Data Center Infrastructure Design

Overview

As a data hub and application carrier, data center is a basic guarantee facility for carrying digital computing power and information systems of various industries, an important precondition for building information platform, and a key link in the development of digital industry chain. Data center infrastructure is also an important part of ensuring stable operation of data centers. Troubleshooting, root cause analysis and remediation of facility problems are common challenges for any infrastructure operation. Therefore, data center infrastructure design becomes critical.

Many emerging trends are impacting all operations within the current data center, such as the increased data center rack density, deployment of virtualization, migration to hosted facilities, and cloud computing. Datacenter infrastructures are changing dramatically. IT managers must move out of their comfort zones to consider higher densities and power consumption, cooling resources, and physical space in the data center, as well as data center operation security.

Data Center Types

Enterprises data centers

A private facility built, owned and operated by an enterprise for internal use. Enterprise data centers include physical computing equipment such as servers, network systems and storage devices, as well as supporting infrastructure such as power, cooling and environmental monitoring systems.

Managed services data centers

Equipment and infrastructure leased by a third-party servicer that offers features and functions similar to a standard data center through a managed services platform (MSP).

Colocation data centers

A data center where white space is leased to individual companies to host their hardware equipment. The colocation provider will provide the building, cooling, power, bandwidth and physical security, while the customer provides the servers and storage.

Cloud data centers

A non-local data center in which users can rent infrastructure managed by a third-party partner and access data center resources over the Internet. The cloud provider is responsible for managing the servers, storage, and network elements.

Edge data centers

Usually placed close to the application users. Edge data centers primarily provide cloud computing resources and cache content to end-users and can allow content and services to be delivered to local users with minimal latency.

Data Center Main Components

Basic IT Infrastructure

IT racks, also known as server racks, LAN racks, or network racks, are used to store critical IT systems and components. Rack size is determined by the type and quantity of rack-mounted equipment and is measured in rack units, U. The standard size for a rack unit is 1.75 inches (44.45 mm), and then the 19-inch IT rack is a standardized size frame or enclosure for mounting equipment as specified in International Electrotechnical Commission (IEC) EIA-310E. Standard-sized IT racks are often chosen because they can accommodate standard-sized components.

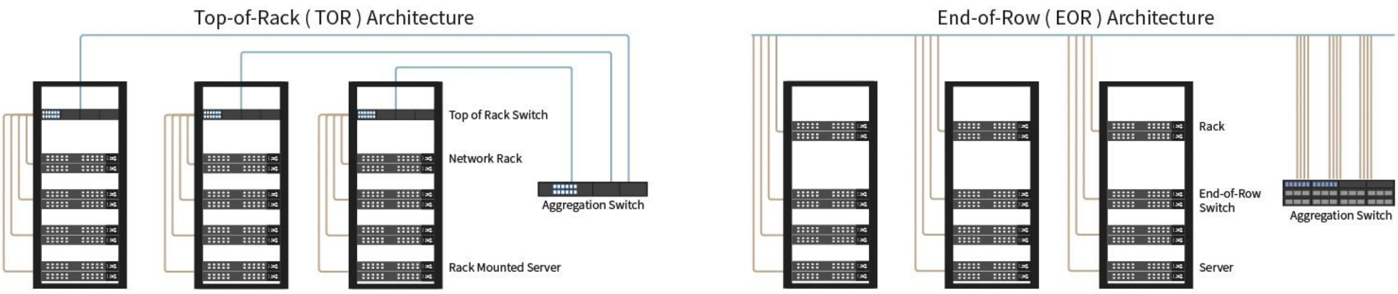
The cabling system of the data center is mainly to connect the devices to each other, so as to realize the mutual communication between the equipment. In the cabling system of the data center, the server cablings methods include the TOR (Top of Rack) architecture and the EOR (End of Row) architecture.

TOR (Top of Rack) architecture

TOR architecture refers to that the fiber patch cables in the server cabinets are connected directly to the switches placed on top of severs cabinets, and then directly to the core switch via the uplink port of the copper or fiber switches.

EOR (End of Row) architecture

The EOR architecture means that the access layer switch installed at the end of the server cabinet completes a permanent connection to other facilities in the cabinet through horizontal cables. The EOR architecture requires a large number of horizontal cables to connect the devices and switch, it also needs to extend the copper or fiber optic cables installed in the patch panel to the edge cabinet where the access switch is installed.

