HUAWEI

Next-Generation Data Center Facility

White Paper

Foreword

The world is undergoing a rapid digital transformation. Innovative services, such as smartphones, smart homes, smart manufacturing, and autonomous driving, are reshaping our work and life. At the same time, with climate change accelerating and a global consensus on carbon neutrality, the pursuit of green and sustainable development has become a common goal.

Data centers are fundamental for the digital economy. However, data centers are currently drawing much attention from all segments of society due to their high energy consumption and carbon emissions. With surging demand for data center construction, information technology evolution and low-carbon requirements, the data center industry is undergoing profound changes, and its development will enter a new era.

What is a "next-generation data center facility" that meets the requirements of the new era? Huawei has worked with global data center industry leaders and technology experts to conduct extensive discussions on industry and technology trends, and reached many important consensuses, including the definition of next-generation data center facilities.

The future is here. It is believed that next-generation data center facilities, defined by the industry's wisdom, will play an important role in the high-quality and sustainable development of the industry.

Contents

01

Intelligent and Low-Carbon Technologies Drive the Rapid and Sustainable Development of Data Centers

Digitalization and low-carbonization are two unstoppable trends in the world. Data centers are the foundation of the digital world. By integrating digital and power electronics technologies, building smart and low-carbon data centers will become the mainstream trend.

1.1 Digital Economy Drives Fast Growth of Data Centers

Currently, digital technologies such as AI, cloud computing, big data, IoT, and 5G are evolving at a fast pace, fueling global GDP growth. According to China Academy of Information and Communications Technology (CAICT), the digital economy accounted for 43.7% of global GDP in 2020 and has expanded at an astonishing speed.

Major economies like China, the EU, and the USA have released digital economy visions.

China has set the digital economy as the main developing direction in the *14th Five-Year Plan (2021–2025) for National Economic and Social Development*.

The European Union has set out digital ambitions for the next decade in *2030 Digital Compass: the European way for the Digital Decade*, which states that by 2030, 75% of European enterprises will use cloud computing, big data and AI; more than 90% of European small and medium-sized enterprises (SMEs) will reach at least a basic level of digital intensity, and Europe will increase investment in renewable energy and digital infrastructure.

In the USA, the National Science Board has recommended increased investments in data, software, computing, and networks to maintain competitiveness in the digital economy in *Vision 2030*.

Nowadays, digital technologies have been evolving and reshaping the way we live and work. In the post-pandemic era, the digital economy has unleashed a powerful force to reshape the global economic structure, and reconstruct the global competitive landscape. Data center construction is expected to develop rapidly.

Data has become a key factor of production now, and the data volume keeps increasing dramatically. According to Huawei's *Intelligent World 2030*, computing power will increase ten times, and AI computing power will increase 500 times. The skyrocketing demand for computing power will unlock the significant potential of the data center market. According to the report of Synergy Research Group, which was released in March 2022, the number of global ultra-large data centers will be more than 1,000 in the coming three years.

1.2 Carbon Neutrality Raises New Requirements for Data Center Sustainability

With global warming problems, carbon neutrality has become a common mission. Until now, more than 130 countries have declared their commitment to carbon neutrality. Energy conservation and carbon emissions reduction have become key trends across all sectors.

In recent years, the data center industry is booming, its market space has been growing rapidly, and power consumption has increased dramatically. According to *Uptime Global Data Center Report 2021*, from 2014, the PUE of large data centers around the world has remained around 1.6 for seven consecutive years. The resource utilization of data centers is low and needs to be upgraded urgently. To advance the sustainable development of the data center industry, a number of countries and international organizations have released data center-related policies. For example, the U.S. government adopts the Data Center Optimization Initiative (DCOI), which requires that the PUE of new-built data centers be lower than 1.4 and that of upgraded old data centers be lower than 1.5. In the *Climate Neutral Data Centre Pact*, European data center operators and industry associations have announced that data centers will be carbon neutral by 2030.

