

Home > Data center design and facilities

FEATURE

How to design and build a data center

Designing an efficient data center is no small feat. Review data center facility and infrastructure components and different standards before you begin.

By **Stephen J. Bigelow**, Senior Technology Editor

Published: 18 May 2022

A data center is the technological hub of modern enterprise operations. The data center provides the critical IT infrastructure needed to deliver resources and services to business employees, partners and customers around the world.

A small or midsize business can often implement a useful "[data center](#)" within the confines of a closet or other convenient room with few modifications, if any. However, the sheer scale involved in enterprise computing demands a large dedicated space that is carefully designed to support the space, power, cooling, management, reliability and security needs of the IT infrastructure.

As a result, a data center facility represents the single largest and most expensive asset that the business will possess -- both in terms of capital investment and recurring operational expenses. Business and IT leaders must pay close attention to the issues involved in data center design and construction to ensure that the resulting facility meets business needs throughout the facility's lifecycle and changing business circumstances.

■ What are the main components of a data center?

There are two principal aspects to any data center: the facility, and the IT infrastructure that resides within the facility. These aspects coexist and work together, but they can be discussed separately.

Facility. The facility is the physical building used for the data center. In simplest terms, a data center is just a big open space where infrastructure will be deployed. Although almost any space has the potential to operate some amount of IT infrastructure, a properly designed facility considers the following array of factors:

- **Space.** There must be sufficient floor space -- a simple measure of square feet or square meters -- to hold all the IT infrastructure that the business intends to deploy now and in the future. The space must be located on a well-considered site with affordable taxes and access. The space is often subdivided to accommodate different purposes or use types.
- **Power.** There must be adequate power -- in watts, often as much as 100 megawatts -- to operate all the IT infrastructure. Power must be affordable, *clean* -- meaning free of fluctuation or disruption -- and reliable. Renewable and supplemental/auxiliary power must be included.
- **Cooling.** The enormous amount of power delivered to a data center is converted into computing -- i.e., work -- and a lot of heat, which must be removed from the IT infrastructure using conventional HVAC systems, as well as other unconventional [cooling technologies](#).
- **Security.** Considering the value of the data center and its critical importance to the business, the data center must include controlled access using a variety of tactics, ranging from employee badge access to video surveillance.
- **Management.** Modern data centers typically incorporate a building management system (BMS) designed to help IT and business leaders oversee the data center environment in real time, including [oversight of temperature, humidity](#), power and cooling levels, as well as access and security logging.



Large enterprises can require very large data centers, like this Google data center Douglas County, Ga.

Infrastructure. An infrastructure represents the vast array of IT gear deployed within the facility. This is the equipment that runs applications and provides services to the business and its users. A typical IT infrastructure includes the following components:

- **Servers.** These computers host enterprise applications and perform computing tasks.
- **Storage.** Subsystems, such as disk arrays, are used to store and protect application and business data.
- **Networking.** The gear needed to create a business network includes switches, routers, firewalls and other cybersecurity elements.
- **Cables and racks.** Miles of wires interconnect IT gear, and physical server racks are used to organize servers and other gear within the facility space.
- **Backup power.** [Uninterruptible power supply](#) (UPS), flywheel and other emergency power systems are critical to ensure orderly infrastructure behavior in the event of a main power disruption.
- **Management platforms.** One or more [data center infrastructure management \(DCIM\)](#) platforms are needed to oversee and manage the IT infrastructure reporting on system health, availability, capacity and configuration.

When a business decides to design and build a data center, the natural focus is on the design and construction of the facility. But IT leaders must also consider the infrastructure that will go into the facility to validate the project.

■ How to design a data center

There are no required or obligated standards for data center design or construction; a data center is intended to fit the unique needs of the overall business, not the reverse. However, the primary purpose of any standard is to establish a common platform of best practices. An assortment of current data center standards exists, and a business can incorporate one or more standards -- or parts of standards -- into a data center project. Standards help ensure adequate attention is placed on these factors, among others:

- conceptual design;
- layout and space planning;
- building construction requirements;
- physical security issues;
- building internals (mechanical, electrical, plumbing and fire systems);
- operations and workflows; and
- maintenance.