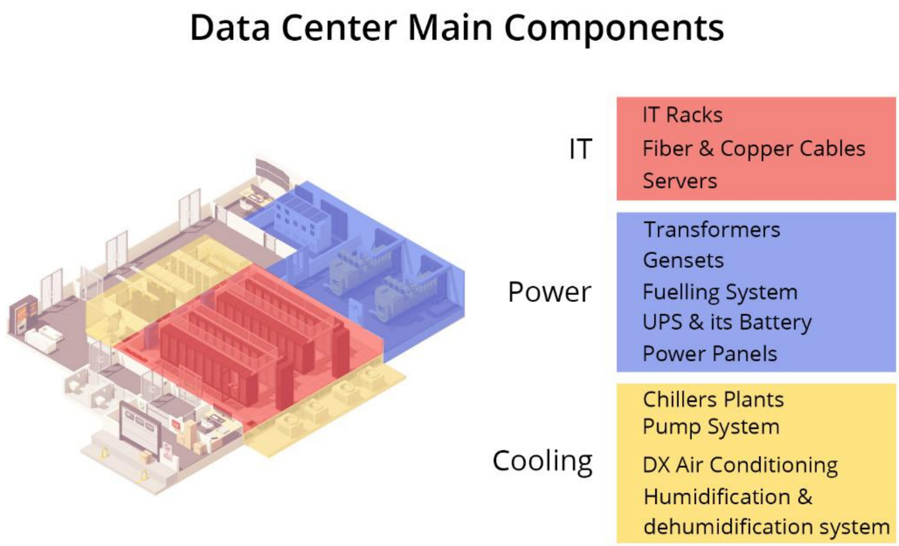


Power Infrastructure

Data center power facilities need to be equipped with one or more transformers to receive power and ensure that the voltage and current types of the input power are accurate. In addition, the uninterruptible power supply (UPS) is one of the important components of data centers that can effectively prevent unplanned power outages. Generators can also be used for power supply supplementation, although this type of power supply is usually fueled by gasoline or diesel.

Cooling Infrastructure

The cooling system is the core facility that provides cooling, heating, humidification, dehumidification, air supply, filtration, and other functions. Data center central air conditioning is provided by chiller units with a double-pipe design to improve the reliability of the system. The air conditioning units in the server room are divided into the lower air supply and the upper air supply according to the air supply method. Also, they are divided into the direct expansion (DX) set and chilled water (CW) set according to whether they have their cooling source. The DX set has its refrigeration system, and the CW set does not have its refrigeration system and needs to use the low-temperature chilled water provided by the chiller to provide the cooling source.



Data Center Systems and Solutions

Physical equipment solutions

When the disorganized distribution of equipment in data centers resulting in high capital and operating costs, the space will become one of the biggest constraints on data centers. To solve this problem, enterprises must optimize the use of existing space. For example, integrating white space and gray space in the data center as well as cable containment, grounding and boundaries, fire doors, aisle containment design, etc.

Physical security

For the design of data center security systems, the most basic protection to prevent the intrusion of outsiders are access control, CCTV, gates, revolving doors, wedge fences, lifting fences, fences, and access control systems. With the high level of sophistication and attacks on the modern data centers, protection strategies should be upgraded. Here are five effective ways to help maintain physical security in your data center: secure location, robust infrastructure & protection system, multi-layered access control, video surveillance, security control tests.