China has released the *Guidance on Accelerating the Construction of the Collaborative Innovation System for National Integrated Big Data Centers* to promote the construction of national integrated big data centers. And the "Eastern Data and Western Computing" program has been launched to channel more computing resources from the eastern areas to the less developed western regions, which promotes the green and sustainable development of data centers, and accelerates the R&D and application of energy-saving and low-carbon technologies. It is required that the PUE of newly-built large data centers be lower than 1.3 by 2025.

Low carbonization of data centers will become an inevitable trend. Large-scale application of clean energy, development of carbon-reduction technologies including PV and energy storage, and waste heat recovery, will effectively advance data centers to achieve sustainable development.

Next-Generation Data Center Facility

02

In the face of the booming development of data centers as well as green and low-carbon requirements, Huawei and global data center industry leaders and technical experts have conducted in-depth and extensive discussions and reached many important consensuses, summarizing the four characteristics as Sustainable, Simplified, Autonomous Driving, and Reliable.

1. Sustainable: Innovative measures will be taken to build green and highly energy efficient data centers benefiting the society.
2. Simplified: In response to the increasing scale and complexity of data centers, the architecture and key subsystems will be minimalist through hardware convergence.
3. Autonomous driving: Digital and AI technologies will be leveraged to address the operation and maintenance challenges of massive data centers and automate data center facilities.
4. Reliable: Reliable operation is the cornerstone of a data center. Through modular redundant architecture and AI active prevention, we can guarantee the reliability of data centers at all levels from components and equipment to systems.

2.1 Sustainable

Looking back on the development of global modernization, we went through three industrial revolutions from "Steam Age", to "Electrical Age", and to "Information Age". These three industrial revolutions mainly relied on fossil energy. With the rapid economic growth, carbon dioxide emissions increased exponentially, which caused global climate change and brought great challenges to the survival and development of human society.

The next-generation data center facilities will be all green, efficient, and recyclable.

All Green: Utilizing Green Sources for Harmony with Nature

Green growth has become a global consensus. It is a sustainable development model based on green, low-carbon, and recycling technologies, which means that while maintaining sustained economic growth, it can reduce the damage to the environment.

Based on the concept of green growth, the next-generation data center construction must utilize green sources.

Electricity:
Renewable energy should be used on a large scale for data centers. The development and utilization of PV, wind power, and hydropower should be given a high priority in order to reduce fossil energy dependence. In addition, distributed PV can be deployed in a data center campus to fully utilize the rooftop and land resources of the campus. Distributed PV can reduce the carbon emissions of the data center by 1% to 2%.

Water:
Reducing the consumption of water resources requires minimal use of clean water, maximum use of reclaimed water, or even without using water. As water resources are increasingly scarce in many areas of the world, data centers need to save precious water resources. For example, all data centers in Ulanqab city of China are banned from using any groundwater for cooling systems.

Land:
The data center scale is increasing and the construction will use land resources intensively. Land should be efficiently used to support more computing power on each meter square of land.

Climate:
Data centers should use natural resources, such as external cold air and cold water. We can build a data center in an appropriate area or use technical measures, such as increasing the supply and return temperature to use more external cold air.

Harmonious coexistence between data centers and nature can be ensured by utilizing green resources.

All Efficient: PUE → xUE, One Dimension to Multi-dimensional System

The Green Grid proposed PUE as an indicator to measure the power efficiency of data centers in 2007, which has been widely accepted and used in the industry. However, PUE alone cannot fully reflect the resource utilization of data centers. For example, even with the same PUE, the carbon emissions from thermal power and green power are completely different, and the water resources consumed by chilled water systems and indirect evaporative cooling systems are also different. With a certain grid power capacity, the number of IT racks that can be deployed varies according to the solutions.

Therefore, the system evaluating whether resources are used efficiently should be changed from PUE-only to xUE, that is, a multi-dimensional system covering CUE, PUE, WUE, GUE, etc.

Carbon Usage Effectiveness (CUE) is a quantifiable indicator for the carbon emissions of data centers. Different energy sources emit different amounts of carbon dioxide. For example, coal emits 1023 grams of carbon dioxide per kWh, and photovoltaic emits only 30 grams per kWh. Therefore, CUEs of different sources of electricity in data centers vary greatly.

