are not necessarily as certain as they may seem.

» AUTOMATIC TRANSFER SWITCH

The ATS is one of the most critical pieces of infrastructure installed within any data center promising any degree of redundancy or uptime. The ATS both receives power from multiple sources—such as generators, battery banks, and the utility provider directly—and is responsible for delivering uninterrupted critical load to the infrastructure within the data center regardless of where the power source is coming from.

Should any one source of power fail, it must be able to instantly transition to a more reliable source and ensure the load does not drop or change drastically enough to affect the infrastructure down the line. Due to the complexity of this piece of equipment, most outages within the data center occur during the testing of this piece of equipment or when the ATS is actually called upon to transition from utility to generator power.

» BATTERY BACKUP AND UNINTERRUPTABLE POWER SUPPLY

Generators take time to turn on, heat up, and start generating electricity. Though this usually only amounts to a few seconds, a backup piece of equipment must carry the load between when utility power stops and the generator kicks in or the entire facility will go dark. The most common power sources during these transition periods are large industrial battery backup UPS systems.

In most cases, the power delivered to the data center floor is constantly flowing through the UPS. This serves two primary purposes:

1. To clean and condition the power after the electrons have made their way through the transformers, ensuring a constant sine wave is delivered to the servers on the data center floor.

2. To ensure the batteries within the system can instantly carry the load delivered to the data center floor before the generators can support the load.

These systems are fairly complex and comprised of large industrial batteries that are stacked high and wide and linked together. The time it takes for the generators to kick on is only a few seconds at most. If a facility is at full capacity, the run time of these battery UPS systems is generally only thirty seconds to a few minutes at most.

Because of this short run time, should the generators not kick on, the engineers at the facility have very little time and very little room for error—if any—to rectify the situation before the load carried in the batteries drains and the facility goes dark. You should pay close attention to the maintenance schedule and redundancy of UPS systems, as well as the run time of the batteries themselves relative to the power load supported by the facility.

» BATTERY MONITORING SYSTEM

It is extremely costly for data center providers to purchase and install entire banks of industrial-grade batteries, which typically need replacing every four to five years. The maintenance of these battery systems can also get expensive, as one "bad" battery within the system will infect and deteriorate the health of the batteries around it if it is not replaced quickly. If unchecked, this one battery can affect the entire system, causing premature aging and replacements.

As a result, systems have emerged that focus on monitoring and reporting the health of each individual battery within the system. Data center providers who install these systems can avoid the need for mandatory costly scheduled replacements every four to five years and focus their time and money on the individual batteries that are in need of replacing instead. Though the monitoring systems themselves are not cheap to install initially, they are far cheaper than having to replace an entire bank of batteries. As such, they have become increasingly popular; it is now rare to find a commercial data center operator or large private data center facility that has not invested in some form of battery monitoring solution.

» TRANSFORMERS

Transformers take an input AC current and "transform" it to a higher or lower voltage. This is extremely important, because utility companies send power to data centers at extremely high voltages that must eventually be converted to 120-, 208-, and 360-volt power feeds on the data center floor.

For example, 1 million watts delivered over 1 million volts at 1 amp (1,000,000 watts = 1,000,000 volts x 1 amp) could be converted within a transformer to 1 million watts delivered over 120 volts at 8,333 amps (1,000,000 watts = 120 volts x 8,333 amps). This equation is not perfect, however, as the conversions result in varying degrees of power loss due to the age of the equipment, heat generation, and the variance of the conversion.

» RECTIFIER AND INVERTER

Rectifiers convert AC power to DC power and are essential for data centers leveraging battery backup UPS systems and those pushing DC power directly to a cabinet, cage, or suite. Within nearly every server, switch, and router is a mini-rectifier converting the power from AC to DC for the device to consume. The little box connected to the power supply for your laptop is an example of a mini-rectifier that converts the AC power delivered from the outlet to DC.

Inverters do the opposite; they convert DC power to AC power. As such, they are also essential for data centers leveraging battery backup UPS systems when the power delivered to the data center floor must be AC.

» STANDARD GENERATOR

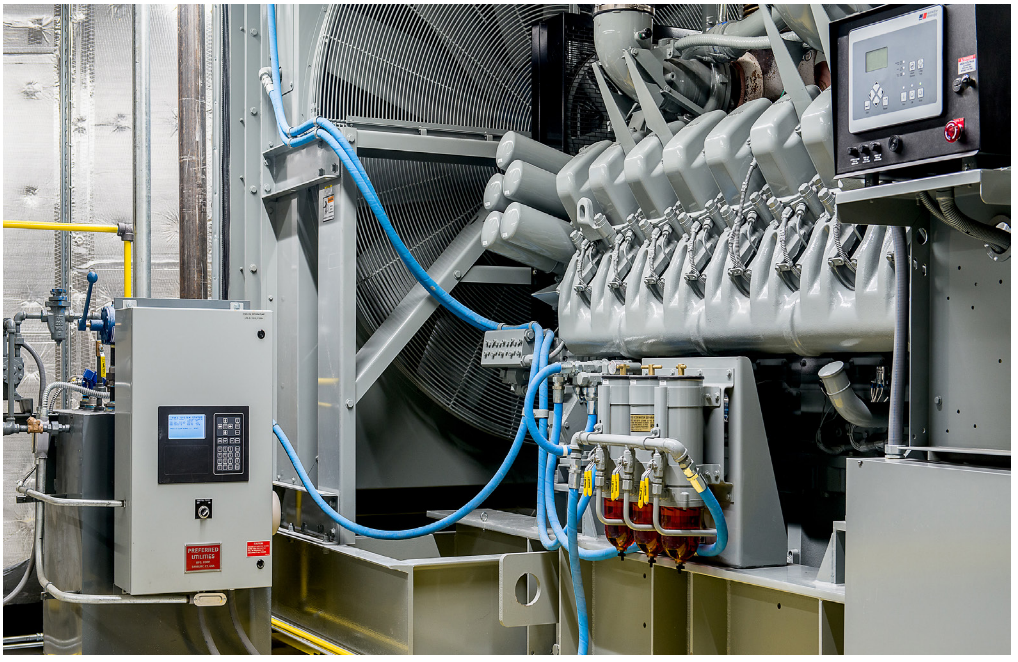


Figure 13-Diesel generator, image courtesy of ViaWest

The generator is a critical piece of equipment for nearly every data center, as it provides power when the utility provider has an outage.

Though there are numerous makes, models, and manufacturers, generators large enough to support a commercial data center generally run on diesel fuel and kick in within a few seconds of the facility losing power from the utility. Key variables to understand with these pieces of infrastructure include:

- Service contracts
- Support contracts
- Maintenance schedules
- Run time without refueling
- Refueling contracts
- Types: Uptime Tier III or Tier IV continuous-run generators are, for example, required versus a prime/standby generator.

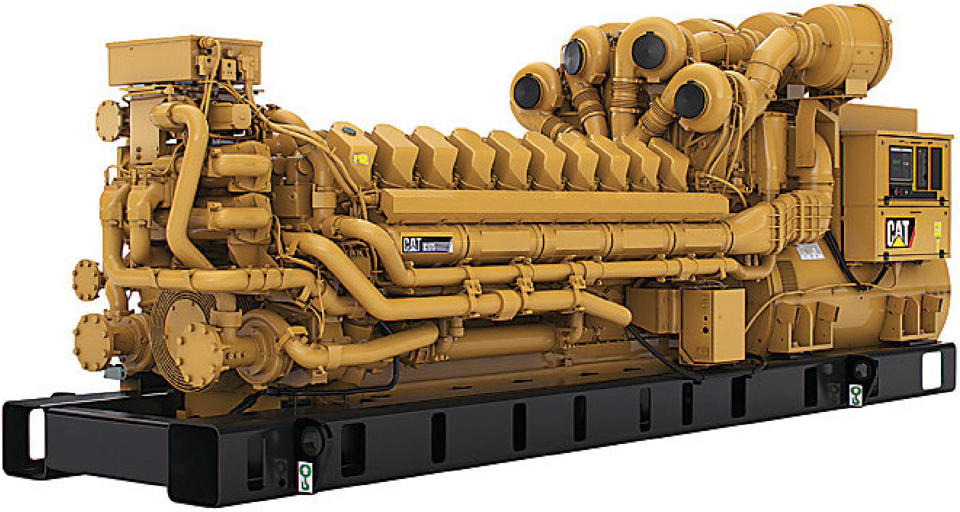


Figure 14-Diesel generator

» DIESEL ROTARY FLYWHEEL UPS

This specific type of generator is designed to replace the need for the standard UPS systems that carry the load of a facility until the generators kick on. It combines the UPS and generator into one unit, with the batteries of the UPS replaced by what amounts to large and heavy power couplings that are always spinning and kick in as the source of electricity until the generator starts. These are fairly complex machines, and are mentioned here due to the popularity of this particular system among existing data center providers.

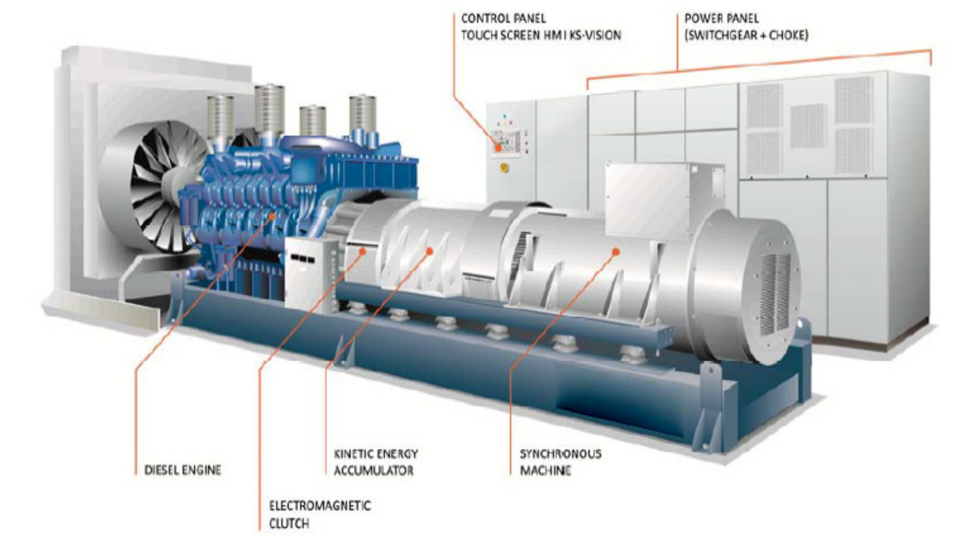


Figure 15-Diesel rotary flywheel UPS

Drawbacks of this type of unit:

- The power couplings only provide a limited amount of time for the generator to turn on, leaving little room for error should the generator not function as expected.
- There are a limited number of units in production relative to the standard UPS/generator configuration, so experienced technicians are harder to find and more expensive.
- Because the device has a number of parts that are constantly on and moving, it tends to require more maintenance than a traditional UPS system.

» RPP AND PDU

The terms Remote Power Panel (RPP) and Power Distribution Unit (PDU) are both used to describe a device that effectively serves as a large circuit-breaker panel. These panels tie back to a UPS. Within these distribution panels, specific circuit types (20 amp, 30 amp, 120-volt, 208-volt, three-phase, etc.) are carved off and pulled to individual cabinets on the data center floor by an electrician. PDUs normally have a standard capacity of power they support. They can do a mixture of power whips, but they have a total capacity limitation.