Below are just some of the major data center design and infrastructure standards:

- **Uptime Institute Tier Standard.** The [Uptime Institute Tier Standard](#) focuses on data center design, construction and commissioning, and it is used to determine the resilience of the facility as related to four levels of redundancy/reliability.
- **ANSI/TIA 942-B.** This standard involves planning, design, construction and commissioning of building trades, as well as fire protection, IT and maintenance. It also uses four levels of reliability ratings, implemented by BICSI-certified professionals.
- **EN 50600 series.** This series of standards focuses on IT cable and network design and has various infrastructure redundancy and reliability concepts that are loosely based on the Uptime Institute's Tier Standard.
- **ASHRAE.** The [ASHRAE](#) guidelines -- which are not specific to IT or data centers -- relate to the design and implementation of heating, ventilation, air conditioning, refrigeration and related fields.

Uptime data center tier standards		
TIER LEVEL	TACTICAL/STRATEGIC APPROACH	DATA CENTER REQUIREMENTS
Tier I	Tactical (cost-/time-to-market-driven, not suitable for real-time or high-value systems)	1. A single, non-redundant distribution path serving IT equipment 2. Non-redundant capacity components
Tier II	Tactical (generally time-to-market-driven, generally not suitable for long-term or high-value systems)	1. All tier I requirements 2. Redundant capacity components
Tier III	Strategic (not suitable for real-time or high-throughput, high-value time-dependent systems requiring highest levels of availability)	1. All Tier I and II requirements 2. Multiple independent distribution paths serving IT equipment rather than one single distribution path 3. All IT equipment dual-powered and fully compatible within the topology of a site's architecture
Tier IV	Fully redundant (most highly available topology)	1. All Tier I, II and III requirements 2. Fully fault-tolerant facility, through electrical, storage and distribution networks 3. All cooling equipment independently dual-powered, including HVAC systems



Data center tier standards outline what is required to ensure reliability and perform needs are met.

In addition, there are many varied regulatory and operational standards that can be applied to data centers. Regulatory standards include HIPAA, Sarbanes-Oxley Act, SAS 70 Type I or II and Gramm-Leach-Bliley Act. Operational standards can include ISO 9000 for quality, ISO 14000 for environmental management, ISO 27001 for information security, Payment Card Industry Data Security Standard for payment card security and EN 50600-2-6 regarding management and operational information.

Standards help ensure proper data center design, construction and operation. In addition, the adoption and careful documentation of relevant standards use can help a business ensure adequate compliance through proper facility resilience, management and business continuance preparations.

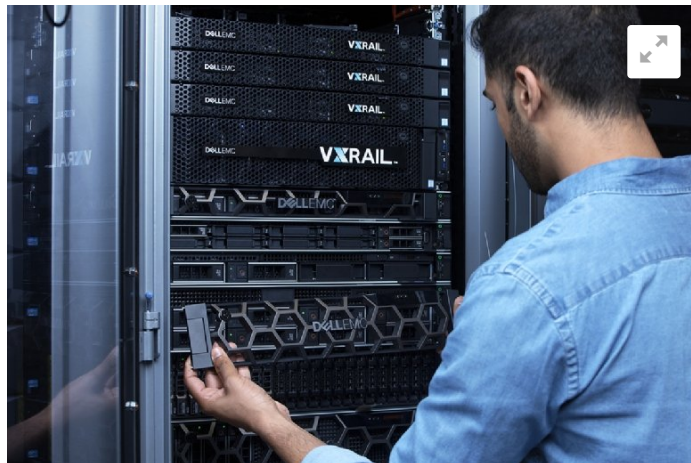
■ Physical space and data center organization

At its heart, a data center facility is little more than a big open space -- a carefully prepared warehouse intended to host and operate demanding IT infrastructure. Although an enterprise-class data center can be a large and complex undertaking, the foremost issue is a simple matter of space expressed as square feet or square meters.