Water Usage Effectiveness (WUE) measures the water consumption of IT devices in a data center during cooling. It can be used to restrict solutions and products that consume a large amount of water.

Grid Usage Effectiveness (GUE) indicates the maximum number of IT devices that can be deployed with a certain grid power capacity. This indicator helps the industry to optimize products and deploy more IT devices.

Different regions and industries pay different importance to the corresponding indicators. Therefore, each indicator can be measured based on the characteristics of regions and industries, and finally an optimal evaluation scheme can be selected based on comprehensive consideration.

All Recyclable: Maximizing Resource Recovery Throughout the Life Cycle

A data center is a big energy consumer and heat emitter. The energy consumed by the data center is converted into heat and discharged into the air, which is not effectively utilized or recycled. In addition, hot spots can be formed around the data center, affecting the cooling and thus the PUE. Waste heat recovery is an important technology in the low-carbon era and it is one of the important features of next-generation data centers as well.

The European *Climate Neutral Data Centre Pact* (CNDCP) has clearly stated that heat recovery is one of the five key measures to achieve carbon neutrality in 2030. The European Sustainable Digital Infrastructure Alliance (SDIA) has put heat recovery on the roadmap, and the proportion of heat recovery in data centers should be greater than 60%. The recovered heat can be used for:

(1) The data center itself, for heating water in office areas and heating a diesel generator room
(2) Ancillary businesses around the data center, such as aquaculture or heating of a commercial complex
(3) The municipal heating network.

In addition to the waste heat recovery of the data center, the recyclability of the data center equipment and the materials of the facilities should also be considered. In the future, more and more regions will promote new prefabricated green buildings and green building materials, with the recyclability rate exceeding 80%. Lead-free and halogen-free green recyclable materials will be used to replace traditional harmful substances, such as lead in components, boards, and auxiliary materials, to achieve a higher recovery rate, and promote a low-carbon circular economy for data centers.

2.2 Simplified

To cope with problems such as slow construction and difficult O&M of traditional data centers, the continuous pursuit of simplicity drives the convergence of components, devices, systems, and data center architectures.

Simplified Architecture: Innovative Forms of Buildings and Equipment Rooms

GB 50174-2017 *Code for Design of Data Centers* defines a data center as "a building that provides a running environment for centralized electronic information devices. It can be one or several buildings or part of a building, including the computer room, auxiliary area, support area, and administration area." Data centers are complex engineering buildings. Traditionally, civil engineering, power supply, and cooling works are conducted one after another. The construction process is lengthy and subject to weather and design changes. In addition, different vendors are responsible for the implementation of different procedures. As a result, on-site construction companies and construction personnel are complex, which increases on-site management difficulty and makes it hard to ensure the construction progress and final quality.

In the future, with the increasing scale of data centers, convergent and simplified buildings and equipment rooms will meet fast service rollout requirements and come in line with the industry trend.

Building prefabrication: A data center is divided into parts, changing the building form and construction mode. The traditional "reinforced concrete + onsite construction" is transformed into "fabricated steel structure building + prefabrication". The engineering is product-oriented, enabling parallel data center construction, fast delivery, and on-demand deployment to meet requirements for fast service rollout in the cloud data center era.

For the traditional building construction, the whole process includes the serial works of the foundation, building, decoration, power supply and distribution, heating ventilation & air conditioning (HVAC), fire extinguishing, and monitoring systems plus installation and commissioning. The prefabricated modular construction can realize all-round parallel construction, which is embodied in two levels. The first is the parallel processes of foundation works and manufacturing, pre-integration, and pre-commissioning of functional modules. At the second level, standard modules are used, and multiple modules can be produced simultaneously in the factory. The efficiency is greatly improved through mass production processes. Modules are pre-integrated and pre-commissioned in the factory, which greatly improves the quality and supports quick onsite deployment and joint commissioning and delivery.

