HUAWEI

Striding Towards
The Intelligent World
White Paper

Data Storage

Building a Data-Centric, Trustworthy Storage Foundation for Diverse Applications

FOREWORD

The evolution of human civilization is a history of how information is stored and transferred. From using knotted strings and hieroglyphics carved on rocks to writing on oracle bones, and from the invention of bamboo slips and paper, to the technology-driven punch cards, hard disks, and flash memory, humans have been pursuing new ways to record key information. Fast forward to the information technology (IT) revolution, and we are witnessing the shift to digital data storage as a new way to store our cultural inheritance assets. Put simply, where there is data, there is data storage.

Since its inception in the 1990s, highly reliable and performant SAN storage has become the first choice for enterprises looking to secure their core databases. Back in 2000, following the rise of the Internet, storage was prioritized in a wide range of industries, leading to the emergence of NAS storage systems to ensure reliable storage and efficient sharing of file data. In 2010, cloud computing became a trending topic, and this drove development like virtual machines (VMs). Meanwhile, resource pooling has become the mainstream architecture in data centers. Unified SAN and NAS storage systems were the new favorite of the era, allowing a single set of storage to support various applications. And this push was fully realized in 2015, when all-flash storage began to replace mechanical hard disks on a large scale. Thanks to advantages in high performance, reliability, and energy efficiency, all-flash created a 10-fold performance increase in diverse workloads. As we moved from the Internet into the mobile age, emerging applications like video, big data, and AI apps, which rely on and produce mass unstructured data, helped develop better distributed storage with stronger scalability.

Digitalization happening globally. In a digital transformation journey, data applications will become more diversified, and increasing volumes of application data will be processed by the production systems, meaning the need for reliable, performant, cost-effective data storage is more important than ever.

For over ten years, Huawei has invested heavily in data storage, to produce a premium portfolio of cutting-edge offerings for 18,000 customers in over 150 countries and regions worldwide, for sectors like carriers, finance, government, energy, healthcare, manufacturing, and transportation. Through cooperation and communication with industry experts, customers, and partners, we write the white paper Striding Towards the Intelligent World – Storage. This report explores the current and future development trends of data storage and offers invaluable insights. I believe this very meaningful research will bring together more industry forces to drive the data storage industry and markets.

Over the past three decades, data storage has become the foundation of high-value data and evolved with the growth of data applications. As we usher in the Yottabyte (YB) data era, Huawei data storage is committed to building a data-centric, trustworthy storage foundation for diverse applications to unleash the power of data-driven productivity.

Dr. Peter Zhou
President, Huawei IT Product Line

CONTENTS

Executive Summary

As new technologies, such as 5G, IoT, cloud computing, and big data, are being applied in digital transformation, enterprise IT architecture is moving towards a hybrid framework of "traditional IT + private cloud + public cloud + edge".

Data storage is the foundation of IT stacks. Storage read and write performance determines that of upper-layer applications and user experience, and enterprises require a reliable data storage platform for their long-term development.

The intelligent world brings with it infinite possibilities. This report provides the following development outlook on the data storage industry:

---

1. By 2025, the global data volume will reach 180 ZB, of which over 80% is unstructured data. Unstructured data is widely used in enterprises and is becoming production and decision-making data.

2. New data applications are continuously created, from traditional database applications to distributed database, big data, and AI applications.

3. Flash storage accounts for more than 50% of the primary storage market, and the era of all-scenario flash storage is coming.

4. Ransomware attacks are becoming the most significant threat for enterprises. Data storage must protect against data security threats, acting as the last line of defense against ransomware attacks.

5. Natural disasters and illegal intrusions threaten data assets and can result in huge economic loss. Data protection is urgently needed to improve the digital resilience of enterprises.

6. AI will optimize storage in two ways: replacing manual management with automatic management and making products smarter, from traditional devices to full data lifecycle.

7. 89% of enterprises use a multi-cloud strategy. Having a combination of public and private multi-clouds is the best choice for cloud transformation.

8. A consumption-based pricing model for cloud computing is driving the emergence of new storage business models. Flexible storage business models are effective measures to cope with explosive data growth and economic uncertainty.

9. Storage devices consume an average of 300 kilowatt-hours per year per 1 TB capacity in data centers. As the data volume increases, energy-efficient storage systems are crucial to ensure zero carbon emissions in data centers.