Perhaps the most significant and perplexing space issue is [right-sizing the data center for the business](#). Data centers are incredibly expensive: too small, and the data center might not meet current or future business needs; too big, and enormous capital can be wasted in providing unused space. It's critical to establish a facility that offers capacity for growth yet optimizes utilization. Data center sizing is sometimes considered an art in itself. Myriad other factors to consider within a data center space include the following:

- **Lighting.** Most data center lighting is low or off without human presence.
- **Temperature.** Cooling demands can keep temperatures low, so humans might need some protective clothing.
- **Noise.** The cooling fans in dozens -- even hundreds -- of servers can produce a cacophony that requires hearing protection.
- **Weight.** Equipment is heavy, and flooring must be designed to bear the extreme weight. Special weight considerations might be needed for raised flooring used to handle cooling airflows.

Beyond the physical space, data center designs must include a careful consideration of equipment locations and layouts -- i.e., where the IT infrastructure is placed within the facility. The most common feature of any data center layout is the *server rack* -- or, simply, *rack*. A rack is an empty metal frame with standard spacing and mounting options intended to hold standardized rack-mounted IT gear, such as servers, storage subsystems, networking gear, cabling, auxiliary power systems such as UPS devices, and I/O options such as keyboards and monitors for administrative access.



This image shows an IT professional installing and maintaining high-capacity, rack mounted systems in a data center.

Racks also play a vital role in data center cooling schemes. Racks of gear are commonly organized to create [hot and cold aisles](#) that can improve cooling efficiency by allowing the introduction of cooled air into a *cold aisle*, which is heated by gear and delivered into a *hot aisle*, where the heated air can be effectively removed from the room. The [organization of aisles can also help](#) facilitate the introduction of additional doors and security measures at the ends of each aisle to limit human access.

■ Data center security

Data center security typically involves the three distinct aspects of access security, facility security and cybersecurity.

Access security. Any discussion of data center facilities must involve a consideration of physical security. Physical security is the management of human personnel and the protection of the physical facility as well as its IT infrastructure. When implemented properly, security ensures that only authorized personnel have access to the facility and gear, and that all human activities are documented. Security can involve the following array of measures:

- badge access into and around the facility (including the equipment areas);
- key access to specific racks and servers;
- logs for employee and visitor/vendor access;

- escorts for non-employees;
- video surveillance; and
- on-site security personnel.

Facility security. Physical security also extends to the integrity of the data center environment, including temperature, humidity and smoke/fire/flood conditions. This aspect of data center protection is often handled by a BMS that monitors and reports environmental or emergency conditions to building managers.

Cybersecurity. Cybersecurity focuses on controlling access to enterprise data and applications hosted within the data center's IT infrastructure. Cybersecurity is intended to ensure that only properly authenticated users can access data or use applications, and that any breaches are reported and addressed immediately. For example, physical security prevents a human from touching a disk in the data center, while cybersecurity prevents that same human from accessing data on the disk from hundreds of miles away across a network. Cybersecurity uses a mix of antimalware, configuration management, intrusion detection/prevention, activity logging and other tools in order to oversee network activity and identify potential threats.

■ Data center power and performance demands

Power is a perpetual challenge for any enterprise-class data center. A large facility can consume about 100 megawatts -- enough to power around 80,000 homes -- and power poses the single biggest Opex for an enterprise-class data center. Therefore, data center operators place the following demands on utility power:

- **Capacity.** There must be [adequate power to run the data center](#).
- **Cost.** Power must be as inexpensive as possible.
- **Quality.** Power must be electrically clean (i.e., free of undesirable electrical noise, surges and spikes).
- **Reliability.** Power must be free of brownouts, blackouts or other disruptions.

These issues are increasingly addressed with locally generated and increasingly renewable options, including wind, solar and on-site generation.

But for a business to understand power issues for any data center site, it's important that data center designers and IT leaders calculate the power demands of the facility and its IT infrastructure. It's this benchmark that enables a business to understand approximate power costs and discuss capacity with regional utilities.

There is no single means of estimating power requirements. For the facility, power is a straightforward estimate of lighting and HVAC demands. IT infrastructure power demands can be more convoluted because server power requirements fluctuate with *workload* -- i.e., how much work the applications are doing -- and the configuration of each server, including the selection of CPU, installed memory and other expansion devices, such as GPUs.