In China, for example, when the prefabricated modular construction mode is used to construct a 1000-rack data center, the construction period can be reduced from more than 18 months to 6–9 months. This difference is more obvious in the Middle East. Due to climate conditions, May to October is the hottest time. Outdoor construction is not allowed from 13:00 to 17:00 every day. The construction time of a large data center is longer than that in other regions. In the Middle East, the construction period of prefabricated modular data centers is shortened by 60% from 30 months to 12 months compared with the traditional construction mode.

The Wuhan AI computing center uses the prefabricated modular data center solution. The building was constructed in four months, and the project was completed in five months. The rollout time was shortened by more than 50%.

Data center modularization: A modular architecture can be used to reshape data center facilities. In traditional equipment room construction, parts such as racks, air conditioners, UPSs, power distribution cabinets, batteries, fire extinguishing devices, and cables are purchased, installed, and tested separately, which brings great challenges to construction, O&M, and energy efficiency management. The modular architecture integrates the racks, cooling, power supply, monitoring, and fire extinguishing subsystems into one module and isolates the hot and cold air. By changing the architecture of the equipment room, the delivery period is shortened, the O&M difficulty is reduced, and the energy efficiency is improved.

Simplified Power Supply: Redefined Components and Links

The trends of data centers are high-density and large-scale. The power supply system is the heart of data centers. To meet the increasing power density and large-scale requirements, the future innovation of data centers is to integrate and innovate power supply links to deliver a simplified power supply system.

Component integration: In the past, the power supply system of a large-scale data center mostly incorporates the UPS parallel system and lead-acid battery solution. There are many complex devices, which result in challenges such as complex on-site installation and commissioning, and a large footprint. Using a data center with 1500 racks and 8 kW/rack as an example, the area for the power supply system (including transformers, UPSs, power distribution cabinets, and batteries) is about 1800 square meters, accounting for 15% to 20% of the total data center area. As the power density of IT racks increases in the future, this proportion will continue to increase if the power supply system remains unchanged.

To meet the challenges of large footprint and complex on-site delivery and commissioning, the system space is reconstructed and optimized by all-in-one integration of originally distributed devices and innovative components without changing power supply links. For example, the switch miniaturization technology is used to reduce the number of switch cabinets without reducing the number of switches. On the other hand, the topology pooling and component optimization technologies are used to improve the power density of UPS modules, effectively reducing the footprint and simplifying onsite delivery.

Simplified link: In response to the call for carbon neutrality, clean energy applications, peak-valley electricity price differences, and virtual power plants (VPPs) will drive data centers to deploy PV and energy storage. In a traditional solution, PV systems (inverters and PV modules) and energy storage systems (converter boxes and energy storage boxes) are deployed on power supply links. More devices increase the complexity of the power supply system, which leads to challenges such as longer links and more difficult maintenance. A next-generation data center integrates all power links innovatively

CTTIC Cloud uses Huawei's PowerPOD solution, which saves the power supply system space by more than 40% compared with a traditional solution, helping the customer deploy 350 more IT racks and save more than 16,000 meters of power cables.

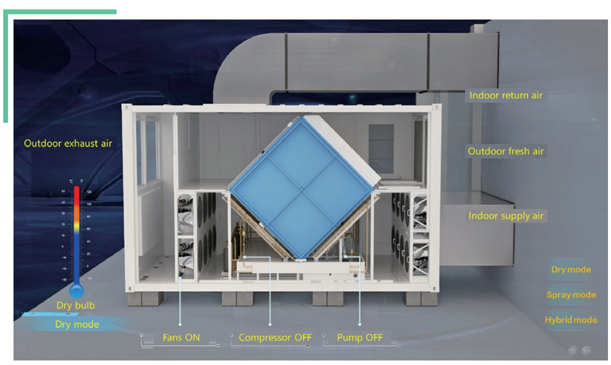
to make them simple. For example, the PV inverter and energy storage converter are integrated with the uninterruptible power system to provide a new medium-voltage uninterruptible power solution. A single system is used to connect the mains, PV, and energy storage at the same time, integrating the mains rectifier and inverter, PV inverter, and energy storage functions. The complexity of the power supply link is greatly reduced.