Our recommendations for enterprise infrastructure and operations leaders (I&O leaders) are as follows:

1. Enterprise IT teams should transform from structured data-centric teams to teams capable of design, plan, and management for mass unstructured data.

2. Use a professional distributed storage system with multi-protocol interworking, hybrid workloads and high-density hardware to improve the efficiency of using mass unstructured data for production.

3. Deploy new and diverse data applications and use the decoupled storage-computing architecture to improve reliability, reduce costs, and increase efficiency; Use the application acceleration engine to implement near-data processing for faster data processing.

4. Take advantage of storage lifecycle changes and new system deployment to accelerate the adoption of all-flash storage.

5. Establish a more comprehensive data protection system within your storage team and build an all-round ransomware protection storage solution to strengthen the last line of defense for data security.

6. Strengthen data protection to ensure that data is not leaked, tampered with, or lost, services are always online, and access is always compliant; and enhance enterprises' digital resilience.

7. Proactively formulate evaluation factors for storage AI management software, strengthen AI capabilities, and develop and train team members.

8. Move emerging innovative services and peripheral services like office automation (OA) to public clouds and retain core competitive services in on-premises data centers.

9. Adopt a centralized data sharing and storage IT architecture, deploy applications across multiple clouds, and plan for a unified cross-cloud data management platform to maximize data sharing.

10. Plan how to obtain IT resources and select the most reasonable business model according to business requirements and future strategies.

11. Deploying storage products with high-density design, converged systems, and efficient data reduction to further reduce data center energy consumption, in addition to reducing power usage effectiveness (PUE).

Outlook I

Unstructured Data

Unstructured data accounts for more than 80% of new enterprise data and is increasingly important to production and decision-making

Unstructured data does not exist in a recognized data structure, such as a relational database table. It includes text, images, documents, and audio/video information. According to Huawei's GIV report, the global data volume will reach 180 ZB by 2025, of which over 80% is unstructured data.

Trend Analysis

Big data, AI and high-performance data analytics (HPDA) give rise to mass unstructured data

With the development of new technologies and applications such as 5G, cloud computing, big data, AI, and HPDA, enterprise unstructured data, like video, images, and files, is rapidly growing. The volume of unstructured data is increasing from PB to EB level.

For example, a world-leading carrier processes up to 15 PB data on average every day. In terms of HPDA, a single DNA sequencer, remote sensing satellite, and autonomous-driving training car generate 8.5 PB, 18 PB, and 180 PB data every year, respectively.

Unstructured data is widely used in enterprises and becomes more important to production and decision-making

With the acceleration of digital transformation, unstructured data has been widely used in enterprises. AI is a typical example. 56% of enterprises use AI for at least one business function, and various scenarios are analyzed and processed based on unstructured data. Improved enterprise data governance capabilities unlock data-driven business growth, and unstructured data begins to move to production and decision-making systems.

In the healthcare industry, historical image archive files are more frequently accessed. AI-powered image reading helps shorten diagnosis from 15 minutes to 20 seconds and increase the diagnosis accuracy from 40% to 95%, greatly improving doctors' work efficiency. In the financial industry, to facilitate online real-time credit extensions, banks use a big data platform to perform real-time big data analytics, interactive analysis, offline processing, and real-time query. This helps banks promptly identify new opportunities and risks, and shorten credit investigation from about one week to real time, creating an ultimate user experience.

Efficient and reliable storage of mass unstructured data underpins enterprise data governance

Mass unstructured data is generated in public clouds, edge sites, IoT terminals, and most frequently in enterprise data centers, where it is predicted that unstructured data will increase at a CAGR of 18% and exceed the amount of unstructured data in public clouds by 2025 to account for 51% of all unstructured data. More and more enterprises are choosing to deploy unstructured data storage in enterprise data centers.

To efficiently and securely store unstructured data in enterprise data centers, a growing number of industries are looking for professional-grade distributed storage solutions. The financial industry uses distributed storage to store image, audio, and video data. In the education industry, distributed storage is an effective way to support HPDA. Most common is the manufacturing industry, where breakthroughs in autonomous driving, industrial Internet, and industrial simulation drive explosive data growth and a greater need for distributed storage.