Traditional [power estimates include rack-based](#) and nameplate-based approaches.

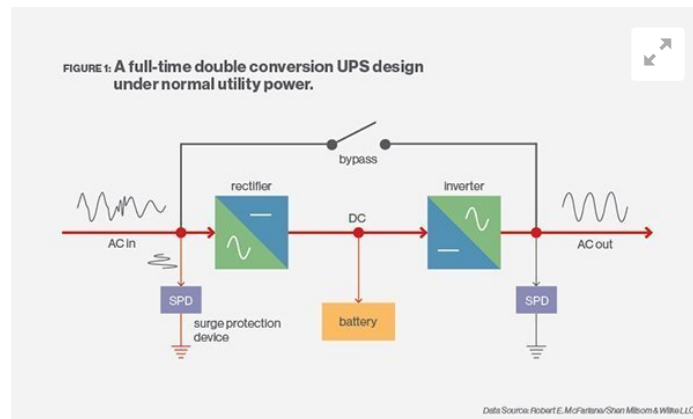
The rack-based approach generally assigns a standardized power-per-rack estimate. For example, an IT leader might assign an estimate of 7 kW to 10 kW per rack. If the data center plans to deploy 50 racks, the power estimate is a simple multiple. A similar approach is a general estimate of data center in watts per square foot (W/ft²). However, since this approach pays little attention to the equipment installed in each rack, it is often the most inaccurate means of power estimation.

The nameplate-based approach enables IT leaders to add up the power requirement listed on the nameplate of each server or other IT device. This is a more granular approach and can typically yield better estimates. Still, the power demand listed on each device nameplate can be notoriously inaccurate and doesn't consider the actual work the device is doing.

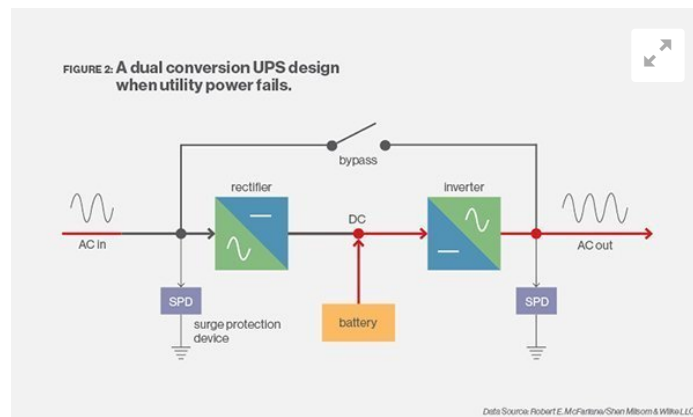
A more recent approach is to use actual power measurements per server, taken with IT power-handling devices, such as [intelligent power distribution units \(PDUs\)](#), located within each rack. Actual measurements can yield the most accurate estimates and give data center operators a better sense of how power demands and costs can fluctuate with workload demands.

Finally, utility power will inevitably experience occasional disruptions in generation and distribution, so data centers must include one or more options for redundant or backup power. There can be several layers of secondary power put in place, depending on which issues the business intends to guard against.

At the facility level, a data center can incorporate diesel or natural gas-powered [backup generators capable of running](#) the entire facility over the long term. Backup power can be supplemented by local renewable energy sources, such as solar or wind farms. At the IT infrastructure level, racks can incorporate UPS options, which provide short-term battery backups to enable an orderly system shutdown when power disruptions become unavoidable.



A full-time double conversion UPS design under normal utility power.



A dual conversion UPS design when utility power fails.

■ Data center cooling systems

The power delivered to a data center is translated into work performed by the IT infrastructure, as well as an undesirable byproduct: heat. This heat must be removed from servers and systems, and then exhausted from the data center. Consequently, cooling systems are a critical concern for data center designers and operators.