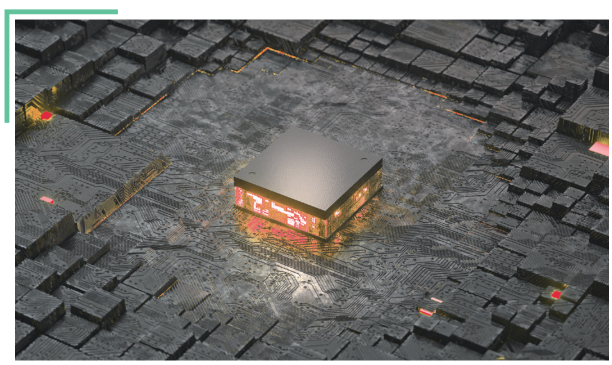
Simplified Cooling: Maximized Heat Exchange Efficiency

To achieve low carbonization, energy saving and consumption reduction in data centers lies in the simplified integration of cooling systems and the maximization of heat exchange efficiency.

Simplified cooling link: In the past, large data centers mainly used mechanical cooling systems. Taking a chilled water system as an example, which includes seven devices (chiller, cooling tower, cold storage tank, indoor unit, cooling water pump, plate heat changer, and management system), four heat exchanges take place in the whole cooling system. The project delivery lasts for more than half a year and is hard to implement. The construction quality affects the cooling efficiency. In the future, a data center cooling system will integrate multiple components into one module, shortening the delivery period and reducing O&M difficulty. In addition, a free cooling mode is available in this system, changing multiple heat exchanges to one, shortening the cooling link, and improving the cooling exchange efficiency.

Air-liquid convergence: As the power density of IT devices increases, especially in ultra-high power computing scenarios involving AI and supercomputing, the servers and chips require more capability and higher efficiency of the cooling system. Close-coupled cooling has become an important trend in the development of cooling systems. The cooling mode ranges from traditional room-level dispersive cooling, to module-level cooling in cold/hot aisle containment, to cabinet-level and chip-level liquid cooling, and finally to direct cooling for chips. In the future, in scenarios where the power density per rack is more than 20 kW, liquid cooling will be used as the primary and air cooling as the auxiliary approach. Converged air-liquid cooling will become the mainstream.





2.3 Autonomous Driving

As the scale of a single data center expands from thousands of racks to ten thousands of racks, the number of devices and their operation status to be monitored are increasing exponentially, causing great challenges in data center facility management.

1. **Relying on manual inspection.** Generally, 15 to 30 professional O&M engineers are required in a 1000-rack data center to perform 6 to 12 on-site inspections every day. The O&M cost is high and the inspection quality is difficult to guarantee.

2. **Passive response.** A data center generates massive alarms every day, including common, redundant, and root alarms. One fault may trigger hundreds of alarms, and it is difficult to find root alarms among them. As a result, fault identification becomes difficult. Traditional data center O&M relies on manual alarm priority identification, alarm root cause identification, and fault analysis, which leads to low O&M efficiency and may cause risks in manual operations.

3. **High operation gap.** Firstly, the actual operation PUE is not as good as the designed value, even though the designed one is attractive. Secondly, the average resource utilization of data centers is usually lower than 65%, causing much waste of resources.

To address the above challenges, data centers need more advanced management methods and systematic, automated, and intelligent technologies to improve management efficiency, and to realize autonomous driving in the data centers with O&M automation, energy efficiency optimization, and operation autonomy.

O&M Automation: Unmanned Inspection

As the scale of a single data center becomes larger, the O&M difficulty increases.

Take a data center with 1500 racks as an example. There are more than 100 types of devices and most of them are dumb, resulting in huge difficulty for manual inspection and much time for fault locating. In addition, O&M relies on labor mainly, leading to a growing proportion of labor cost. According to the report of *Uptime Institute global supply-side survey 2021*, the data center O&M labor cost increased from 4.5% in 2015 to 10% in 2020. On the other hand, the aging population of the world keeps increasing and labor shortage becomes a prominent issue, making it even harder for enterprises to employ proper O&M engineers. According to the above report of Uptime Institute, 47% of data center enterprises have difficulty in finding qualified O&M engineers.