However, it is clear the storage capacity previously purchased by enterprises is now insufficient. They need to first alleviate their worries over storing vast amounts of data. The traditional multi-copy technique is a capacity barrier to unstructured data storage. To optimize storage space utilization, data reduction techniques implemented by professional distributed storage are needed, such as high-ratio elastic erasure coding (EC), deduplication, and compression. Replacing general-purpose servers with high-density storage hardware also helps reduce the footprint, power consumption, and O&M complexity, achieving optimal TCO. In addition, the industry uses professional distributed storage that integrates software and hardware to provide enterprise customers with end-to-end solutions featuring high reliability, performance, and scalability. This simplifies deployment, management, and services and reduces OPEX.

To deal with data mobility issues, professional distributed storage implements hot, warm, and cold data tiering and automatically relocates data to different tiers for optimal ROI. Unstructured data management becomes increasingly complex. It is difficult to manually allocate data to proper storage space in a timely and flexible manner, resulting in inefficient data management and costly O&M. The hot, warm, and cold data tiering technology of professional distributed storage can store data in proper storage space based on policies with automatic data migration. This solution solves the problems encountered by enterprise customers and is widely used across industries.

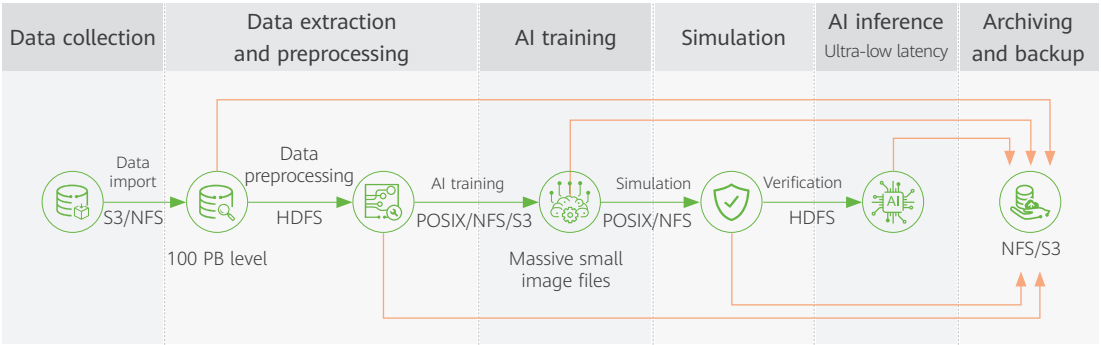

Figure 1: Data processing flow of autonomous driving training

Unstructured data-powered technologies often involve multiple access protocols (file, object, and HDFS) in one data processing flow. To ensure premium usability, preferred solutions adopt multi-protocol interworking without data copying to reduce data redundancy. Figure 1 shows the data processing flow of autonomous driving training.(Figure 1: Data processing flow of autonomous driving training)

Suggestions

Enterprise IT teams strengthen their capabilities of processing mass unstructured data

As enterprises use unstructured data more widely, especially in their production and decision-making systems, the ability to efficiently store mass unstructured data and extract its huge value to help make informed decisions becomes a key competitive edge. Therefore, it is necessary for enterprise IT teams to strengthen their capability of processing mass unstructured data, and transform their structured data-centric capabilities to design, planning, and management of mass unstructured data.

Choose professional distributed storage to build a foundation for mass unstructured data

To improve the efficiency of using mass unstructured data for production, use a professional distributed storage system to build a global unified data storage foundation centered on unstructured data. It is best to choose a distributed storage system that supports hybrid workloads, multi-protocol interworking (file, object, and HDFS), data reduction, and high-density hardware to ensure sufficient capacity, superb data mobility, and premium usability.

Evaluate multiple factors and then determine whether to use enterprise data centers or public clouds to deploy unstructured data

When designing and planning the deployment of mass unstructured data, consider the full data lifecycle management (data generation, storage, access, and migration) as well as data sharing and mobility between service platforms and even across cloud platforms. In addition, it is recommended that enterprises evaluate factors like TCO, performance, and security before selecting enterprise data centers or public clouds.