Figure 16-RPP and PDU

» POWER BUS-LINE SYSTEM

This power delivery system was developed specifically to reduce the number of individual copper lines pulled from any given network of RPP/PDU systems and to allow for quicker circuit provisioning. Effectively, the power bus-line system is an above-rack, preinstalled power panel capable of delivering a variety of voltage and amperage circuit types. You simply need to attach a single circuit box directly above the cabinet to the “bus-line” without needing to install the cabling between the cabinet and the RPP/PDU.

This system is specifically popular with clients who have large footprints and who know ahead of time exactly what their cabinet layout and topology will be. Although the initial upfront cost to deploy the system can be expensive, the total cost of ownership is in fact cheaper than having to pay for the labor and materials for multiple circuit runs.

This solution is almost never found in managed commercial retail colocation facilities due to needed flexibility and the variety in cabinet and cage designs across the data center floor. The upfront cost to install the system also serves as a barrier for most providers, who are hoping to keep their upfront investment as light as possible.

Electricians perform the bulk of the work performed during the "turn-up" phase of a new install. Electricians install and pull the copper cables required to create the primary or primary/redundant configurations of circuits and feed those cables from different distribution panels on the floor.

» IN-CABINET PDU

Though there are some cases where a specific piece of hardware will plug directly into the power circuit delivered to a client in a cabinet, almost all other hardware will plug directly into what amounts to a very large power strip, no different than the one likely under your desk at work or home.

Figure 17- "Smart plug" power strips in cabinet

Some even afford the client the ability to log in remotely to the specific power strip itself to view how much power is being drawn within each individual outlet on the device.

These same devices, known as "smart strips," also give clients the ability to turn individual outlets on and off. This means you can remotely reboot a device without ever having to contact the data center provider—or even visit the site itself.

There are three types of cabinet-level PDUs; basic, monitor, and management.

Basics provide power but do not provide basic information via a web interface or SNMP. Monitor PDUs allows users to monitor the power utilization and normally have a LED readout on them so power readings are available locally if someone is in the cabinet. Managed PDUs even go a step further and allow technicians to powercycle an outlet—reboot a device—without being local to the device. Options in outlet types, power connection type, and management abilities give technicians the ability to manage to the level they require or can afford.

Chapter Summary

- Utility substation: Delivers power from a utility provider's main power grid to local clients.
- Automatic transfer switch: Instantly switches power source from a source that has failed to a redundant source.
- Battery backup and uninterruptable power supply system: Conditions power to ensure consistency for servers and supplies power to bridge gap between potential utilities failure and generator ignition.
- Battery monitoring system: Monitors health of individual batteries within a UPS.
- Transformer: Changes the voltage of an alternating current.
- Rectifier versus inverter: Rectifiers convert AC to DC; inverters do the opposite.
- Generator: Typically a combustion engine that generates electricity. A backup system.
- Remote power panel and power distribution unit: The circuit breakers of data centers.
- Power bus-line system:
- In-cabinet PDU: A really big power strip/surge protector.

CHAPTER EIGHT: POWER DELIVERY AND REDUNDANCY

After studying this chapter, you should be able to:

- Define what the "N" jargon actually means—and what it doesn't mean.
- Describe why redundancy matters to clients.
- Communicate to a client the redundancy capabilities of a specific facility.

The symbols N, N+1, N+2, 2N, and 2N+1 are commonly used to describe the level of redundancy built into the architecture for the power and cooling infrastructure within a facility. The definitions below provide a quick definition so you can understand what each of these actually mean for power delivery; for our purposes, "outage" is defined as a loss of power to the IT load on the data center floor:

» N

Power delivered along a single path from the utility to the end user. Should any piece of infrastructure fail within this path, the entire path will fail and cause an outage.

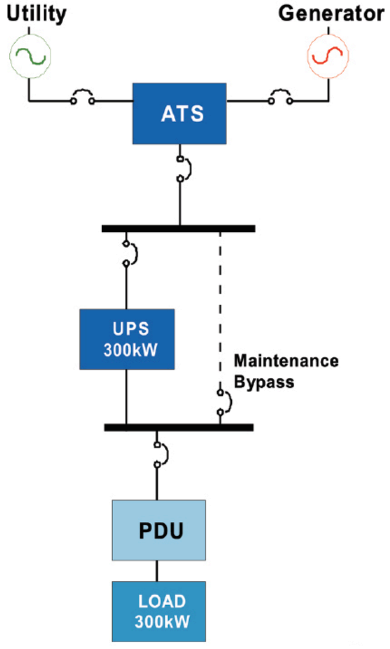


Figure 18-capacity diagram

» N+1

This design requires two failures of the same equipment within the same segment during the same timeframe to cause an outage.

» N+2

Same as N+1, but with two—not just one—extra pieces of critical infrastructure added to each segment. This design requires three failures of the same equipment within the same segment during the same time frame to cause an outage.

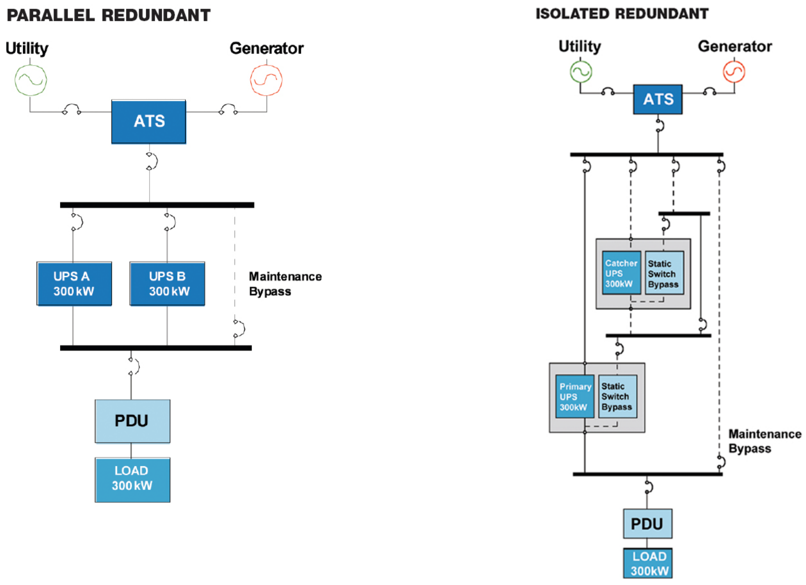


Figure 19-Parallel and Isolated redundant capacity diagram

» 2N

Power delivered from the utility to the end user along two separate and diverse paths.

If the end user is using both paths in a primary/redundant fashion, any failure along one path will not cause an outage. This design requires any one failure within any segment on both paths for the end user to experience an outage.

» 2N+1

Diverse paths with one extra piece of critical infrastructure added to each segment along both paths. If the end user is using both paths within a primary/redundant fashion, any one failure on both paths will not cause an outage.

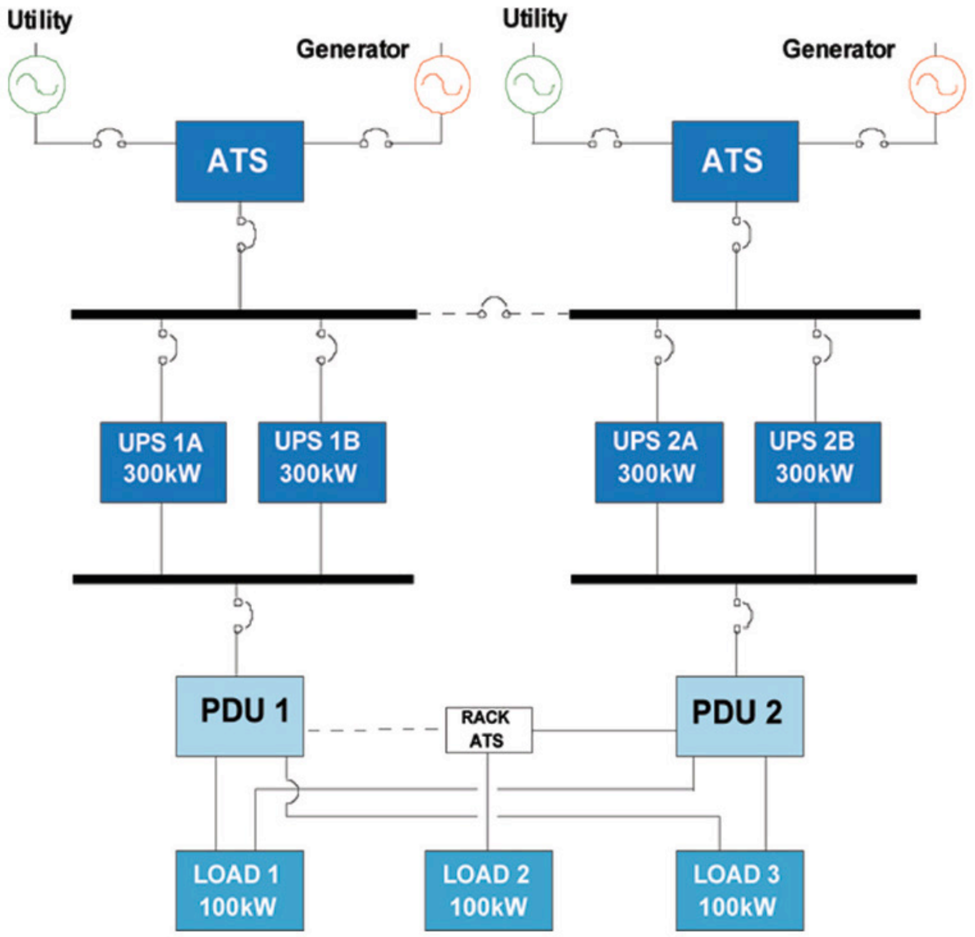


Figure 20-2N capacity diagram

» WHY REDUNDANCY MATTERS

There are a few reasons why redundancy is important. To the customer, there is a cost for downtime. If the cost of downtime for a twenty-four-hour period is $5,000 or less, they may not look at a high level of redundancy being worth the cost. But, if the cost of downtime is $1,000,000 for two hours then yes, the additional redundancy is worth the cost. In the business it is considered the cost of "insurance." If a customer is looking at a disaster recovery site versus a production site, then the redundancy may not be that important. For that same customer, the production site's redundancy may be very critical. Why does it matter? Because the cost to deliver these levels of redundancy drastically increases as you move from N to 2N+1 and because more physical pieces of infrastructure and links within the infrastructure are required to be deployed, maintained, and managed. There is also a higher cost for maintenance and electricity use will be higher due to the additional components.

The SLA and committed uptime a data center provider will offer will be determined by the level of redundancy designed and managed or actually contracted and delivered to a client. A facility may have a 2N+1 design in place; however, if a client chooses to only accept a primary feed from one available path, the uptime guarantee will be less reliable than if it accepted feeds from both paths available. More often than not, a provider will reduce the price of power as well if a lower level of redundancy is requested and can be delivered.

To understand exactly how power is delivered from the utility to a cabinet on the floor of a given facility, ask to see the electrical "one-line" drawings used by those who manage the physical infrastructure itself.

If uptime is critical to your company and its applications, learning how to read these drawings will give you the ability to instantly see if the redundancy of the facility is in fact what the provider is claiming.