Traditional manual O&M cannot meet complex O&M requirements of data centers. Automatic O&M will become an important feature of next-generation data centers. To build a digital visualized foundation, AI technologies will be applied to implement unmanned inspection through intelligent sensing, sound recognition, and image recognition. In addition, the O&M process will be standardized, and expert experience will be shared on a cloud platform and incorporated it into the process, quickly improving the skill sets of O&M engineers.

In the past, it took two hours for one engineer to finish an on-site inspection for a data center with 2000 racks. In the future, by using automatic inspection methods, such as indicator collection, camera image analysis, and infrared sensing, engineers can complete the inspection of 2000 racks in 5 minutes remotely.

Automatic Energy Efficiency Optimization: Enables Smart Cooling

Cooling system is the second largest power consumer in a data center, next only to IT devices. Challenges in traditional building automation (BA) systems include less than 10 collection parameters, less than 3 adjustable parameters, slow and inaccurate adjustment, 2 hours for each optimization, and short effect of adjustment. In addition, the traditional BA systems rely heavily on manual optimization and expert experience, with high skills required. For increasingly complex cooling systems, manual adjustment based on expert experience cannot be performed in real time based on environmental parameters (such as temperature and humidity) and IT load. In addition, to ensure system reliability, cooling requirements often expand, resulting in energy waste.

In the future, AI dynamic modeling technology will be used in data center energy efficiency optimization to build a machine learning model between energy consumption and varying parameters such as IT load, climate, and device running parameters. In this way, the energy consumption of each subsystem can be diagnosed in real time while device and system reliability is ensured. The technology automatically and accurately infers and configures the optimal control logic for a data center, and builds a PUE model through deep neural network (DNN) training to obtain appropriate instructions. An optimal cooling policy can be inferred from 1.4 million original combinations within 1 minute, and feedback on the effect is provided after the policy is delivered, achieving optimal energy efficiency of the data center.

AI energy efficiency optimization is used in Huawei Cloud Langfang Data Center. There are more than 700 collection points. After optimization, the annual average PUE decreases from 1.42 to 1.25, saving 33.36 million kWh or US$3.83 million ($0.115/kWh).

Operation Autonomy: Maximizes Resource Value

Resource Optimization @AI

During data center operation, a large number of devices need to be added, removed, or changed to match service changes, causing great challenges to the management of rack space resources. In a data center, some racks are not fully used or even idle, and some racks are overloaded. How to manage data center resources more efficiently and maximize their utilization should be a big consideration in the next-generation data centers.

AI-based resource optimization technologies can be used to manage data center resources throughout the life cycle, establish a management model and analysis platform centered on device management, automatically count data center resource status through AI simulation and service prediction, and accurately locate data center assets based on asset management. The technologies will comprehensively analyze the available space, power, cooling, and network (SPCN), and intelligently recommend an optimal position for devices on the racks. In this way, data center resources can be visualized and managed, preventing capacity stranding. This feature maximizes the utilization of data center resources, improves resource usage, and boosts operation revenue.

Energy scheduling @AI

With the increasingly diversified energy input and usage of future data centers, an AI-powered autonomous driving platform will enable flexible scheduling of various energy resources and implement on-demand scheduling in terms of green power, PV deployment, and energy storage, reducing manual calculating and operations and maximizing resource value.

2.4 Reliable

Reliability of data center facilities is still jeopardized by many factors.