There are two primary cooling issues. The first issue is the [amount of cooling required](#), which ultimately defines the size or capacity of the data center's HVAC subsystems. However, designers must make the translation from the data center's power demand in watts (W) to cooling capacity gauged in *tons* (t) -- i.e., the amount of heat energy required to melt one ton of ice at 32 degrees Fahrenheit in one hour. The typical calculation first requires the conversion of watts into British thermal units (BTU) per hour, which can then be converted into tons:

$$W \times 3.41 = \text{BTU/hour}$$

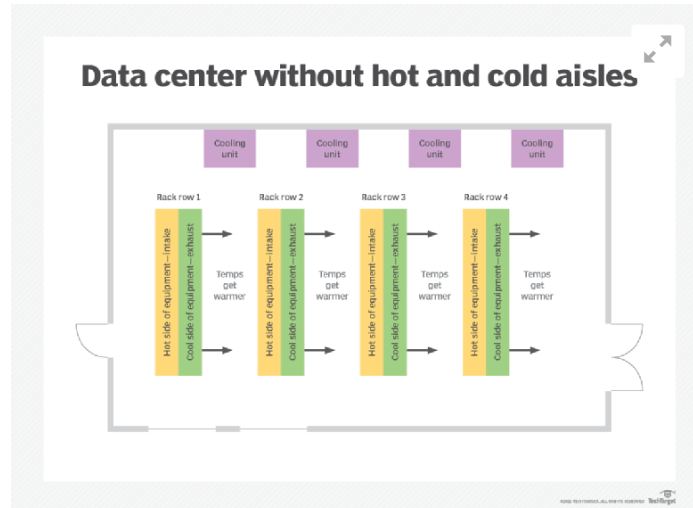
$$\text{BTU/hour} / 12,000 = t$$

The key is understanding the data center's power demands in watts and planned scalability, so it's important to right-size the building's cooling subsystem. If the cooling system is too small, the data center can't hold or scale the expected amount of IT infrastructure. If the cooling system is too large, it poses a costly and inefficient utility for the business.

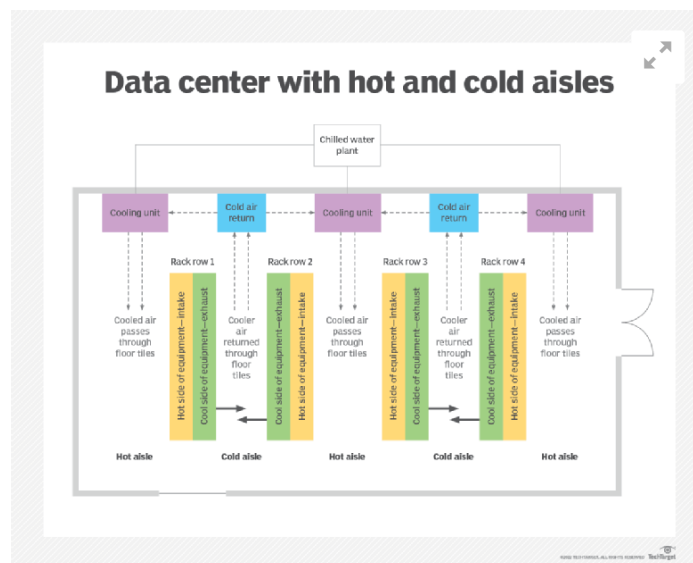
The second cooling issue for data centers is efficient use and handling of cooled and heated air. For an ordinary human space, just introducing cooled air from one vent and then exhausting warmed air from another vent elsewhere in the room causes mixing and temperature averaging that yields adequate human comfort. But this common home and office approach doesn't work well in data centers, where racks of equipment create extreme heat in concentrated spaces. Racks of extremely hot gear demand careful

application of cooled air, and then deliberate containment and removal of heated exhaust. Data center designers must take care to avoid the mixing of hot and cold air that keeps human air-conditioned spaces so comfortable.

Designers routinely address server room air handling through the use of containment schemes, such as hot aisle/cold aisle layouts. Consider two rows of equipment racks, where the rears face each other (see second diagram below). Cold air from the HVAC system is introduced into the aisles in front of each row of racks, while the heated air is collected and exhausted from the common hot aisle. Additional physical barriers are added to prevent the heated air from mixing with the cooled air. Such containment schemes offer a very efficient use of HVAC capacity.