Fire detection and suppression

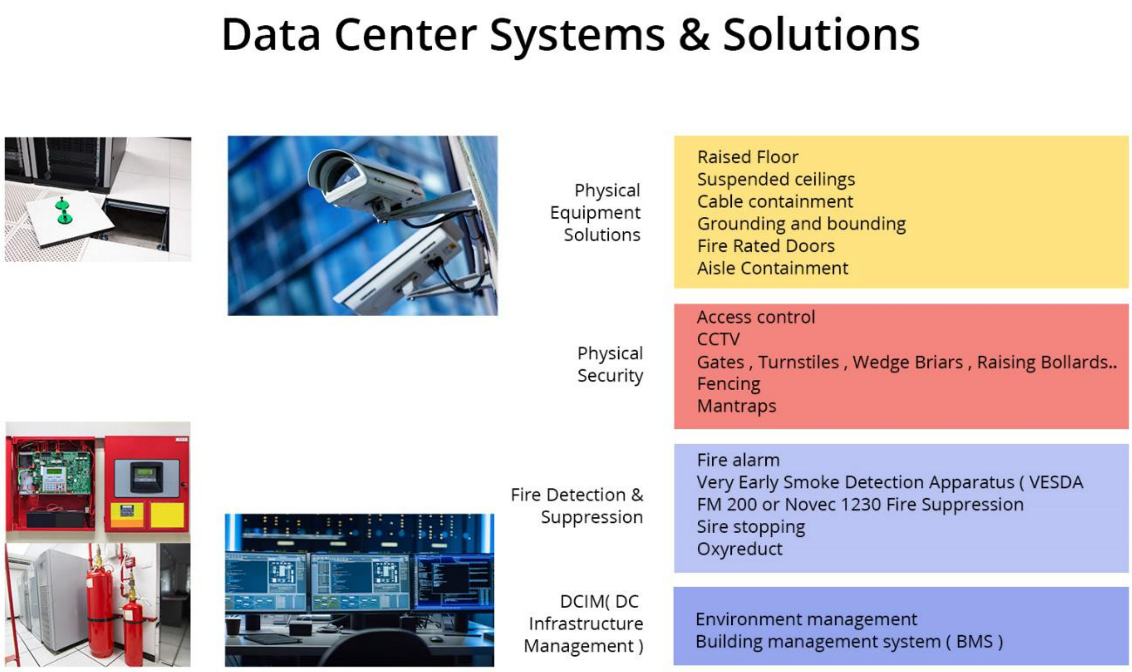
The fire suppression system involves various facilities such as fire alarms, VESDA, FM200 or Novec1230 fire suppression, bull stop, oxygen reduction, etc. Fire alarms are designed to detect and extinguish a fire as soon as it begins. In addition, VESDA systems and water or gas fire suppression systems are critical to data center operations, accurately detecting fires and providing timely alarms to keep data centers safe.

DCIM

DCIM stands for Data Center Infrastructure Management. DCIM software is commonly used to monitor, measure and manage data centers, covering IT equipment and infrastructure, such as power and cooling systems. It is designed to help data centers improve energy efficiency and prevent equipment problems that cause downtime.

DCIM provides the following advantages:

- Access to accurate, actionable data generated during data center operations.
- Stay informed about changes to program standards.
- Facilitate asset management.
- Effective forecasting and planning for space, power, and cooling capacity.
- Efficiently understand the operational status of power and cooling infrastructure and environment, improving overall data center availability.
- Reduce operating costs by improving energy efficiency and effectiveness.



Tiers Topology Categories

A data center tier is a system used to describe a specific type of data center infrastructure in a consistent way, with a total of four layers of structure. Tier 1 is the simplest infrastructure, while tier 4 is the most complex and has the most redundant components. Each tier includes the necessary components for all the layers below it.

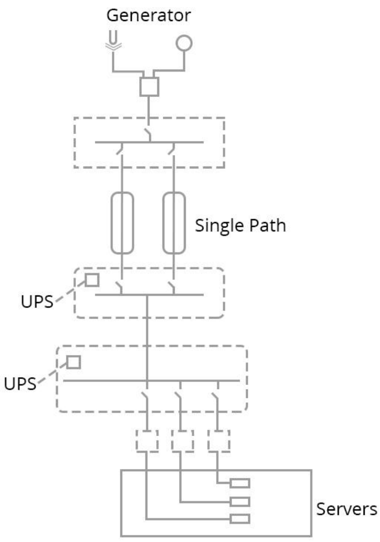
Tier 1 data center has a single power and cooling path and no backup components. The expected uptime for this tier is 99.671% per year. The facilities include：

- 2 Uninterruptible power supplies (UPS)
- IT system areas
- Dedicated cooling equipment that operates outside of office hours
- Engine generators

Tier 1
Basic Capacity

N = 2 UPS modules

Single path



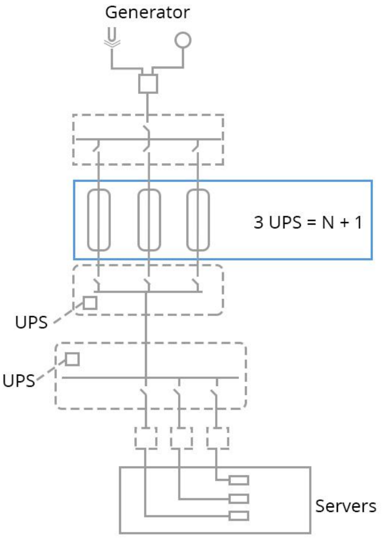
Function	Tier I
Active Capacity Components to Support the IT Load	N
Distribution Paths	1
Concurrently Maintainable	No
Fault Tolerance	No
Compartmentalization	No
Continuous Cooling	No