Unfortunately, due to a lack of knowledge as to what these terms actually mean, or in some cases due to deceptive sales tactics, some providers will claim the redundancy offered is much higher than what has actually been designed and/or delivered to customers. For example, a company may claim its facility has a 2N+1 design when only one specific segment of the power or cooling delivery mechanism is designed with that type of redundancy. This is why buyers should look for specific certifications such as Uptime, TIA-942, or BICSI-002. Having a third party review the design and do a site visit will give potential customers a feeling of confidence in the redundancy and resiliency of a facility.

Chapter Summary

- N redundancy: Power on a single path. If one element fails, it all fails. Think Christmas lights. Or don't, if you have traumatic memories.
- N + 1 redundancy: Each element of key infrastructure is backed up by a redundant element, but power is supplied from only one unique source.
- N + 2 redundancy: Each element of key infrastructure is backed up by TWO redundant elements, but power is supplied from only one unique source.
- 2N: Two independent/diverse power supply paths.
- 2N + 1: Same as 2N, with redundant pieces of critical infrastructure added along both paths.

CHAPTER NINE: COOLING

After studying this chapter, you should:

- Be able to describe the key elements in a data center's cooling system.
- Be able to accurately describe how these elements work together.
- Be able to communicate a facility's cooling strengths or limitations to a client.

A data center effectively requires the same infrastructure as a large refrigeration warehouse to cool/remove the heat generated from the thousands—if not millions—of servers housed within. There are dozens of cooling designs in use today, and even more being designed and deployed as the industry continues to evolve and grow. One of the most essential and basic concepts to remember when reading this chapter is this: cold air is denser than hot air. In essence, this translates to the fact that hot air rises and cold air falls. Most facilities engineers estimate that one ton of cooling infrastructure can adequately cool 3.5 kW of compute.

Chemistry Corner: What is heat? Heat is how energy expresses itself as motion. Quite literally, as a substance gains energy, its molecules increase in speed of motion at the atomic level and spread farther apart from each another. That is why as energy is added to a substance, it can change physical states: think of ice changing from a solid to a liquid (water) to a gas (steam). Then there is also plasma, such as lighting. Every single element in the periodic table, which is what everything is made of, is technically capable of existing in these four different states. When something is a solid, think of it as having stored energy.

It is also important to know that in physics, nothing "sucks" and everything "blows." What this means is that when a high-pressure/energy system comes into contact with a low-pressure/energy system, the high-pressure system will expand, or blow into the low-pressure system until both systems are equalized at a middle pressure/energy. In this way, motion is always traveling in the direction of the low pressure system…or blowing.

The human body is a great example of this equalization action. Your body temperature—in effect, your energy—is constantly trying to equalize with your environment. When you are cold, it's because your body is warmer than the surrounding air, and so your body gives off energy to that air, reducing your body temperature and increasing the air temperature. As that happens, your body loses energy and your molecules slow down. The now-heated air touching your body has absorbed more energy, and its molecules are now moving faster than the surrounding cooler air is, becoming less dense (because there are fewer molecules in the same physical space) and rising above the denser air (where there are more molecules per the same physical space.)

When you sense that an area is hot, what you are really feeling is the surrounding air putting energy into your body—again, trying to equalize. As the heated air cools, by equalizing energy between your skin and itself, it becomes denser and sinks below the lower density more energized, or heated, air surrounding it.

Due to these core principles of physics in our universe, we have discovered different ways to equalize systems for our own benefit—namely the below "cooling" or "energy removal" systems in data centers.

» COMPUTER ROOM AIR CONDITIONER MECHANICAL COOLING

As the name computer room air conditioner (CRAC) implies, we are talking about ***very*** large air-conditioning units designed to relocate a ***massive*** amount of energy/hot air from a data center and to deliver a massive amount of low energy/cold air back into the data center. Cooling is accomplished by blowing air over a cooling coil filled with refrigerant. CRAC units are self contained, and so allow for single unit failures without impacting other CRACs in a facility. This allows for easy capacity increases, but lacks the efficiency of scale that many larger, unified cooling systems (CRAHs) have. PUEs for data centers using CRACs are typically higher.

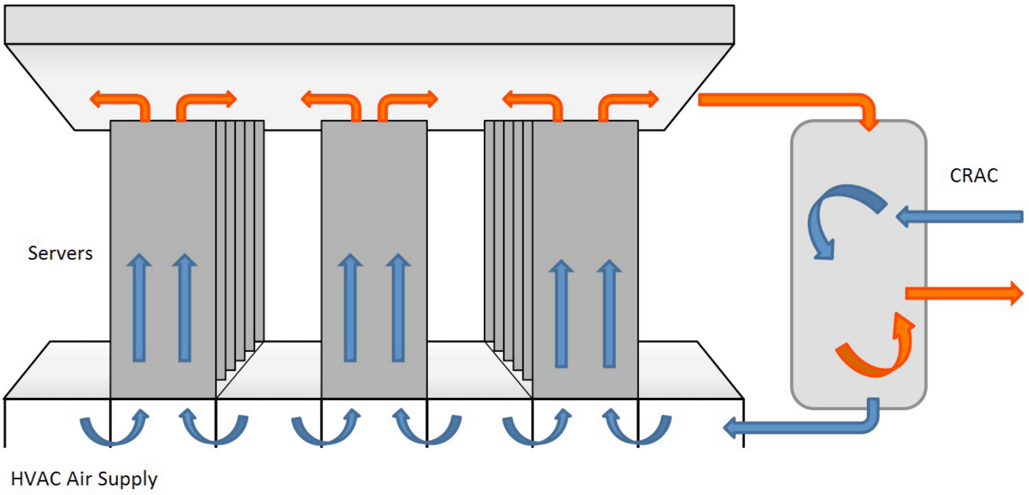


Figure 21-CRAC

CRACs use a liquid/gas refrigerant, and force it to change states (or energy storage capacity) between liquid and gas. In order for a lower energy state to become a higher energy state, energy must be added. The reverse is also true: in order for something to move to a lower energy state, energy must be removed. Using electricity and compressors, CRACs mechanically compress refrigerant into its liquid state (they lower the liquid's energy). Then that liquid is exposed via a "heat exchanger"—think of your car's radiator—to a data center's hot air.

That hot air is able to send its energy into the liquid, allowing that liquid to once again expand and become a gas, and removing energy/heat from the air itself. The gas is then moved away from the data center—into the outside world—where it loses energy by mingling with the surrounding air and recondenses via a compressor. This cycle is one of the best ways to superchill an area, but it requires lot of additional energy to compress the gas mechanically and to force the energy exchange.

That is why data centers that use this technology have a higher PUE than data centers using the other methods described below. For redundancy, many facilities add one additional unit for every four or five primary cooling units.

Broadly speaking, this type of cooling is one of the lower-impact types of cooling for computer equipment. All the air in a data center is recirculated, and nothing is being added to it, such as moisture or debris. However, with the ability to supercool, aka make ice, it is possible to condense gas vapor from the air into a liquid state at the CRAC unit itself. So it is important that a data center utilizing CRACs has a leak-detection system under the units to catch any liquid being released from the cooling method. This is often overlooked, and is why a humidity control feature should be considered as well. CRACs also have a "zone of influence," that is, a CRACs effectiveness decreases relative to a certain volume of space. The CRAC's effectiveness increases proportionally with the size of the CRAC unit.

» COMPUTER ROOM AIR HANDLER

Computer Room Air Handler (CRAH) units serve the same purpose as CRAC units. However, CRAHs use fans and a cooling coil only. They often use a chilled water loop from a larger unified chilling system—using the same compressor and refrigerant systems found in CRACs, but with better economies of scale—removing heat from the data center floor and placing it outside the building.

Figure 22-CRAH

CRAH systems aren't often configured to superchill a facility, rather they help to remove some of the issues with CRACs/condensation within the actual data hall. However, they do require larger distribution networks, which may mean large water pipes have to run through a data center. This of course poses a risk for water spills if the pipes were to break or leak. This is another reason to include leak-detection systems in data centers. CRAH systems utilize cold water, allowing multiple types of cooling methods to chill that water loop, which is physically separated from the data center hall. That is why CRAHs are known as indirect systems: the energy removal is not done directly on the data center floor. CRAHs, like CRACs, have a zone of influence that decreases as room volume increases. The same 1:4 redundancy standard applies for both CRACs and CRAHs.

Also note that some units can reduce air temperature to lower degrees than others. One design may only reduce air temperature by eighteen degrees, while another may reduce air temperature by thirty degrees. Hotter outside temperatures require more powerful units.

» RAISED FLOORS

The picture below shows a raised-floor environment where cooling is delivered via CRAC/CRAH units—along with power and networking, in some cases—directly under the data center floor. Cool air is thus forced up from under the floor through perforated tiles that are placed in front of each cabinet or row of cabinets.

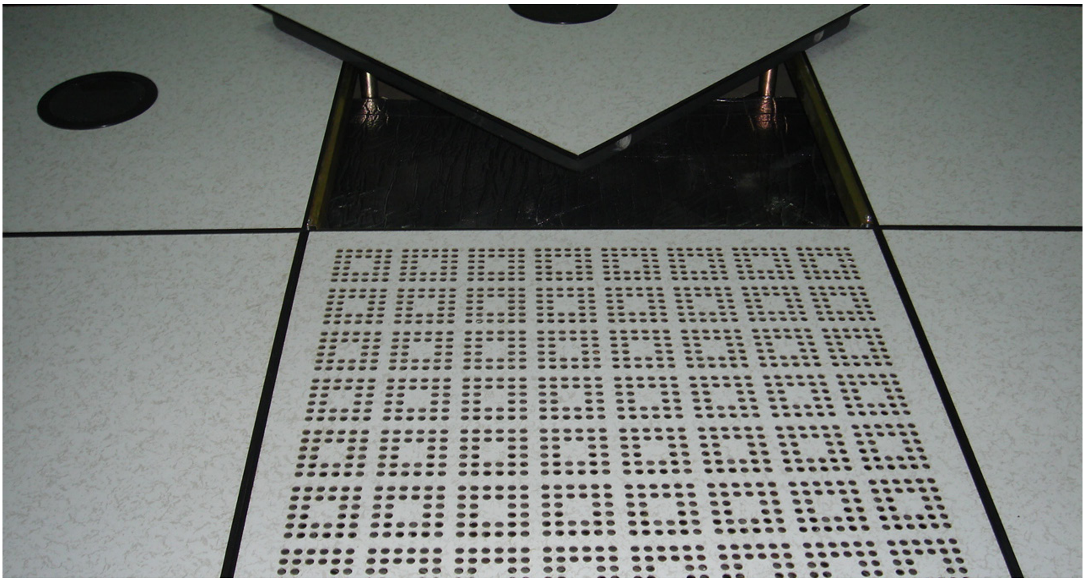


Figure 23-Photograph of raised floor

If the perforated tiles lie too close to a cooling unit, however, those tiles may experience a sort of Venturi effect. This is when air pressure of a different temperature above the tile forces the tile under the raised floor. The physics at work here resemble an undertow on a beach. Due to the proximity of the CRAC or CRAH unit, air speed above the tile increases above-floor pressure and reverses the flow of air through the perforated tile.