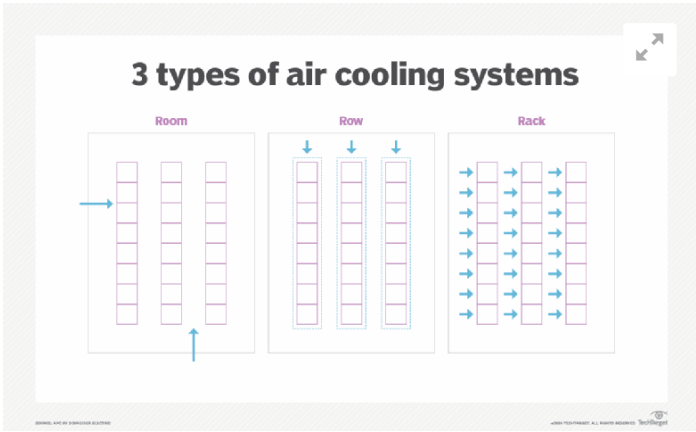
In data centers that don't use a hot/cold aisle design, the cooling units aren't always able to efficiently cool equipment.



In data centers designed around hot/cold aisles, the cooling units are able to more efficiently cool equipment.

Other approaches to cooling include *end of row* and *top of rack* air-conditioning systems, which introduce cooled air into portions of a row of racks and exhaust heated air into hot aisles.

Some data centers even embrace emerging liquid cooling technologies that immerse IT gear in baths of chilled, electrically neutral liquids, such as mineral oils. The liquid chiller is small and power-efficient, and [liquids can offer many times more heat transfer efficiency than air cooling](#). However, liquid cooling faces other challenges, including leaks/flooding, part corrosion or susceptibility to liquid intrusion, liquid filtering and cleanliness, and human safety.



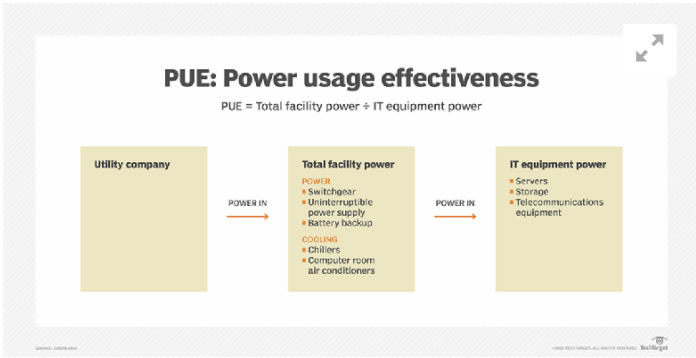
This diagram shows the basic concepts of room-, row- and rack-based cooling architectures. The blue arrows indicate the relation of the primary cooling supply p the room.

■ Data center efficiency and sustainability

Today's concerns about the environmental impact of CO2 emissions from power generation have prompted many organizations to place new emphasis on the efficiency and sustainability of the data center.

Efficiency is fundamentally a measure of work done versus the amount of energy used to do that work. If all that input energy is successfully converted into useful work, the efficiency is 100%. If none of that input energy results in successful work, the efficiency is 0%. Businesses seek to improve efficiency toward 100% so that every dollar spent in energy is driving useful data center work.