Tier 2 data center still has a single power and cooling path but some redundant and backup components. This tier provides 99.741% of expected uptime per year. The facilities include：

- Engine generators
- Energy storage
- Chillers and Cooling units
- 3 UPS modules
- Pumps
- Heat dissipation equipment
- Gasoline tank and Fuel Cell

Tier II Redundant Components

N = 2 UPS modules
Single path
Adds redundant UPS module



Function	Tier II
Active Capacity Components to Support the IT Load	N+1
Distribution Paths	1
Concurrently Maintainable	No
Fault Tolerance	No
Compartmentalization	No
Continuous Cooling	No

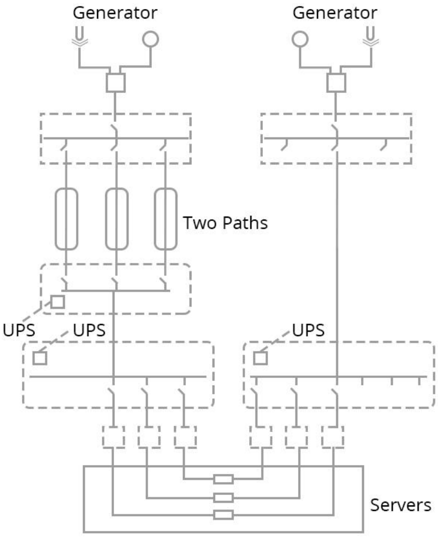
Tier 3 data center has multiple power and cooling paths and a redundant system. Expected uptime for this floor is 99.982% per year. Tier 3 facilities require all of the components present in a Tier 2 data center, but these facilities must also have N+1 redundancy. The "N" is the capacity required to support the full IT load. The "+1" represents the additional components used for backup.

Tier III Concurrently Maintainable

N = 2 UPS modules

Adds second critical distribution path

Adds means to maintain UPS system



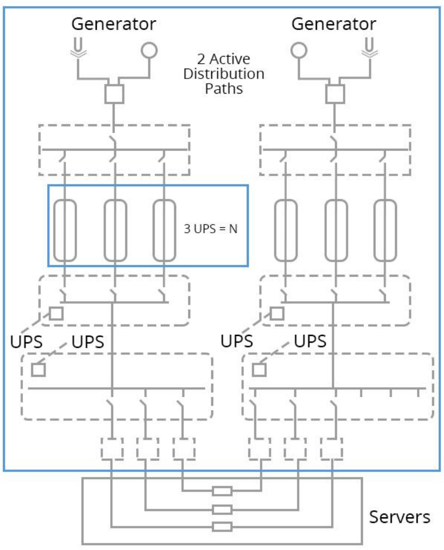
Function	Tier III
Active Capacity Components to Support the IT Load	N+1
Distribution Paths	1 Active & 1 Alternate
Concurrently Maintainable	Yes
Fault Tolerance	No
Compartmentalization	No
Continuous Cooling	No

Tier 4 has a fully fault-tolerant data center with redundancy for every component. The expected uptime for this tier is 99.995% per year. It can realize 2N or 2N+1 Redundancy. 2N or N+N redundancy indicates that tier 4 has a fully mirrored and independent system as a backup. If there is a condition in the primary component, the redundant equipment will immediately act to keep the facility up and running. 2N+1 redundancy provides twice the operational capacity and an additional backup component in case the second system fails while it is running.