» SLAB FLOORS

Some data centers are designed with concrete slab floors, where cooling is forced down from above and along the front of each cabinet. Some research has indicated that in high-density environments—more than 250 watts per square foot—a slab-floor environment is capable of delivering more cooling to the cabinets than a raised-floor environment.

Figure 24-Photograph of slab floor, image courtesy of QTS

Most research indicates that there is no real demonstrable cost or power savings to be had between one design or the other; however, you may find some facility managers who swear that one design is far superior to the other. This is probably due to the fact that different facility managers have experienced different optimal designs in different environments. However, no research indicates conclusively that either design outperforms the other.

Figure 25-Photograph of hot and cold aisle containment

» COLD AND HOT AISLE CONTAINMENT

Despite the raised floor versus slab floor debate, there is no question that the most efficient way to deliver cooling to server cabinets is through some form of hot or cold aisle containment.

The type of system discussed above ensures the cooling payload delivered by the CRAC and CRAH units reaches the intended aisle, where servers are “sucking in” air and no heat from outside the environment is introduced. This lowers PUEs, as dead space that is not in use—as on the left side of the above picture—is neither cooled nor heated, and hence taken out of the temperature control calculation for the data center. If you have less air to condition, your workload is reduced, which provides lower PUEs.

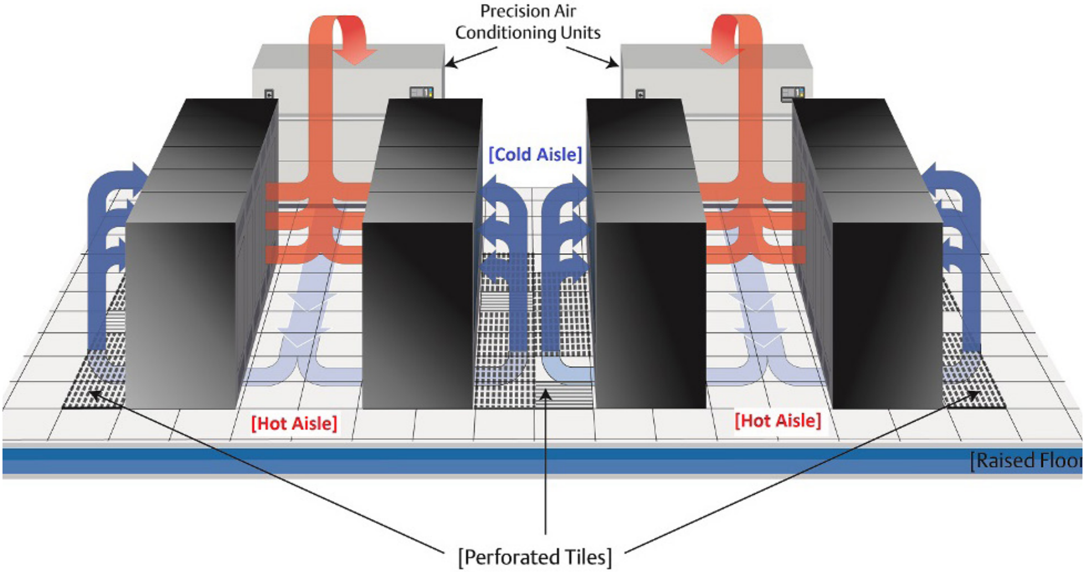


Figure 26-Hot and cold aisle diagram

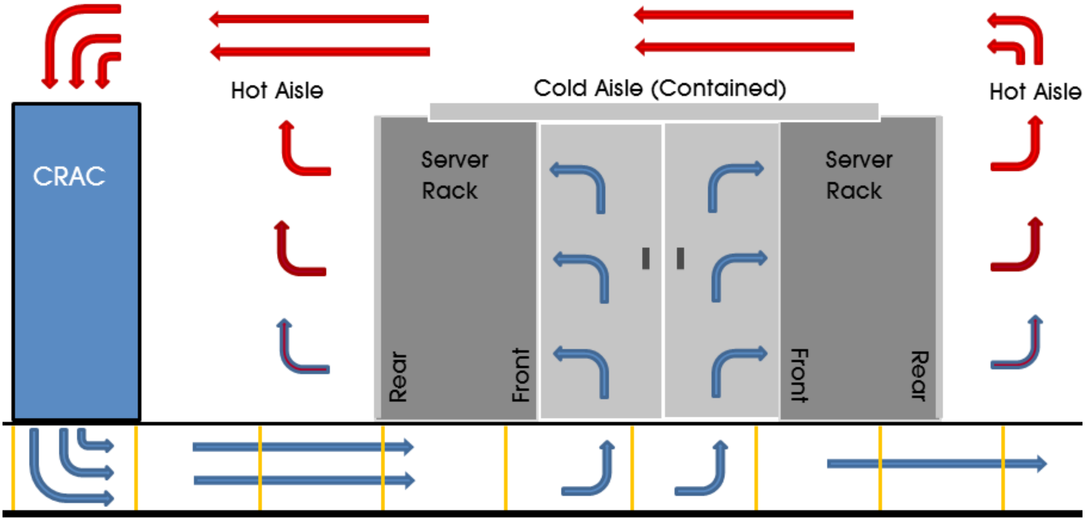


Figure 27-Cold aisle containment diagram

Such a system, once optimized, can also afford the facility the ability to deliver a higher power density than a facility or environment without a containment system in place.

» AIR-SIDE ECONOMIZATION

An air-side economizer brings outside air into a building and distributes it to the data center floor. Instead of being recirculated and cooled, the exhaust air from the servers is simply directed outside. This is also known as "outside air cooling" or "ambient air cooling." This method works most often in conjunction with normal chillers to bring air from outside the building into the building. Newer data centers built in regions that maintain an average air temperature between fifty and seventy-five degrees tend to use ambient air for the majority of their cooling needs. The primary purpose for leveraging this method of cooling is to save money, as significant savings can be achieved when a facility does not need to artificially cool the air within the building with traditional mechanical air conditioners.

If the outside air is particularly cold, the economizer may mix the outside air with the exhaust air. As a result, the exhaust air's temperature and humidity will fall within the desired range for the equipment.

The air-side economizer is integrated into a central air-handling system with ducting for both intake and exhaust. Its filters reduce the amount of particulate matter, or contaminants, that are brought into the data center as well.

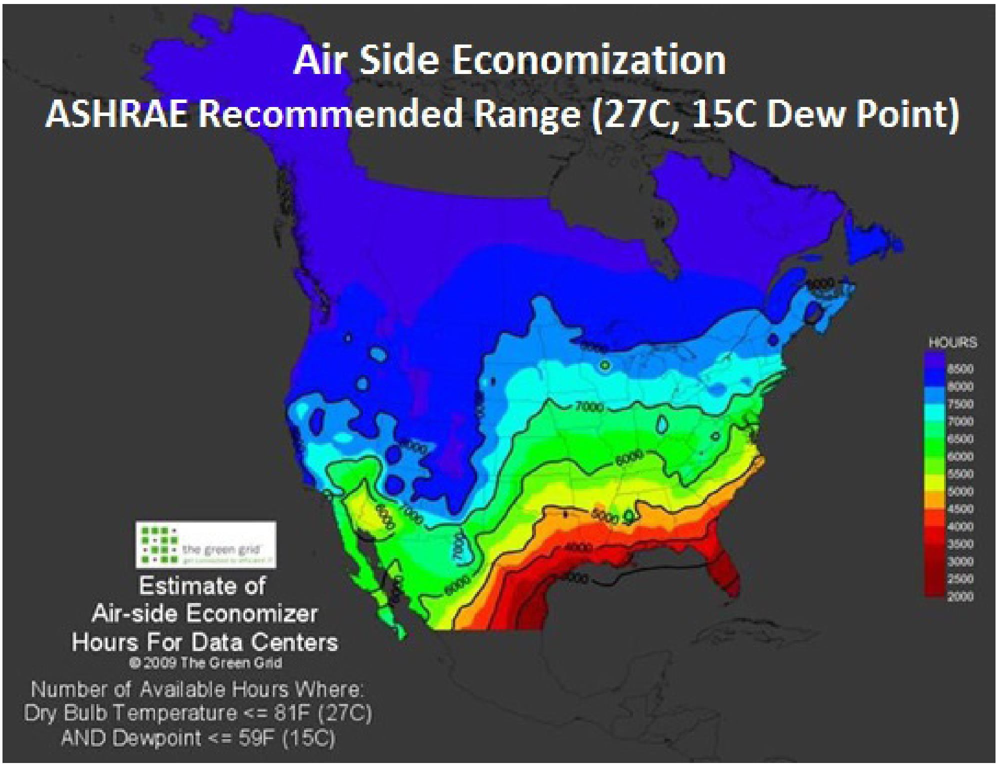


Figure 28-Air side economization map

One thing to note about these systems is that they do bring in outside debris—be it in a physical state or a gas state. As the air is brought into the data center, it is filtered, but it is not controlled like a contained system is. So depending on what is in the surrounding air, this may lead to condensation of undesirables within the data center at the smallest levels and potentially causing "misery" issues over time whose sources cannot be easily tracked down or identified. Filters and humidity control measures are critical for long server life in these instances.

» WATER-SIDE ECONOMIZATION/DIRECT WATER EVAPORATIVES

This method of cooling uses water instead of refrigerant to remove energy when changing states from a liquid to a gas. Commonly known as a "swamp cooler," this is when water is poured over an exposed water-to-air coil system, creating a large surface area. Water naturally evaporates in all states—for our purposes, it will turn into a gas, and will take energy/heat with it. This leaves the air and remaining liquid water colder/lower energy after passing through the system. That cooled air is then passed directly into the data hall, with the water vapor in tow. This is the biggest issue with this method, because the heated water vapor will seek out a piece of metal colder than it is—such as your server—and will attempt to equalize energy states with it. When energy is removed from the vapor, it condenses back into a liquid—on your server. This is why humidity control is very difficult with direct evaporative systems.

Direct water evaporatives require remarkably low energy to run. It only relies on a water pump—provided by public water pressure—and an electric fan—to force air movement. Large, single-tenant data centers explore this type of method when used at a very large scale. Companies investigating this system must weigh the costs of additional hardware failures due to humidity against the higher energy costs of daily operations via other cooling methods. Multitenant data centers cannot do the same, as they do not own the servers within the data centers and will be expected to protect their customer's gear from premature failures.

You should not find any direct water evaporative cooling systems in an enterprise-grade, multitenant data center. However, this technique is very regularly used outside of the data center, to chill water loops—the remaining water—for a CRAH system. In this case, the data center's operator must keep the liquid physically separated from the data center's air to prevent any added humidity from being added to the enclosed data hall. When set up in this manner, it is a safe and lower-cost energy system, although it does require additional water pumps and fans in the CRAHs over direct evaporative. This is often called "free cooling via indirect water-side economization\evaporative."

Another fact of physics is that hot air rises due to its lower density/weight than the surrounding air. Most CRACs "suck" the hot air from the ceiling and deliver cold air from under the floor. Some CRACs, however, suck the hot air from up high in the hall and force cold air down from the ceiling as well. The delivery method selected is up to the data center provider to manage.