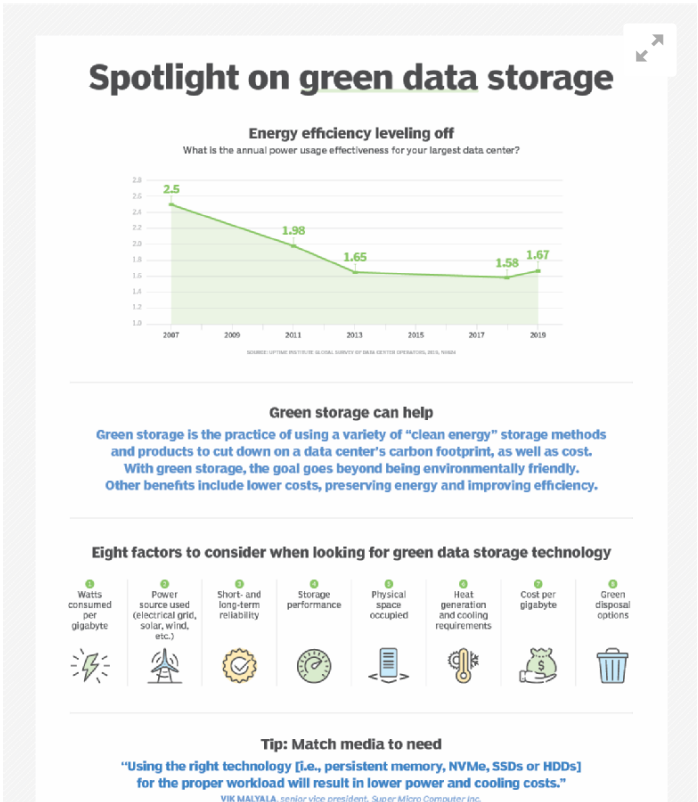
Measures such as power usage effectiveness ([PUE](#)) are available to help organizations gauge efficiency. PUE is calculated as the power entering the data center divided by the power used in the IT infrastructure. This yields a simple ratio that approaches 1.0 as efficiency approaches 100%, and the corresponding percentage is expressed as data center infrastructure efficiency. Businesses can improve the PUE ratio by reducing the amount of energy in non-IT uses, such as reducing lighting and cooling in non-IT spaces and implementing other energy-efficient building designs.



Power usage effectiveness is a metric used to assess the efficiency of a data cent

Sustainability is another concern. Power generation creates pollution that is believed to drive climate change and reduce the health of the planet. Creating a sustainable or [green data center](#) means to strive for net zero carbon emissions for the power that drives data centers. *Net zero* means that energy is obtained from renewable sources that add zero CO2 to the atmosphere.

In some cases, the business can choose to [approach net zero by using power from nonpolluting sources](#), such as solar or wind farms. In other cases, power can be purchased from power providers capable of capturing or recovering an equivalent amount of CO2 emitted in energy production, yielding net zero emissions. To achieve net zero, businesses must embrace energy conservation, energy efficiency -- such as PUE initiatives -- and renewable nonpolluting energy sources.



Reduce data storage energy consumption to run a greener data center.

■ **Data center design best practices**

There is no single way to design a data center, and countless designs exist that cater to the unique needs of each business. But the following strategies can [produce a data center design with superior efficiency and sustainability](#):

- **Measure power efficiency.** Data center operators can't manage what they don't measure, so use metrics such as PUE to oversee the efficiency of the data center. PUE should be a continuous measurement taken at frequent intervals, year-round, as seasons and

weather can affect power usage.

- **Revisit airflow.** Cooling is essential to safe operation of IT infrastructure, but airflow must be managed and optimized. This can include limiting hot air/cold air mixing, using hot aisle/cold aisle containment schemes and even using blanking plates to cover unused rack openings, which prevents cooled air from flowing to places that don't cool any gear.
- **Raise the temperature.** The colder a server room is, the more power-hungry and expensive it is. Rather than keeping the server room colder, evaluate the effect of actually raising the temperature. For example, rather than running a cold aisle at 68 to 72 degrees Fahrenheit, consider running the cold aisle at 78 to 80 degrees Fahrenheit. Most IT gear can tolerate elevated temperatures in this way.
- **Try alternative cooling.** An HVAC system might be standard for data centers, but consider ways to reduce or eliminate dependence on conventional HVAC. For example, data centers in cooler climates might reduce HVAC use and introduce cooler outside air -- termed *free cooling* -- into the facility. Similarly, HVAC can be supplemented or replaced by water-cooled chillers -- i.e., *economizers* -- or other heat-exchange technologies that use far less energy.
- **Improve power distribution.** Data center power efficiency is often lost due to [inefficiency in power-handling](#) and distribution devices, such as equipment transformers, [PDUs](#) and UPS gear. Use high-efficiency power distribution gear and minimize gear -- both of which result in fewer steps and opportunities for loss.