Tier IV Fault Tolerant

N = 3 UPS modules

Dual path



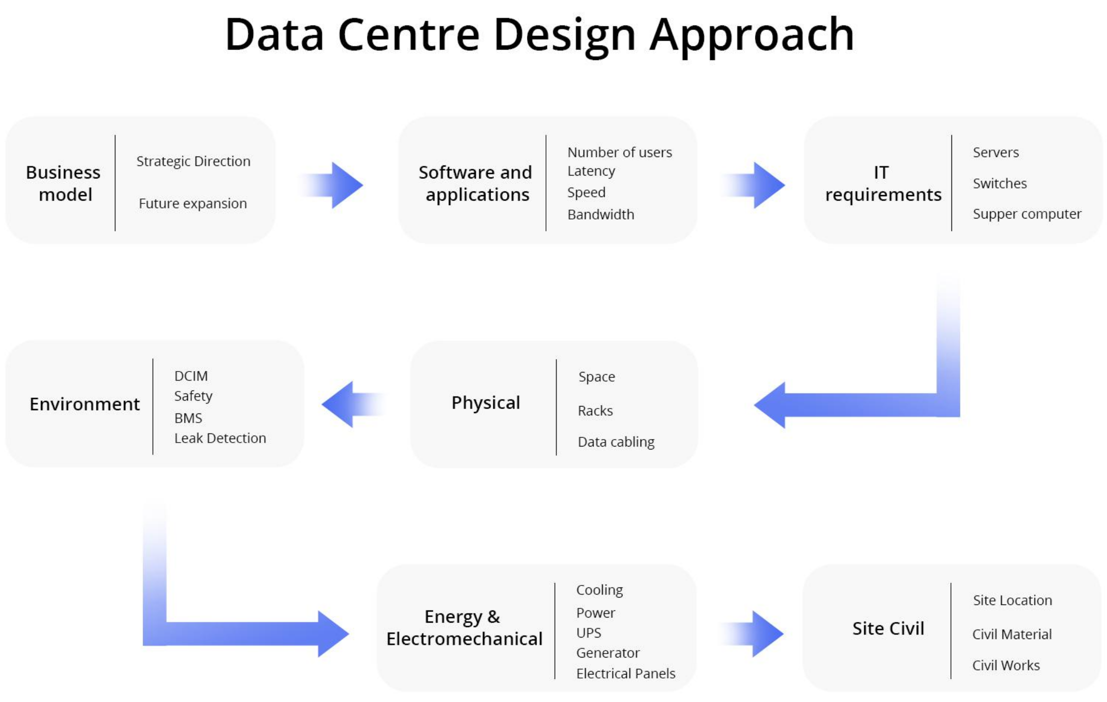
Function	Tier IV
Active Capacity Components toSupport the IT Load	N After any Failure
Distribution Paths	2 Simultaneously Active
Concurrently Maintainable	Yes
Fault Tolerance	Yes
Compartmentalization	Yes
Continuous Cooling	Yes

Data Center Design Approach

The data center designer should first consider the strategic direction of the enterprise and future business expansion, and then clarify the degree of matching between the strategic direction and the business needs. The next is the application design. Data center designers need to know the number of users to choose reasonable applications. Of course, network bandwidth and transmission speed should also be taken into consideration to avoid network delays and improve user experience. For IT infrastructure, the main thing is to plan the usage and backup quantity of equipment, such as servers, switches, supercomputers, etc. Finally, it is the racks and cabling systems quantity, as well as the efficient use of space.

To ensure the stable operation of the infrastructure, the data center environmental monitoring solutions should be designed according to the data center architecture. DCIM (data center infrastructure management) software can be a good choice to complete part of the environmental monitoring to prevent data vulnerabilities. In addition, BMS (Building Management System) as a building IT control system possesses performance of controlling and monitoring mechanical and technical elements. We know that data center cooling, power, UPS, and generator facilities are critical components. Using green energy can minimize business operating costs to the greatest extent.

At last, a good data center location can better protect the data center and extend its life of the data center. Geographic and power factors are major considerations when sitting in a data center for avoiding natural disasters can prevent force majeure factors and reduce losses while electricity is one of the biggest costs of data center operation. Taking into account quality and capacity requirements, we can choose an empty location where electricity is relatively cheap. Renewable energy sources can be a good choice for reducing energy consumption and helping enterprises build a more environmentally friendly corporate image. Additionally, other factors such as local data protection laws, tax structure, and land policies need to be considered as appropriate.



Data Center Design Recommendations

Comply: Comply with the standards

Use: Use modularity for easy maintenance and upgrades

Consider: Consider affordable capital and minimize life cycle cost

Target: Target a compact and efficient footprint or layout

As cloud computing continues to heat up and Internet applications increase dramatically, data centers need to configure more information equipment to meet application demands, which also increases power supply, cooling, and other infrastructure equipment. Hence, it results in the energy consumption and operation and maintenance costs of data centers reaching unprecedented levels. Modular data centers can control energy consumption and reduce human resource consumption in the later operation phase through centralized management of IT facilities, effectively controlling the overall operation cost of data centers.