Chapter Summary

- CRAC: Huge air-conditioning unit that cools air by blowing that air over a cooling coil filled with refrigerant. CRACs recycle hot air.
- CRAH: Just fans and a cooling coil usually chilled with water. CRAHs push hot air outside the building.
- Air-side economization/direct air cooling: If the outside air is an appropriate temperature, the economizer simply blows that cool air directly into the facility, reducing energy requirements.
- Water-side economization/direct water evaporative cooling: Think "swamp cooler." Heat from the data center converts water in the system to water vapor. But then the water vapor condenses on your servers. You won't find this in large facilities.
- Raised floor: Cool air is delivered from below the servers. The servers rest on a raised floor.
- Slab floor: Cool air is forced down from above the servers. The servers rest on a slab floor.
- Hot and cold aisle containment: Delivers cool air only where it's needed, rather than cooling dead space in addition to servers.

CHAPTER TEN: FIRE-SUPPRESSION SYSTEMS

After studying this chapter, you should:

- Be able to describe the basic components of a fire-suppression system.
- Be able to describe how these elements work together.
- Know how to communicate to a client the strengths and limitations of a facility's systems.

The first step in choosing a business-friendly approach to fire suppression is to know what you are trying to achieve in terms of the safety of your people and/or equipment and your maximum tolerable downtime.

» COMPONENTS OF A FIRE-SUPPRESSION SYSTEM

Detection System

This must be a reliable system that detects both smoke and heat, along with other hazardous gases. The detection system should be calibrated so that it does not react to false alarms caused by other contingencies in your data center. It should also incorporate an emergency suppression system delay or cancel feature so that false alarms may be addressed before the suppressing agent is delivered.

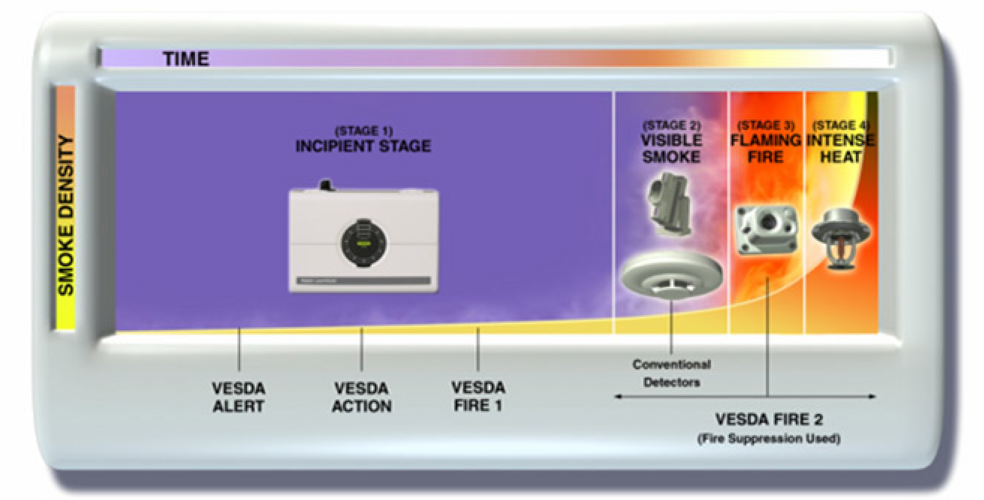


Figure 29-VESDA ASD fire suppression diagram

For example, the Very Early Smoke Detection Apparatus Aspirating Smoke Detection (VESDA ASD) by Xtralis continuously samples the air to detect even the most minimal levels of smoke and pre-smoke particulate. Early warning adds crucial time to determine the best first response that will minimize damage to the facility and/or disruption to business. This system can also be outfitted for the detection of harmful gases and environmental monitoring.

Alarms

These simply produce a loud noise, usually along with flashing lights. The alarm should be coordinated with a suppression system to allow employees enough time to evacuate the area before the fire is put out.

Fire Extinguishers

These must be properly placed around the building, according to regulations, so that a fire can be extinguished prior to impacting a larger area and also so that people can evacuate, even when there is fire blocking an exit.

Suppression Agents

Wet Pipes: Water sprinklers installed throughout the building that spray water after the automatic detection of heat and smoke, or by manual operation.

Dry Pipe/Pre-Action: This is the industry standard. Like wet pipe, except water is not kept in the pipes; this means that pipes don't collect condensation that could drip on equipment.

Gaseous Agents: These deprive a fire of heat and oxygen for immediate suppression. Advantages are that gaseous agents do not do substantial damage to equipment and allow for efficient business recovery. Some of these systems are harmful to humans, however, and agents such as carbon dioxide and Inergen remove oxygen from the air, causing an unsafe atmosphere for people to try and breathe in. These systems cost more to install than water systems, require annual maintenance, and require floor space for the canisters.

Some examples:

- HFC-227ea: The first fire suppression agent to replace Halon 1301 as a non-ozone-depleting agent. It's colorless, odorless, and the EPA allows it to be used in occupied areas.
- HFC-125: Not only is this agent colorless, odorless, and without residue, but this agent also is non-conductive and non-corrosive.
- FM-200: Extinguishes fire very rapidly to significantly reduce downtime; can also be used with people present. This has been the industry standard for the past few years.

- HI-FOG: Utilizes a central pumping unit that can deliver fire suppression to any area within the facility. HI-FOG is a cost-effective solution when compared with other suppressing system lifecycles and also allows for expansion should the data center need to be augmented. HI-FOG is a good option since it releases less water than sprinkler systems and it suppresses only the area where there is an actual fire.

Emergency Power-off

The EPO is the "big red button" that cuts the power to an entire facility and was primarily designed to give firefighters the ability to shut down power quickly. Though the EPO was formerly made a requirement for all data centers by the National Fire Protection Association (NFPA), as of 2011 this is no longer the case. Most current EPO systems now have some form of dual-activation method to ensure accidental activations do not occur, and some facilities that meet more stringent criteria are able to remove the EPO entirely so long as other methods are available to cut power to the entire facility if required.

Chapter Summary

- Detection system: Something that recognizes smoke or heat without triggering costly false alarms.
- Alarms: Notifications that allow employees to respond to a fire and evacuate the premises.
- Fire extinguishers: Tools used to put out fires. Duh. They should be positioned all over the place.
- Suppression agents: Your basic suppression agents are wet pipe, dry pipe/pre-action, and gaseous agents.
- Emergency power-off: A big red button that shuts everything down. This is actually a thing.

CHAPTER ELEVEN: UPTIME INSTITUTE CERTIFICATIONS

After studying this chapter, you should:

- Be able to accurately define the entity Uptime Institute.
- Be able to accurately describe the key features of each of the Uptime Institute's tier categories.
- Be able to determine a client's actual tier needs and communicate why.

At some point while working in or around the data center industry, there will be mention of what "Tier" a specific facility is—or is not.

This section serves a very brief overview of what this "Tiered" classification system amounts to, and you should be familiar with what is being referred to when a provider states that the facility is a Tier 2 or Tier 3 design. This section also serves to define the different certifications the Uptime Institute offers.

It is important to note that a Tier 1 and Tier 2 network provider is ***very*** different from a Tier 1– and Tier 2–designed data center. For those inclined to dig beyond the information provided in this chapter, please refer to the *Uptime Institute Data Center Site Infrastructure Tier Standard: Topology* whitepaper. It's readily available online.

» WHO OR WHAT IS THE UPTIME INSTITUTE?

Uptime Institute, an independent division of The 451 Group, provides education, publications, consulting, certifications, conferences and seminars, independent research, and thought leadership for the enterprise data center industry and for data center professionals.

The Uptime Institute is an unbiased, third-party data center research, education, and consulting organization focused on improving data center performance and efficiency through collaboration and innovation. The Uptime Institute serves all stakeholders of the data center industry, including enterprise and third-party operation, manufacturers, providers, and engineers.

Founded in 1993, the Uptime Institute pioneered the creation and facilitation of end-user knowledge communities to improve reliability and uninterruptible availability—or uptime—in data center facilities and information technology organizations. In 2009, The 451 Group acquired the Uptime Institute. The 451 Group analyzes the tech industry and focuses on enterprise IT innovation.

The Uptime Institute continues to expand its services and activities around the world. It offers Tier certifications in 26 different countries, Accredited Tier Designers and Accredited Tier Specialists representing 60 different countries, and Uptime Institute Networks are currently growing in Brazil, the Asia Pacific region, and EMEA. The Uptime Institute has permanent offices in Brazil, Dubai, London, and Singapore, in addition to the United States—in New York, Denver, Seattle, and San Francisco.

» TYPES OF CERTIFICATIONS

Tier Certification Of Design Documents

This certification answers the question, "Is the facility in construction or about to be constructed Tier compliant based on the architecture/design documents the general contractor is working with?…And if so, to what Tier Standard are they?"

Tier Certification Of A Constructed Facility

This certification answers the question, "Has the facility been constructed as designed and does it have the capability to meet the target Tier objective?"

Operational Sustainability Certification

This certification answers the question, "Is the facility managed/operated consistent with the behavior and risk objectives outlined within the Constructed Facility Tier Certification received?"

A Certification of Design Documents is required before a Certification of Constructed Facility will be awarded, and a Certification of Constructed Facility is required before an Operational Sustainability Certification will be awarded.

- Tier I: Basic Site Infrastructure (non-redundant). Site has N redundancy across all critical infrastructure components related to power and cooling.
- Tier II: Redundant Site Infrastructure Capacity Components. Site has N+1 redundancy across all critical infrastructure components related to power and cooling.
- Tier III: Concurrently Maintainable Site Infrastructure. Site has N+1 redundancy across all critical infrastructure components related to power and cooling ***and*** multiple distribution paths serving the critical environment on the data center floor, though only one path is required to be active at any time.
- Tier IV: Fault Tolerant Site Infrastructure. Site has 2N+1 redundancy across all critical infrastructure components related to power and cooling and multiple active distribution paths serving the critical environment on the data center floor.

Other Relevant Information

A company claiming a specific tier ***design*** level is much different than a company that's received a tier ***certification*** from the Uptime Institute. A data center will not be granted "public acknowledgment of the [Tier] achievement in the form of foil and website listing" until it completes both the Tier Certification of Design Documents ***and*** the Tier Certification of Constructed Facility. If a facility only completes Tier Certification of Design Documents, it will receive a letter only, no public acknowledgment.

A facility will only be certified and rated as high as its weakest link, meaning that a Tier IV power distribution design coupled with a Tier II cooling design and Tier III designs everywhere else within the facility will still render a Tier II rating and certification. This is important to note, as most providers who have not undergone the certification process will present the opposite in sales materials, and will emphasize the systems designed with the highest rating only.

Uptime Institue Tier Classifications				
	TIER 1	TIER 2	TIER 3	TIER 4
Active Capacity Components to Support the IT Load	N	N+1	N+1	N After Any Failure
Distribution Paths	1	1	1Active and 1 Alternate	2 Active Simultaneously
Concurrently Maintainable	No	No	Yes	Yes
Fault Tolerance	No	No	No	Yes
Compartmentalization	No	No	No	Yes
Continuous Cooling	No	No	No	Yes

Figure 30-Uptime Institute tier requirements table

Other Data Center Standards

Other common certifications/standards include TIA-942 and BICSI-002. These standards have the same defined levels, but cannot utilize the Tier labeling; the term "class" may be used, or just verbiage concurrently maintainable. For example, TIA-942 has a checklist that any mechanical/ electrical company can utilize to identify whether a facility is compliant with a standard. BICSI-002 provides significant information for the proper design of a facility to meet a certain redundancy level, but no checklist. Each standard is somewhat different and everyone needs to understand that just because a facility meets the standard for TIA-942, for example, that does not mean it also meets Uptime Institute standards. Uptime is the only group that can inspect/review or award the Uptime certification. For TIA-942 and BICSI-002, any mechanical and electrical firm can review a facility and state whether it is in compliance with either standard. Uptime is definitely the most well-known data center certification.