Outlook II

Diverse Data Applications

Diverse data applications such as distributed databases, big data analytics, high-performance computing (HPC), and AI are booming

Over three decades, data storage has evolved to become the optimal foundation of high-value data in line with the data application development. As industries are going digital, new data applications are continuously generated. Besides traditional database applications, distributed database, big data, and AI applications are emerging. According to statistics, each enterprise has over 100 types of data applications

Trend Analysis

Immature IT stacks for new data applications, fueling urgent need for tiering standards

When new data applications emerged in the past, the integrated server architecture with coupled applications and local disks was often used to rapidly deploy innovative services. However, as innovative services become production services, the gap between the computing power lifecycle and data lifecycle increases greatly. Computing and storage resources need to be planned and maintained flexibly and independently. Enterprises begin to focus on reliability and cost-effectiveness. The architecture based on local disks of servers is far less reliable than external storage. In addition, compute and storage cannot be independently expanded, resulting in high hardware costs, idle resources, and low utilization.

Consider the distributed database as an example. At the early stage of its emergence, enterprises adopted the coupled storage-compute IT stack architecture that causes resource waste. Both compute and storage resources need capacity expansion, when either are insufficient. As technologies become mature, distributed databases, such as Amazon Aurora, run on a decoupled storage-compute architecture that makes IT resources more flexible to expand and saves more than one third of server resources. Meanwhile, enterprise-grade all-flash storage notably improves data reliability.

Big data analytics is essential to informing operational business. Since 2010, enterprises have begun to turn to a new data analytics technology, Hadoop. Its local data storage design creates a coupled storage-compute architecture. As data processing scales increase from TBs to

PBs or even tens of PBs, the disadvantages of the coupled storage-compute architecture have gradually been exposed, specifically those of low resource utilization and high deployment costs.

Some enterprises are exploring IT architecture innovations. For example, a carrier adopted the new decoupled storage-compute architecture to replace the original coupled one. The decoupled model slashes hardware and software costs of cabinets, servers, and electricity by 40%, and the annual carbon emissions by over 50% on average.

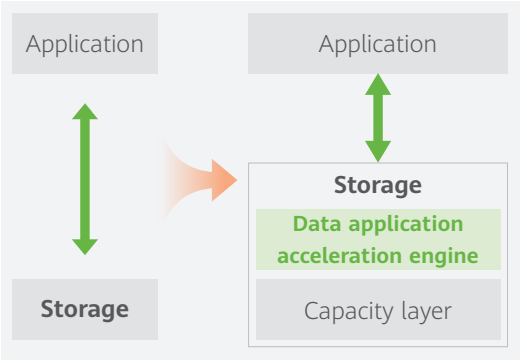

Figure 2: Application acceleration for near-data processing

The digitalization level of enterprises is determined by the efficiency of diverse data applications, which therefore should be improved through joint innovation

As emerging applications such as big data and AI are widely used in enterprises, the efficiency of data applications determines the digitalization level of enterprises. With the development and segmentation of new application scenarios, joint innovation of applications, compute, and storage is required to build efficient solutions for growing sub-scenarios. For example, storage vendors are building acceleration engines as joint innovation carriers for diverse data applications.(Figure 2: Application acceleration for near-data processing)

Distributed database application acceleration engine: Currently, most mainstream distributed databases can only enable write-once and read-many, requiring database/table sharding. The database application acceleration engine can enable write-many and read-many for distributed databases, minimizing application reconstruction.

Big data application acceleration engine: The data access latency of the traditional Hadoop big data platform is hundreds of microseconds, and the data analysis latency reaches days. Now, a distributed storage high-speed cache is built to move application operators to the storage layer, reducing the data access latency to 10 microseconds and accelerating the big data analysis efficiency to minutes.

HPDA application acceleration engine: High-performance data analytics involves mixed data types of large and small files, posing high requirements on bandwidth and OPS. Traditional storage devices can only support the access performance of a single data type. To resolve this, parallel data access clients that support cache acceleration are deployed at the compute layer, and metadata access is accelerated at the storage layer, to meet high performance requirements for mixed access of large and small files.

AI application acceleration engine: AI training evolves to large and multi-modal models, increasing the number of modeling parameters by 10 to 100 times. The AI acceleration engine is deployed to accelerate feature processing and intelligent scheduling of pipeline tasks, improving the AI training efficiency by dozens of times, accelerating training period, and controlling time costs.