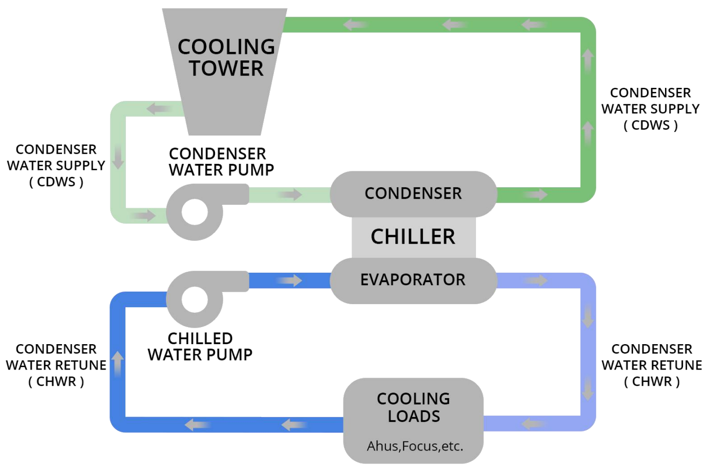
The temperature has an impact on the operational efficiency of a data center. It is recommended to always adopt a compact data center with an efficient footprint and layout, as every space within the data center requires investment. We need to focus on two points of the data center energy efficiency for making the most of the space within the data center: the cooling system and the power system.

Data Center Cooling Technologies and Design

Data centers have to operate at an optimum temperature for highest equipment efficiency. 18°C-27°C is the recommended inlet temperature of IT equipment by ASHRAH (American Society of Heating, Refrigerating, and Air Conditioning Engineers). Here are four cooling technologies as follows.

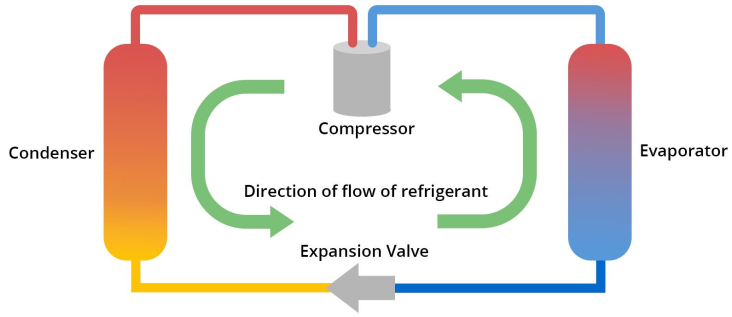
Chilled water system

A chilled water system acts as a centralized cooling system to provide cooling for an entire building or even multiple buildings. The chilled water system principle involves providing cooling to a building by using chilled water to absorb heat from the building space. Chillers can be located at the heart of the water-cooling system to remove heat from the water through the refrigeration cycle. There are two main types of chilled water systems: air-cooled chillers and water-cooled chillers. Air-cooled chillers are typically placed outside the building and remove heat from the chilled water by discharging it directly into the surrounding air. Water-cooled chillers are generally placed inside the building. They work in almost the same way as air-cooled chillers. The difference is that they remove heat from the chilled water by discharging it into a second, separate water line called the condenser water line.



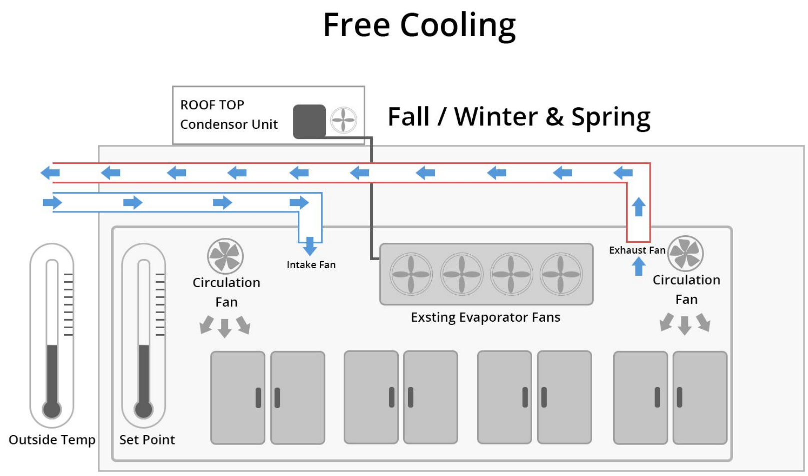
Direct expansion

Direct expansion (DX) systems are among the most compact refrigerant systems. Instead of using a secondary coolant circuit, the direct expansion system works by using refrigerant to cool the desired payload directly through the cooling plate. The direct expansion system consists of the four basic components of a vapor compression refrigeration system: compressor, condenser, expansion valve and evaporator. In the DX system, the evaporator absorbs heat directly from the payload.