Chapter Summary

- Tier 1: Basic infrastructure without redundancy.
- Tier 2: N + 1 redundant facility.
- Tier 3: N + 1 redundancy ***and*** multiple power and cooling distribution paths available, although only one path is required to be active.
- Tier 4: 2N + 1 redundancy ***and*** multiple active paths serving site infrastructure.

CHAPTER TWELVE: DISASTER RECOVERY OPTIONS

After studying this chapter, you should:

- Be able to define basic recovery strategies.
- Be able to identify which recovery strategy lends itself to which circumstances.
- Speak with a client about recovery strategies while understanding their standard terminology.

The phrase Disaster Recovery, or "DR" for short, is thrown around often and sometimes confused with the phrase "business continuity." To be clear, DR is simply a ***subset*** of business continuity planning, and is focused almost exclusively on the process of saving data and ensuring the ability to recover data in the event of an outage. Business continuity is more broadly focused on the business at large. It refers to the systems, operations, management, and processes in place to ensure a company's entire staff and facilities can continue to do business during an outage or disaster. IT is thus a critical component of a complete business continuity strategy, but not the only concern. Data recovery, on the other hand, is the ***sole*** objective of a DR strategy.

» MANAGED RECOVERY PROGRAM

If a client has spent the time and effort to put together a complete MRP, it is fair to assume they have taken into consideration and documented how all the critical and noncritical applications used within their organization will be managed and restored should a failure occur across any component of the IT infrastructure supporting the business.

As such, a complete MRP is very complex and time-consuming to put together. It is, however, essential for any company taking DR seriously to spend the time and effort to develop a MRP. As a result, only those businesses that stand to lose a significant amount of money in the event of an outage—credit card companies, banks, trading firms, large e-commerce companies, etc.—tend to receive the budget approvals to compile and execute a complete MRP.

» RECOVERY TIME OBJECTIVE

RTO is an essential component of the MRP and specifically speaks to how long a company can go without the use of the systems and applications running within the business. Some systems, such as corporate e-mail or the applications that verify credit card transactions will have a mandated RTO of as close to zero seconds as possible. Other systems and applications, such as file/e-mail archives or noncritical corporate applications, may have a mandated RTO of closer to two to three days.

It is safe to assume that the closer to zero seconds your RTO is, the more expensive it will be to achieve the desired RTO. Understanding that different systems and applications should have different RTOs is the first step toward designing a functional MRP.

» RECOVERY POINT OBJECTIVE

An RPO is another essential component of the MRP, and specifically speaks to how long you wish to go between making a backup of specific data or data sets, such that if there were an outage, you could recover data from a chosen point in time. It's similar to RTO, in that some data requires an RPO as close to zero seconds as possible, such as stock market buy/sell transactions where millions of dollars change hands every second. Other data sets are not as essential, such as the backup of your personal laptop, where the files generally don't change every day. The cost of having your RPO as close to zero as possible will also be high, much like RTO.

Cost will increase based on:

1. The size of the data sets you need to backup.
2. The RTO assigned to the specific data sets.
3. The number of times data must be recorded over time.
4. How far back this data must be saved.
5. The volume of data added to the total saved file over time.

» ACTIVE-ACTIVE

Much like a 2N power configuration, this refers to an environment where there is an exact mirror of data across two systems. If either system were to fail, there would be no downtime or loss of data since it exists and operates in synchrony elsewhere. RTO and RPO are close to zero seconds in this configuration.

» ACTIVE-PASSIVE

Much like an N+1 power configuration, this refers to an environment where there is a mirrored system that is on standby to receive regular updates of data. If the primary system were to fail, the standby environment would be "turned on" to serve as the new production environment. It would restore the original production environment to whenever the last update was received, based on the RPO implemented. There are a number of ways to configure this type of environment, so the RTO can vary from a few minutes to a few hours or a few days for noncritical systems, if data must be recovered from tape that is sitting off site.

» PHYSICAL-TO-VIRTUAL RESTORATION

A P2V refers to an active/passive environment where server configurations and applications of a standby system mirror the configurations of a production system. More often than not, the configurations exist as virtual instances on shared infrastructure. Many companies are concerned with the reliability and stability of a virtual environment mirroring the primary physical production environment. However, the cost is relatively low as a disaster recovery solution. And, with proper and regular testing, this solution has proven viable and effective—so long as the initial physical primary environment can be brought back live within a few hours or days.

» PHYSICAL-TO-PHYSICAL RESTORATION

This refers to an active/passive environment where the physical servers, switches, routers, storage devices, etc. of the standby system mirror the production system almost exactly. As such, the cost of a P2P environment far exceeds the cost of a P2V, and this system is rarely implemented. The nominal cost of converting a P2P environment into an active-active environment also makes such a solution less popular.

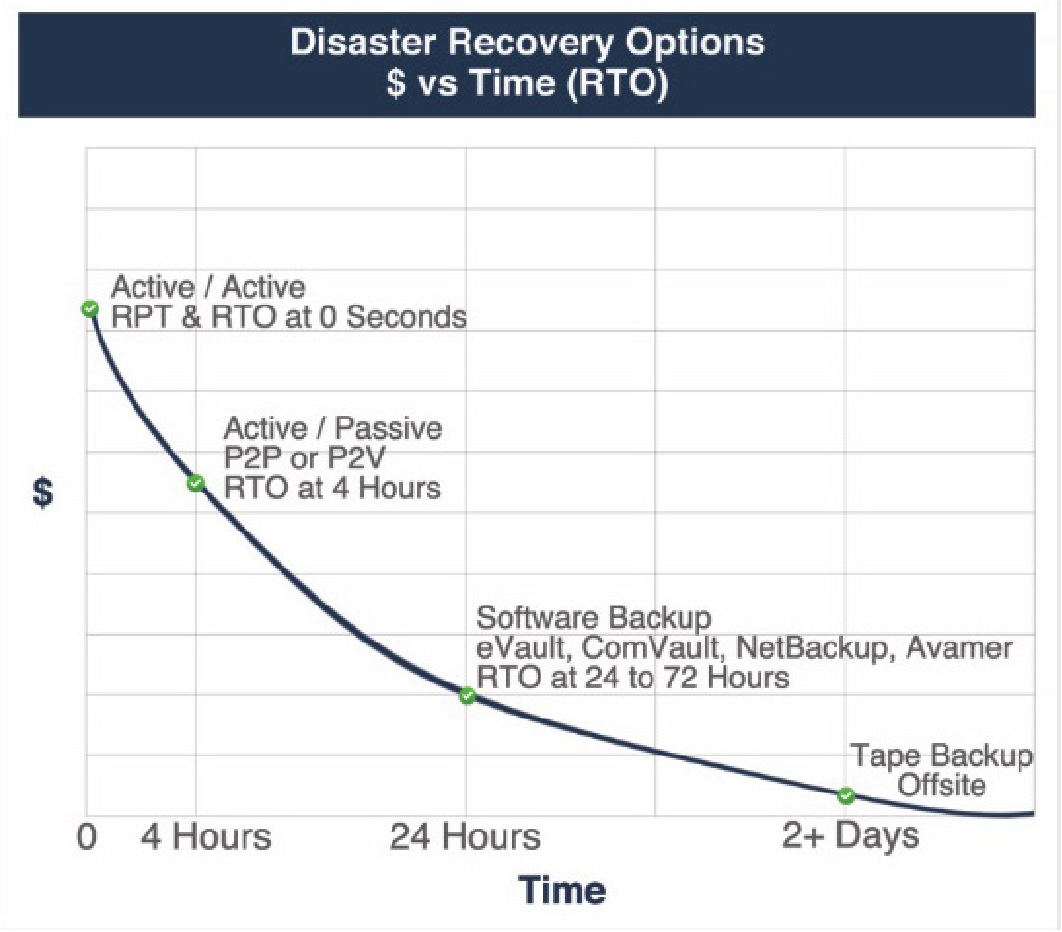


Figure 31-Disaster recover options graph

Chapter Summary

- Managed Recovery Program: A roadmap to recovery that an organization produces before recovery is needed. Super expensive and time-consuming.
- Recovery Time Objective: Part of the MRP. This describes how long an organization can go without certain applications and systems.
- Recovery Point Objective: Part of the MRP. Describes how frequently you'll back up your systems.
- Active-active: Think virtual 2N. There are two exact copies of data across two systems, ensuring no downtime whatsoever.
- Active-passive: Think virtual N + 1. There's a copy on standby, and the copy receives regular updates.
- Physical-to-virtual restoration: The data copy exists as a virtual instance on shared infrastructure.
- Physical-to-physical restoration: The data copy exists on a physical copy of the primary infrastructure. Super pricey.

CHAPTER THIRTEEN: GEOLOCATION

After studying this chapter, you should:

- Be able to accurately describe how a data center's location can produce unique competitive advantages and disadvantages.
- Be able to accurately identify when a client may want a data center located nearby rather than farther away.

The location where a company hosts its infrastructure can play a critical role in that company's success or failure. It is therefore important to take several different variables into consideration before recommending one location over another:

» LATENCY

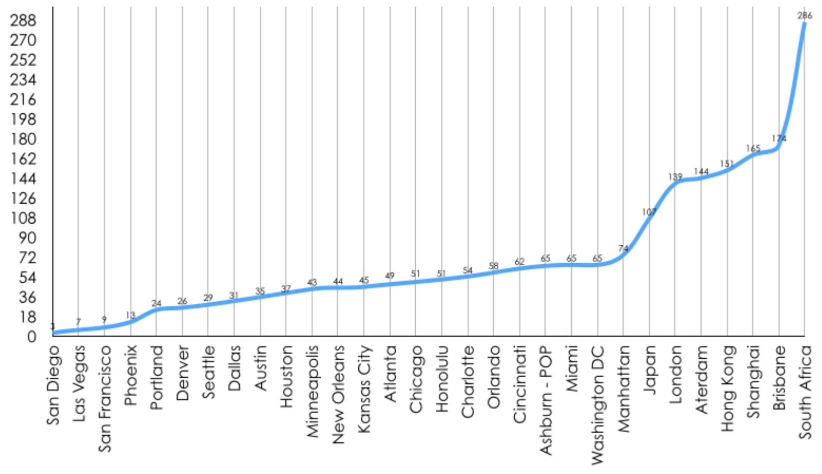


Figure 32-Latency from Los Angeles Graph

Video streaming, mobile applications, voice services, credit card processing, and financial trading firms are examples of services that live and die by their ability to deliver data milliseconds before a user loses interest and tries a different service.

Though data can theoretically travel across a fiber line at the speed of light, the distance and infrastructure that sits between your data and your end users will add milliseconds of latency that can become significant over multiple hops.

Content Delivery Networks (CDNs) have carved a niche in the market, understanding this reality and affording clients the ability to host their data as close to their end users as possible around the world.