Suggestions

Use the decoupled storage-compute architecture for emerging data applications to improve reliability and cost-effectiveness

With the deepening of digitalization, diverse data applications are becoming new production applications, posing increasing requests on data reliability. In addition, as the gap between the computing power lifecycle and data lifecycle increases greatly, compute and storage resources need to be planned and maintained flexibly and independently.

It is recommended that elastic, reliable, and cost-effective professional storage running on a decoupled storage-compute architecture should be used for diverse data applications, to rapidly deploy innovative services, and improve reliability and cost-effectiveness. (Figure 3: Decoupled storage-compute architecture for emerging data applications)

Deploy IT stacks with acceleration engines to better support diverse data applications

To facilitate diverse data applications, future storage will no longer be just a persistence carrier of data, but data infrastructure that combines the persistence layer and data acceleration engines. The acceleration engines connect the storage layer to different applications for near-data processing, improving processing efficiency by several times, and significantly accelerating applications.

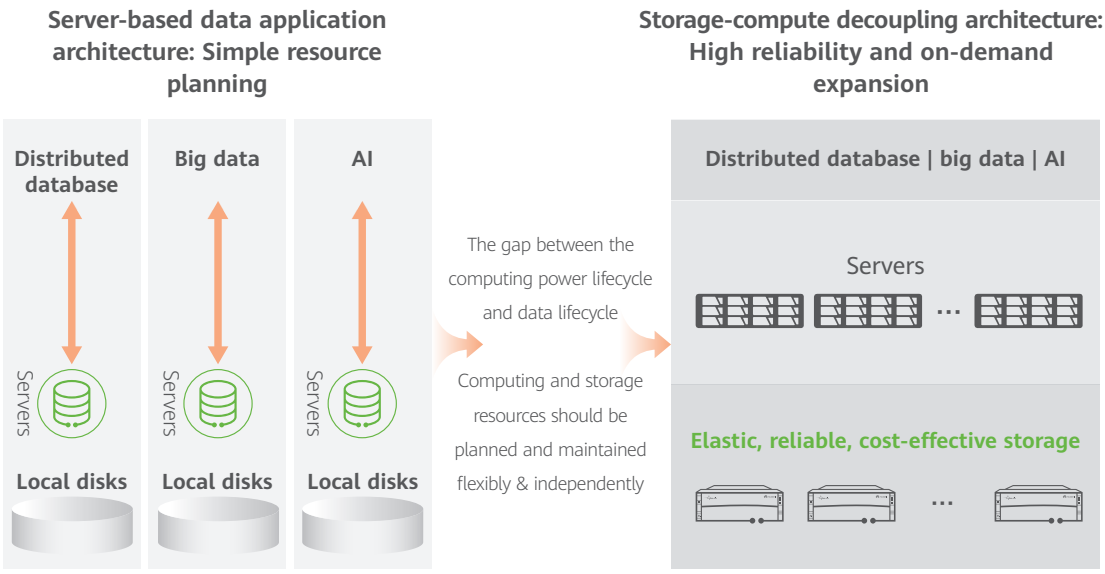

Figure 3: Decoupled storage-compute architecture for emerging data applications

Outlook III

All-Scenario Inclusive All-Flash

Worldwide inclusive flash storage drives larger all-flash storage market

In 2014, Gartner released the very first Magic Quadrant for all-flash arrays, reporting that all-flash storage accounts for over 50% of the primary storage market (2021Q4). Now the evolving all-flash storage technologies are driving larger all-flash storage market.

Trend Analysis

Upgraded SSD tech promotes all-flash storage to diverse industries

The evolution of SSD technologies has enabled all-flash storage to meet a wider range of service scenarios in various industries.