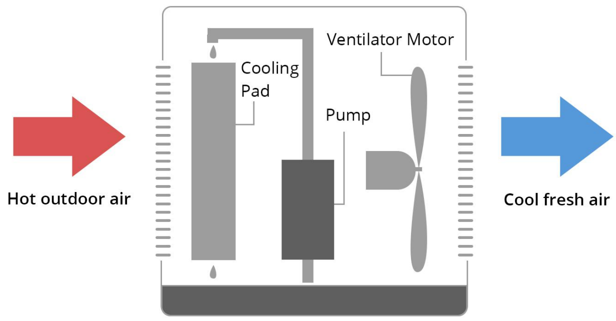
Free cooling

Free cooling is a process that uses the external ambient temperature to dissipate heat. It is effective when the temperature difference between the external water supply and return water is as small as 1°C. It means that in a 24/7 data center with a typical room temperature of 24°C, a natural cooling system can be used for more than 95% of the year. Natural cooling is typically achieved by installing a naturally cooled water coil near the DX coil. The chiller monitors the external ambient temperature. Once the temperature falls below the return water temperature, the water is diverted to the natural cooling coil to reduce or eliminate the need for mechanical cooling. In addition to saving energy by reducing the need for mechanical (DX) cooling, simultaneous natural cooling maximizes the part-load efficiency of components such as electronically commutated (EC) fans, variable frequency drive pumps, and centrifugal compressors.



Evaporative cooling system

The heart of a modern evaporative cooling system is the cooling pad, where the water evaporates and then the air that passes through is cooled. A pump that delivers water to the cooling pad continuously wets the pad. Depending on the effectiveness of the evaporative medium, the exhaust air can be cooled to 60 to 90 percent of the wet bulb. Evaporative cooling systems, either indirect or direct, are a highly sustainable and energy efficient method of cooling that can provide comfort and efficiency to production facilities, distribution centers and office buildings.



Cooling system design

Cooling systems typically include cold and hot aisles, cold aisle containment and hot aisle containment. The cold air rises through the server from the floor, while all the hot exhaust air collects in the hot channel and rises to the ceiling, where it is sent to the computer room air conditioning (CRAC) unit. This principle indicates that the server receives only fresh cold air, while the CRAC unit receives hot exhaust air, which increases the temperature difference in the heat exchanger of the CRAC unit, thus increasing the efficiency of the machine. Cold air containment is a very popular choice because it is easy to implement and very low cost. Cold air fills the cold aisle and then the hot air is discharged to the rest of the room, where the air conditioning unit pulls it in for retrofit. Another containment strategy is hot aisle containment, where cold air fills the room while hot exhaust air is pushed into another void within the ceiling. Hot aisle containment provides superior performance and a slight cooling buffer in the event of electrical or cooling system failure.

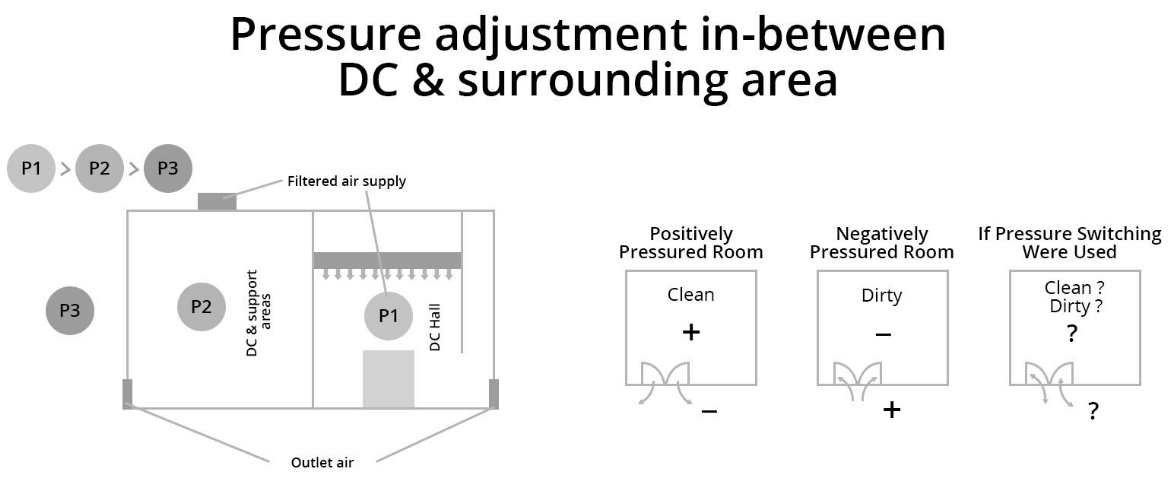
In-row cooling allows precise cooling and conditioning of the air in the server cabinet. It can be installed closer to the actual rack, such as on the floor or suspended overhead. In this way, neither the cold air nor the hot exhaust air travels very far and the equipment can dissipate high heat loads faster.