Understanding the latency sensitivity of the applications and data a company will be serving is therefore imperative when determining the location of a data center. It could be a nonissue, or it could be the primary decision criteria for a client.

» ACTS OF GOD, FORCE MAJEURE

If 100 percent uptime is absolutely critical to your business and a DR strategy has not been put in place with an acceptable and tested RTO, you should be well aware of the potential for acts of God and force majeure to affect the facility that houses your critical infrastructure. First and foremost, however, it is important to note exactly what acts of God and force majeure mean, as most often the SLA tied to a data center services contract becomes null/void should these actually happen.

Geological and natural disasters are some of the largest factors tied to the phrase "acts of God." There is a mountain of historical data indicating that Mother Nature has a track record of acting more often in specific regions of the country than others. Devastating earthquakes are inevitable on the West Coast, as depicted in the map below, and blizzards that shut down facilities and make them impossible to reach are inevitable in the North. Tornadoes that wipe buildings clear off the map are inevitable in the Midwest and South, and hurricanes that flood entire cities happen up and down the East Coast and along the Gulf of Mexico shoreline.

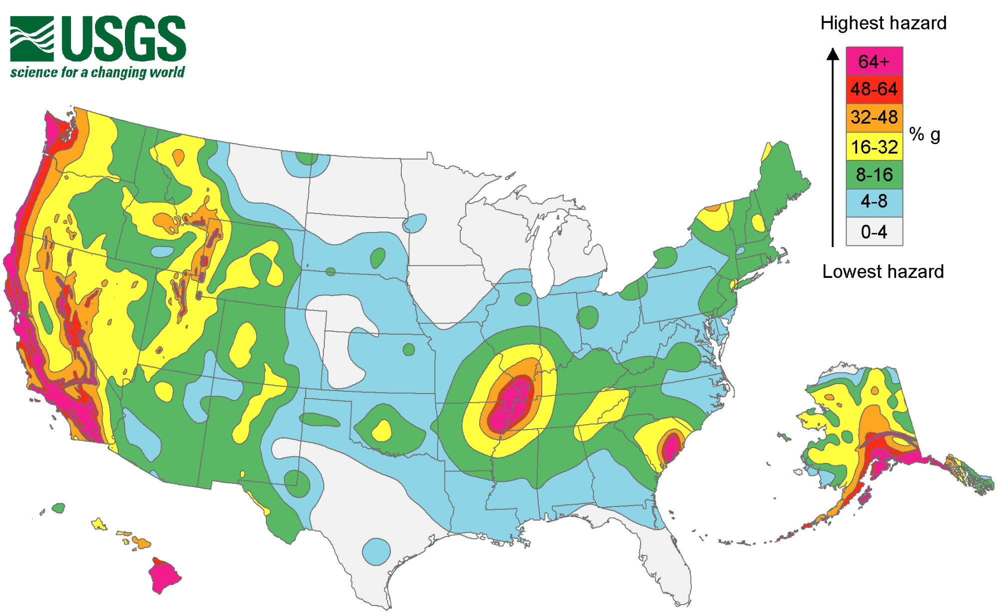


Figure 33-Map of seismic hazard zones in the US

Risks beyond the control of the service provider that are generally not geological in nature, such as crimes, riots, strikes, wars, and terrorism, fall under the "force majeure" label. As a point of reference, bad weather and large sporting events are examples that would not fall under this category, as they are not in their nature unusual incidents that might interrupt work. It should not be surprising to learn that large cities with dense populations and critical infrastructure built throughout are at a higher risk of a force majeure incident occurring.

» TAX INCENTIVES

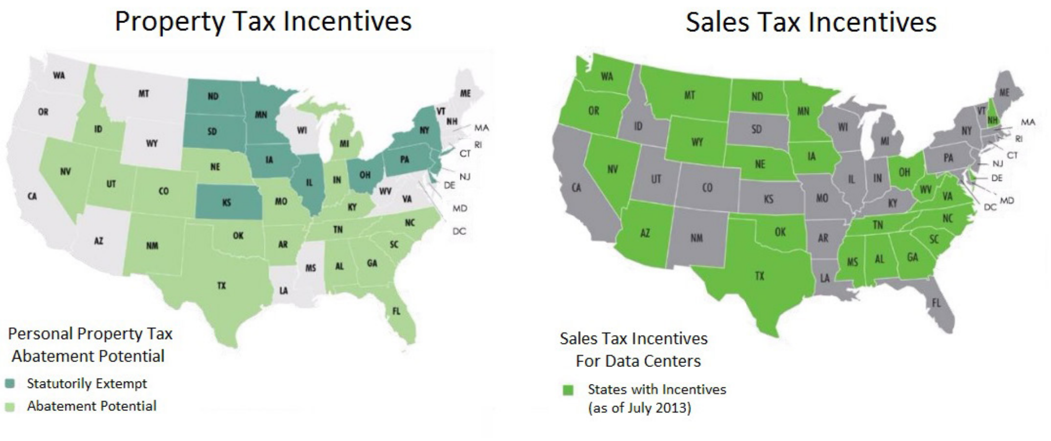


Figure 34-Maps of tax incentives by state

Over twenty-five states now offer tax incentives for clients and providers to locate their data facilities within their borders. Even more cities offer additional incentives to do the same. Primarily, these amount to a rebate on power costs for designing and managing "efficient" facilities and sales-tax rebates for infrastructure purchases such as hardware, generators, PDUs, UPSs, and so on.

For a small company, this may only amount to a few hundred dollars a year in rebates; however, to those who spend many millions on infrastructure and power each year, this can amount to a few hundred thousand dollars, or more, in tax credits.

» COST OF POWER

The numbers within the chart below speak for themselves, and this is why there have been so many large data center deployments by Dell, Yahoo, Facebook, Google, Microsoft, and others in the Pacific Northwest and Northeast, where the cost of power in some locations is sub 3 cents kW/hour! Since power is the majority of the ongoing costs associated with maintaining a data center, these savings become extremely significant.

» COST OF COOLING AND CLIMATE

Considering the numbers above and the fact that the generation of cooling constitutes the majority of the additional power consumed by data centers, geographic regions that maintain a relatively constant temperature at or below sixty degrees afford providers substantial savings by eliminating the need for mechanical cooling generation.

Per kW/ hour	Cost per 1 kW	Cost per 10 kW	Cost per 100 kW	Cost per 1 mW
$0.02	$14.60	$146.00	$1,460.00	$14,600.00
$0.03	$21.90	$219.00	$2,190.00	$21,900.00
$0.04	$29.20	$292.00	$2,920.00	$29,200.00
$0.05	$36.50	$365.00	$3,650.00	$36,500.00
$0.06	$43.80	$438.00	$4,380.00	$43,800.00
$0.07	$51.10	$511.00	$5,110.00	$51,100.00
$0.08	$58.40	$584.00	$5,840.00	$58,400.00
$0.09	$65.70	$657.00	$6,570.00	$65,700.00
$0.10	$73.00	$730.00	$7,300.00	$73,000.00
$0.11	$80.30	$803.00	$8,030.00	$80,300.00

Figure 35-Costs per various kW amounts

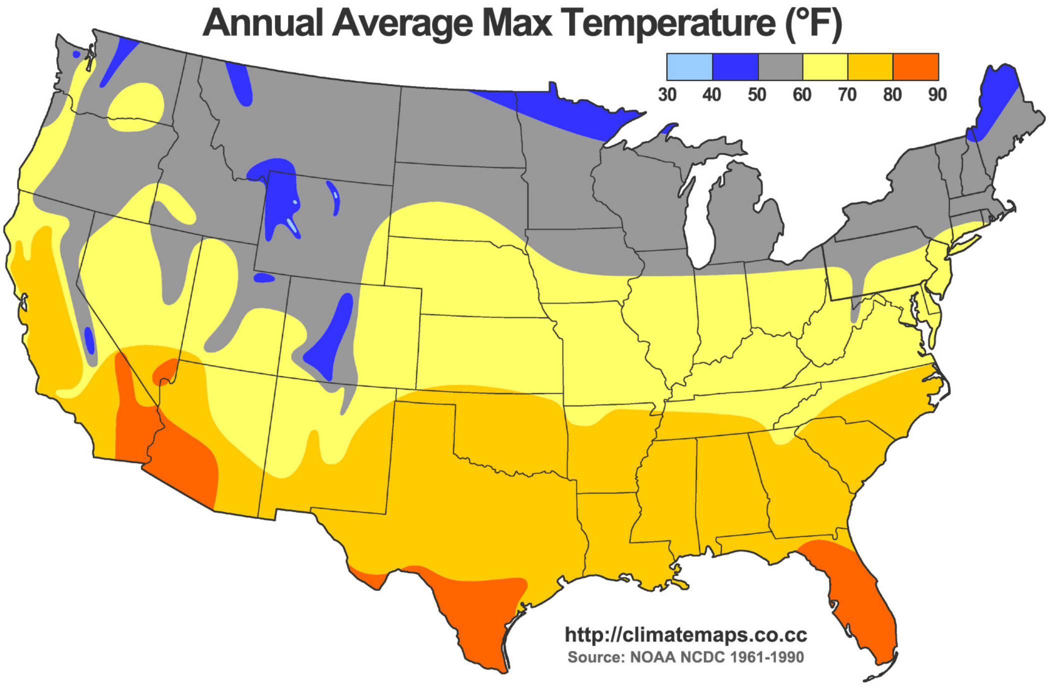


Figure 36-Annual Average Max Temperatures by Region

Specifically, this can mean the difference between a PUE of 1.6 and 1.3 for a provider with the exact same infrastructure deployed and same utility rate per kW/hour, but with one facility deployed in Las Vegas and one deployed in Seattle. This is because the Washington facility can make use of outside air to cool the facility for the majority, if not the entirety, of the year.

» COST OF REAL ESTATE

Though a colocation client will not directly see the impact of a drastic per square foot cost differential, the provider most definitely will. This is why, all things being equal from a power, cooling, and infrastructure perspective, a data center in San Francisco or NYC will generally be 50% to 200% more expensive than a data center in Sacramento, CA or Buffalo, NY.

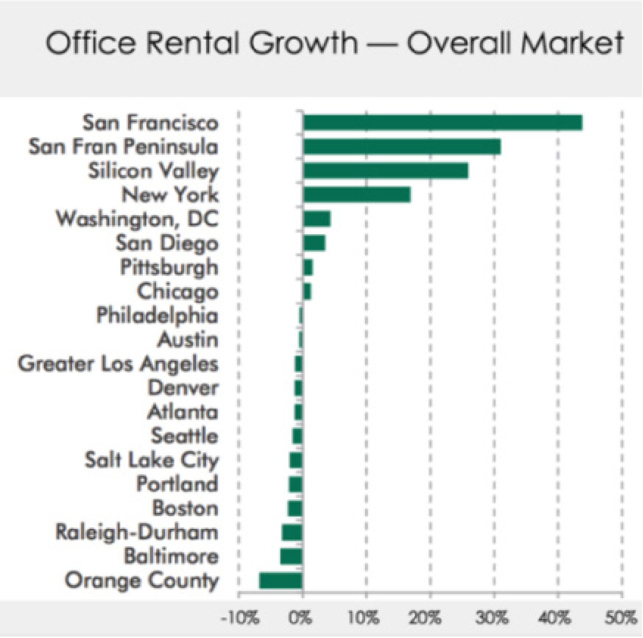


Figure 37-Office rental growth in various markets

» COST OF LABOR

Though mostly overlooked, a company should also take into consideration the cost of having smart hands on site, if required, to manage and maintain a facility. For the provider who will require multiple facilities managers or contractors on site daily, or a client who must maintain a staff on site or near the facility, this 25–50 percent differential between talent in San Francisco and Rancho Cordova, California, for example, is very real and adds up over time.