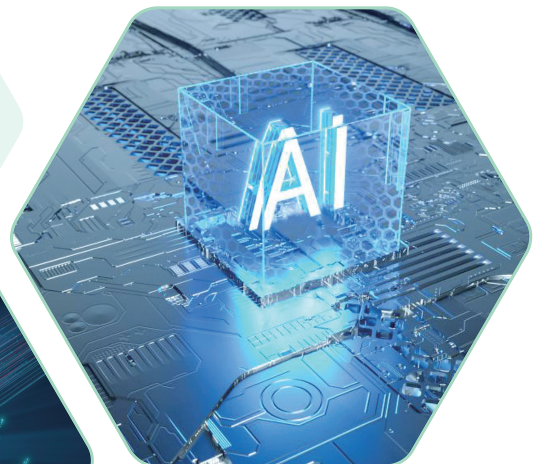
According to the *Uptime Institute global supply-side survey 2021*, failure in the power supply system and cooling system are still the main causes of data center operation interruption, accounting for 57% of them, of which 43% are from the power supply system and 14% are from the cooling system.

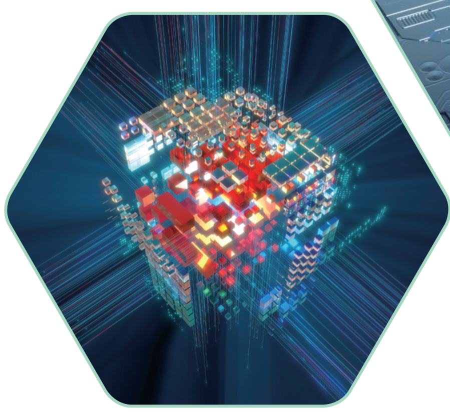
Facility reliability problems would cause huge consequences and impact.

For example, in 2021, a fire broke out in the data center of a cloud service provider in Europe. As a result, 3.6 million websites were down and some data was permanently lost, causing huge losses.

Next-generation data center facilities will be sustainable, simplified, and supports autonomous driving. However, these features are feasible only when data centers run reliably.

Proactive security and architecture security should be built to ensure the reliability of next-generation data centers.





Proactive Security: Early Warning and Quick Fault Closure

According to the Heinrich's Law, any accident can be prevented; for every serious accident, there are 29 minor accidents, 300 near-misses, and 1000 potential accidents.

Based on real-time fault detection, visualized panoramic display, and AI-assisted locating, faults can be quickly detected, analyzed, and troubleshot, and services can be quickly recovered.

Fault prevention is part of high-level reliability.

Traditional data center maintenance usually relies on manual passive response, which makes it difficult to detect minor risks and prevent faults in advance.

Proactive security means using big data and AI technologies to implement predictive maintenance from components to data centers, based on the visibility and perception of all domains in the data centers.

For example, for key components and vulnerable components, such as capacitors and fans, big data collection and AI model training can be used to predict their service life, calculate the failure time of the components in advance, and remind O&M personnel to replace the components. This feature avoids failure before it happens.

On the other hand, the fault response mechanism should change from work order-driven manual response to automatic fault response.

Secure Architecture: Safeguards All Ranging from Components to Data Centers

In the future, ICT technologies will be more widely used in the data center facility field. They will be deeply integrated with power electronics technologies to ensure reliability at various levels, such as components, devices, and systems, to effectively enhance facility resilience and build a comprehensive security defense line. For example, at the component level, key components that adopt modular design are hot-swappable to quickly recover from faults. At the device layer, the full redundancy design enables seamless switching to redundant modules within 0 ms delay after a single point of failure (SPOF), ensuring that devices run without any interruption. At the system level, an end to end (E2E) visualizable, manageable, and controllable platform enables the system availability to reach "Five Nines" (99.999%).

To sum up, the architecture-level optimization design reduces energy conversion layers and removes extra components. In addition to reducing fault points, hitless and interruption-free switching ensure that systems are always online and secure.

Summary

In the pursuit of next generation data center facilities, technological innovation will be a key force in ensuring sustainable development. Looking towards the future, Huawei will keep making breakthroughs in products and technologies through continuous investment in R&D and with the all-round cooperation with customers, ecosystem partners, and industry organizations. We can jointly usher in a new era of data center development.

HUAWEI TECHNOLOGIES CO., LTD.

Huawei Industrial Base

Bantian Longgang

Shenzhen 518129, P. R. China

Tel: +86-755-28780808

www.huawei.com