Downflow air handling methods send the server exhaust air to the top of the air handling unit (AHU), where it is treated by cooling coils and blown by internal fans to the data center's under-floor ventilation system. Downstream units are deployed using a combination of two cooling systems from chilled water, an evaporative cooling system, or free cooling. Up-flow works oppositely to downflow units, rarely deploying in new data centers anymore because of the advancement in aisle containment technology and data center design.

Direct chip cooling uses a liquid or phase change heat transfer mechanism that runs over the silicon surface of the chip or in proximity to the chip. This method provides maximum power and heat dissipation at the lowest operating temperatures allowing the processor to run significantly faster than processors cooled by other conventional methods and traditional methods.

Pressure Adjustment In-between DC Surrounding Area

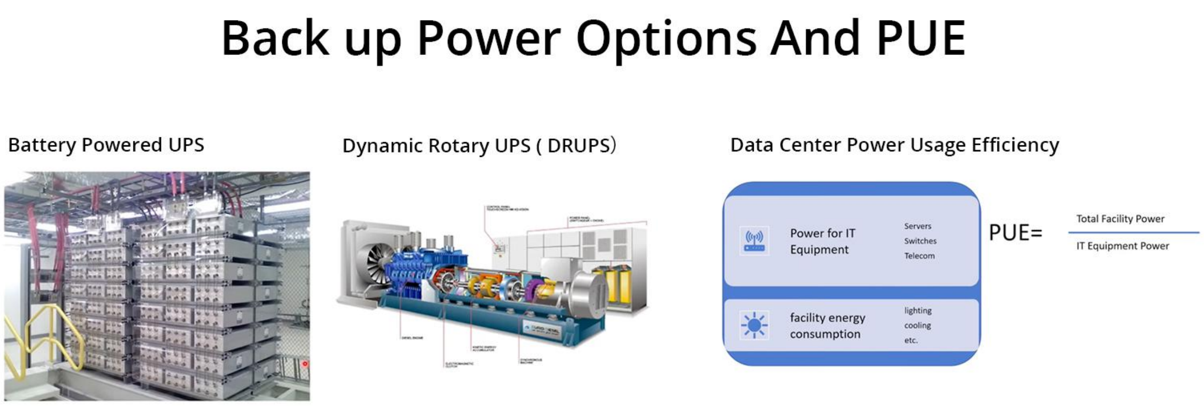
In terms of pressure and air concepts, the data center design in such a way that the pressure inside the hole is greater than the support area, and the pressure in the support area is greater than the pressure outside the data center. If we open a door between areas P1 and P2, air will travel from inside the data bore to the support area as well as outside the facility. Air flows from the inside to the outside, so we always have an air supply that can push air into the data center units and there are air outlets running to extract air from the outside, which keeps the data center air flowing. So, this serves two purposes, the first is to have air circulation in the data; and the second is the air changes to increase the pressure in the room.



Backup Power Options

Actually, there are two options for uninterruptible power supplies (UPS) for data center power back up. Battery-powered UPS and dynamic or rotating UPS.

One of the main purposes of a data center is to provide maximum uptimes for all of its equipment. But electric utilities cannot guarantee 100% uptime for their electrical services, so data centers can stay up and running by using UPS. The main advantages of using UPS are to prevent power surges, provide uniform power and reduce electrical loads. Dynamic or rotary UPS with higher capacity and higher voltage generation possibilities (ratings from 200kW to 3000kW) are ideal for industrial environments. The PUE (Power Usage Effectiveness) of a data center is a measure of the energy efficiency of the data center. In other words, PUE evaluates the energy performance of a data center by calculating the ratio of overall energy usage to the energy used by IT equipment only. the PUE varies depending on the data center, its design and occupancy, and other criteria, including external temperature.



Summary

To sum up, data center infrastructure design should create the most secure operating environment by following data center environmental standards for those facilities placed in data center tiers. This requires us to have an in-depth understanding of data center design specifications and then take into consideration various aspects to achieve efficient data center management, including space planning, cooling technology, power systems, energy consumption, and security systems.

https://www.fs.com

The information in this document is subject to change without notice. FS has made all efforts to ensure the accuracy of the information, but all information in this document does not constitute any kind of warranty.