An IT operations manager in San Francisco costing $80K/year, for example, might cost only $60k/year in Rancho Cordova, due to the cost of living being so much lower between the two locations. This $20K differential will equate to a $1,666/month additional premium per person to do business in San Francisco.

Chapter Summary

- Effects of geolocation
- Latency: Data travels as fast as light, but with all the cabling in the world, the distance between a data center and the data's user can mean more or fewer milliseconds of latency.
- Acts of God: Some regions are more disaster prone than others, so a data center's location may place infrastructure in greater or less risk.
- Tax incentives: Different states and regions offer different tax incentives that change the economics of data center operation. Some places may be cheaper than others.
- Cost of cooling and climate: If a data center can utilize external air to cool servers, that facility will run more efficiently than others elsewhere. Not all regions, however, enjoy cool, stable temperatures.
- Cost of labor: Labor costs vary from region to region, and these costs affect data center economics.

CHAPTER FOURTEEN: FLAVORS OF CLOUD COMPUTING

After studying this chapter, you should:

- Be able to accurately define "the cloud."
- Be able to accurately describe what people mean—and what they think they mean—when they toss around cloud jargon.
- Be able to discuss actual needs with clients and the limitations of any given cloud infrastructure.

Unfortunately, the phrase "cloud computing" now means so many different things to so many different people. You must constantly ask, "What ***exactly*** do you mean by cloud computing?" As we all should know by now, there is no cloud magically storing data in the sky, although I am routinely floored to discover that for many newcomers to the industry this is still unclear.

To be fair, we have been experiencing "cloud-washing" within the IT industry, which is no different than how we have been experiencing "green-washing" over the past decade across almost all industries. Companies such as Salesforce.com have been leading this campaign, spending millions of dollars by branding and promoting the phrase "cloud computing." For the majority of companies like Salesforce.com, the phrases "we are in the cloud" and "we are a cloud company" now merely means "our data is hosted on a server somewhere that you do not own." Salesforce.com, for example, is simply a Software as a Service (SaaS) company with a Software Platform (PaaS) that their application partners can also leverage to sell software tied to their core database applications. Internally, Salesforce.com leverages a variety of virtualization technologies to ensure a scalable and seamless delivery of their software and their partners' software to their end users.

It is important to know that this definition is different than what most technical people mean when they use the phrase cloud computing. Those working within the industry, who are directly responsible for designing and managing physical and virtual compute resources, use cloud computing when speaking to ***elastic*** and ***on-demand*** hosting services. These services are achieved and delivered by the virtualization and automation of physical IT infrastructure that can quickly respond to fluctuations in the demand for resources. These services also afford the service provider the ability to monitor and bill customers only for the resources they use each month and do not require long-term contracts.

Elastic and on-demand hosting services are thus only one of many possible options that can be leveraged for organizations with IT infrastructure needs. Thus, when you hear the buzz phrase "cloud computing," please ignore it. Below are definitions for the different hosting and cloud options available for customers today.

» HOSTING

Defined by services delivered on month-to-month or long-term contracts. The client leases the ability to use a specific and contracted amount of storage, CPU, and RAM, and is billed for the ability to use these resources, regardless of actual usage.

Hosting: Dedicated Physical

The client leases physical hardware, though they do not have direct physical access to the machine, and is provided root access with clearly defined storage, CPU, and RAM. The entire machine is dedicated to the user, and any operating system or applications can be installed or reinstalled at any time.

Hosting: Managed Physical

The client leases the physical hardware, though they do not have direct physical access to the machine, and there is clearly defined storage, CPU, and RAM. Though the entire machine is dedicated and private to the user, the hosting provider is responsible for managing the operating system and can also manage other applications for the client that have been installed. The user generally does not have root access to the machine.

Hosting: Virtual Server

The client leases virtual storage, CPU, and RAM on hardware they do not own. This virtual "instance" or "VM" most often resides on hardware that is hosting multiple VMs across a number of different users. Though the user can install or reinstall any operating system it wishes on the VM, it does not have root access to the physical hardware. The user is likely unaware of the hypervisor or technology leveraged on the back end to create the virtual instance(s) and is also likely entirely unaware of who else is sharing the same physical or virtual resources.

Hosting: Managed Virtual Server

The client leases virtual storage, CPU, and RAM on hardware they do not own. Exactly like a managed physical solution, the hosting provider maintains root access to the operating system and can manage different applications installed on the VM as well if needed.

This VM can reside on hardware that is hosting multiple VMs across a number of different users (managed shared) or reside on a managed physical server, providing exclusive use to the compute resources for the client (managed private). Either way, the user does not have root access to the physical hardware. Same as a standard Virtual Server, the user is likely unaware of the hypervisor or technology leveraged on the back end to create the virtual instance(s) and is also likely entirely unaware of who else is sharing the same physical or virtual resources.

» CLOUD: INFRASTRUCTURE AS A SERVICE (IAAS)

This is defined by hosting services delivered on month-to-month or minute-to-minute contracts, providing access to an undefined and seemingly endless or "elastic" pool of storage, CPU, and RAM resources. The client can have multiple VMs set up but will only pay for the actual storage, CPU, and RAM consumed over the course of the month. Some data center providers will claim they are also IaaS providers, which in their context is definitely the case; however, the term has come to primarily refer to the ability for consumers to leverage the specific storage, CPU, and RAM compute resources on demand as outlined above.

Cloud: Unmanaged Public

Similar to a Virtual Server, however the billing model in this case is "on-demand," and the provisioning of new resources generally takes only minutes. Physical storage, CPU, and RAM resources are shared across multiple users, leveraging specific hypervisor and virtualization technologies hidden from the user. The client is responsible for managing, monitoring, and turning up and down specific resources and setting thresholds on actual utilization.

Cloud: Managed Public

Same as unmanaged public, but the cloud provider is responsible for managing, monitoring, and turning up and down specific resources and setting thresholds on actual utilization.

Cloud: Unmanaged Private

This solution could also be called a Dedicated Virtual Server solution. It generally requires the client to commit to and pay to reserve a finite amount of storage, CPU, and RAM to truly have access to dedicated resources on dedicated hardware. One of the benefits of this type of solution is that the client can reserve additional dedicated resources if needed. However, the solution as a whole is much more expensive than if these resources were within a shared environment.

Cloud: Managed Public

Same as unmanaged private, but the cloud provider is responsible for managing, monitoring, and turning up and down specific resources and setting thresholds on actual utilization. Also more expensive than if these resources were within a shared environment.

Cloud: Owned Private

This type of solution is *not* a hosting solution but deserves mention due to its increasing popularity. The client is responsible for owning and managing the physical hardware that the virtualization and "cloud" software sits on. In this case, the cloud refers specifically to the software leveraged to aggregate and deploy storage, CPU, and RAM resources within all machines through an interface similar to that found within the "on-demand" cloud solutions mentioned above. CloudStack, OpenStack, vCloud, and Eucalyptus are examples of the technology leveraged to "cloudify" your own infrastructure. Many hosting and cloud service providers, in fact, leverage these solutions on their back end to deliver the cloud and hosting solutions outlined above.

Cloud: Owned Managed

Same as owned private, but the client has outsourced the bulk of the system's management, monitoring, and turning up and down specific resources, along with setting thresholds on utilization and advocating for the purchase of additional hardware for more capacity.

Cloud: Hybrid

This solution affords a client the ability to gain visibility, manage, and deploy resources across multiple owned or leased hosting and cloud environments. Companies tend to leverage this kind of solution to manage separate production and development or test environments, or to give their own production environment the ability to "burst" into a public cloud, should storage, CPU, or RAM resources exceed those within the existing owned environment.

Chapter Summary

- Hosting: dedicated physical. Client leases physical hardware, but does not have physical access to physical hardware. They access servers remotely.
- Hosting: managed physical. Same as above, but the provider manages the OS and other root applications.
- Hosting: virtual server. Client leases virtual storage on hardware that likely contains multiple virtual instances.
- Hosting: managed virtual server. Same as above, but the provider manages the OS and other root applications.
- Cloud: unmanaged public. Similar to a virtual server, but billing model is on demand. Client manages, monitors, and throttles resource consumption. Storage and resource availability is elastic, or able to grow seamlessly with demand.
- Cloud: managed public. Same as above, but the provider manages, monitors, and throttles resource consumption.
- Cloud: owned private. The client owns and manages physical hardware and creates a virtual or cloud instance on said hardware.
- Cloud: managed private. Same as above, but the client outsources much of the system's management.
- Cloud: hybrid. Client accesses and manages multiple cloud instances and can shift resources between them.

CASE STUDY EQUATIONS AND SOLUTIONS

Case Study 1

Provider A	Equation	(8 x $700)+(14 x $1,200)+(14 x $600)+($795)+($397.50)
	Solution	$31,992.50 MRC x 36 months = $1,151,730 TCO
Provider B	Equation	(9 x 7)+3 x $275
	Solution	$18,150 MRC x 36 months = $653,400 TCO
Provider C	Equation	((9 x 7)+3) x $145 = $9,570 "rent" + ((9 x 7)+3) x 730 x $.12 x 1.5 = $8,672.40 "variable"
	Solution	$18,242.40 MRC x 36 months = $656,726.40 TCO

Case Study 2

Provider A	Equation	(2 x $500)+(2 x $2,750)+(2 x $1,375)
	Solution	$9,250 MRC x 36 months = $333,000 TCO
Provider B	Equation	((30 x 208 x 1.73 x 80%)/1000) x 2 = 17.27 = 17 x $600
	Solution	$10,200 MRC x 36 months = $367,200 TCO
Provider C	Equation	((30 x 208 x 1.73 x 80%)/1000) x 2 = 17.27 = 17 x $400
	Solution	$6,800 MRC x 36 months = $244,800 TCO
Provider D	Equation	((30 x 208 x 1.73 x 80%)/1000) x 2 = 17.27 =17 x $275 = $4,675 "rent" + ((30 x 208 x 1.73 x 80%)/1000) x 2 = 17.27 =17 x 730 x $.11 x 1.6 = $2,184.16 "variable"
	Solution	$6,859.16 MRC x 36 months = $246,929.76 TCO

Case Study 3

Provider A	Equation	4.5 x 100 x $210
	Solution	$94,500 MRC x 60 months = $5,670,000
Provider B	Equation	4.5 x 100 x $135 = $60,750 "rent" + (4.5 x 100) x 730 x $.06 x 1.6 = $31,536 "variable"
	Solution	$92,286 MRC x 60 months = $5,537,160 TCO
Provider C	Equation	4.5 x 100 x $135 = $60,750 "rent" + (4.5 x 100) x 730 x $.06 x 1.4 = $27,594 "variable"
	Solution	$88,344 MRC x 60 months = $5,300,640 TCO

FIGURE AND PHOTO INDEX

ADDITIONAL NOTES

"Try not. Do, or do not. There is no try."

—Master Yoda