■ Data center design challenges

Although there is no single uniform formula for data center design and construction, there are numerous perpetual challenges faced by data center designers and operators. Below are several broad considerations and challenges:

- **Scalability.** A data center is a long-term installation that can remain in service for decades, but the data center operating today may differ significantly from the data center operating one or two decades from now. Designers must consider ways of handling today's workloads and services, while also considering how those resources should scale well into the future. The challenge is in providing room for growth in space, power and cooling while mitigating the costs of such capacity until it's needed.
- **Flexibility.** A data center is a bit like a heavy manufacturing floor: Equipment is put into place but can be almost impossible to move and change as demands evolve. An inability to move gear and shift aisles can prevent businesses from adapting and changing to meet new business demands. The challenge is meeting the needs for change without downtime or costly and time-consuming
- **Resilience.** A business relies on its data center. If the data center doesn't work, the business doesn't work. Power disruptions, network disruptions, environmental catastrophes and even hacking and other acts of malice can take down a data center. Designers face the challenge of understanding the most prevalent threats and designing appropriate resilience to meet those threats.
- **Change.** New computing technologies and new requirements are always being developed and introduced. Data center designers must consider how to adapt and incorporate often-unforeseeable changes without the need to fundamentally redesign the IT

Data
Center



infrastructure to accommodate each change.

■ Data center infrastructure management software and tools

Data centers are complex organisms that require continuous monitoring and management at both the facility and IT infrastructure levels. Data center operators typically [employ DCIM tools](#) to provide perspective into the operation of both the facility and infrastructure. An array of common management tasks needed to operate a data center includes the following:

- **Observation and oversight.** Observation tasks include [monitoring power, temperature and humidity](#) within the facility. Observation within the infrastructure can include available capacity, meaning which systems are utilized and which are free; application health to monitor the proper operation of key enterprise workloads; and overall uptime or availability. Observation tasks are commonly linked to alerting and ticketing systems for prioritizing and remediating problems as they are detected.
- **Preparation and remediation.** Data center management also involves preparation tasks, such as disaster recovery and backup processes, as well as workload migration capabilities for enabling timely system service tasks. Remediation tasks include routine service, along with periodic system upgrades and ad hoc system troubleshooting and repair.
- **Capacity and capability.** Data center management is also about planning for the future. Management tools can oversee current capacity -- i.e., used versus free resources -- and help data center operators track utilization in order to plan for more capacity. They also help support regular improvements to data center capability, such as system upgrades, technology refreshes and the introduction of new data center technologies.

Management is a pivotal element in business service assurance and service level agreements (SLAs). Many data centers are bound by some form of SLA -- either to internal departments or divisions or to external business partners and customers. Monitoring and management with DCIM and other tools are essential in [guaranteeing adherence to an SLA](#) or identifying SLA violations that can be promptly isolated and remediated. In addition, comprehensive monitoring and management help ensure business continuance and disaster recovery, which can be vital for today's regulatory compliance obligations.

🔍 Dig Deeper on Data center design and facilities

Data center cooling systems and technologies
and how they work

power usage effectiveness (PUE)

By: Julia Borgini

How to calculate data center cooling requirements

By: Julia Borgini

-ADS BY GOOGLE

By: Alexander Gillis

How to choose the right DCIM tool for better configuration

By: Robert McFarlane

Latest TechTarget resources

WINDOWS SERVER

CLOUD COMPUTING

STORAGE

SUSTAINABILITY AND ESG

SearchWindowsServer

How to use ChatGPT with PowerShell to write code for admins

Use of artificial intelligence tools is on the rise in many industries, and administrators can tap into these digital helpers to ...

How to work with the new Windows LAPS feature

Microsoft updated this automated method to manage and back up passwords for local administrator accounts on Windows desktop and ...

About Us	Contributors	Guides
Editorial Ethics Policy	Reprints	Opinions
Meet The Editors	Answers	Photo Stories
Contact Us	Definitions	Quizzes
Advertisers	E-Products	Tips
Partner with Us	Events	Tutorials
Media Kit	Features	Videos
Corporate Site		
All Rights Reserved, Copyright 2000 - 2023, TechTarget		
Privacy Policy		
Do Not Sell or Share My Personal Information		

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.