NVMe SSDs adopt the NVMe standard to offer twice the performance of traditional SAS SSDs. Equally, NVMe over Fabrics, also known as NVMe-oF, is an enhanced protocol employed by SSDs to provide sub-ms latency for storage networks. Enabling improved performance with every generation, all-flash storage is perfect for real-time ultra-low latency scenarios, such as transaction systems. (Figure 4: Architecture and performance comparison between traditional and NVMe storage)

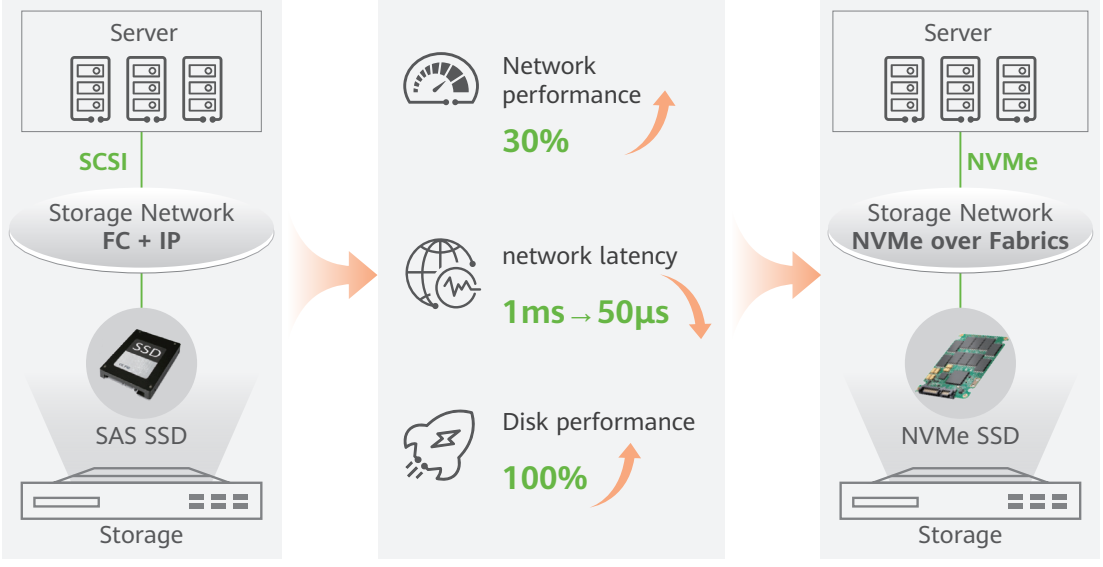

Figure 4: Architecture and performance comparison between traditional and NVMe storage

The global wear leveling technology uses algorithms to evenly write data to cells of all SSDs in the storage system, greatly extending the SSD service life and enabling enterprises to use all-flash storage in core service scenarios.

Similarly, data reduction technology is used to reduce data volumes without compromising the data, thereby reducing the number of disks and devices required in a storage environment. Today's data reduction ratios are as high as 4:1 or 5:1, helping slash the purchase costs of all-flash storage and allowing all-flash storage to enter more non-core service deployments.

These technologies are a catalyst for all-flash storage to be perfectly deployed in enterprises' core transaction and production systems, and can also be used in decision-making, operational support, and backup systems.

Inexpensive all-flash: Quad-level cells (QLCs) and extra cell layers in 3D NAND

NAND cells are the core component of enterprise-level SSDs, and determine the cost of SSDs. Most mainstream vendors now adopt NAND cells with 176 layers, and even released the 200-layer (nearly double that of 2018) design roadmap. For example, the Micron roadmap[2] states the cell layers in a 3D NAND device will exceed 200, 300, and even 400, which in turn will slash the price of every TB of SSDs.

In addition to prioritizing stacking layers, triple-level cells (TLCs) are becoming a mainstream choice for enterprise-level SSDs, which have given rise to QLC SSDs. One such market offering is the PLC technology released by Toshiba at the Flash Memory Summit 2019. With this tech, every cell of PLC stores five electrons, making it a cheaper option than TLC/QLC but with a shorter cell life. This trend of optimized enterprise-level PLC SSDs will help further reduce the purchase costs of flash storage.

A combination of many stacking layers with TLC/QLC/PLC can significantly reduce the price of a single SSD. IDC predicts that by 2025, the cost per unit of SSD capacity will be lower than 10K RPM HDDs and higher than large-capacity HDDs used for cold data storage.(Figure 5: SSD price trend per unit capacity predicted by IDC)

Breakthroughs in cell technology and PCIe 5.0 are improving the capacity of SSDs, with many enterprises now adopting 15.36 TB SSDs or even 31 TB SSDs. Thanks to lower procurement costs, large-capacity SSDs will help develop storage environments in data centers.

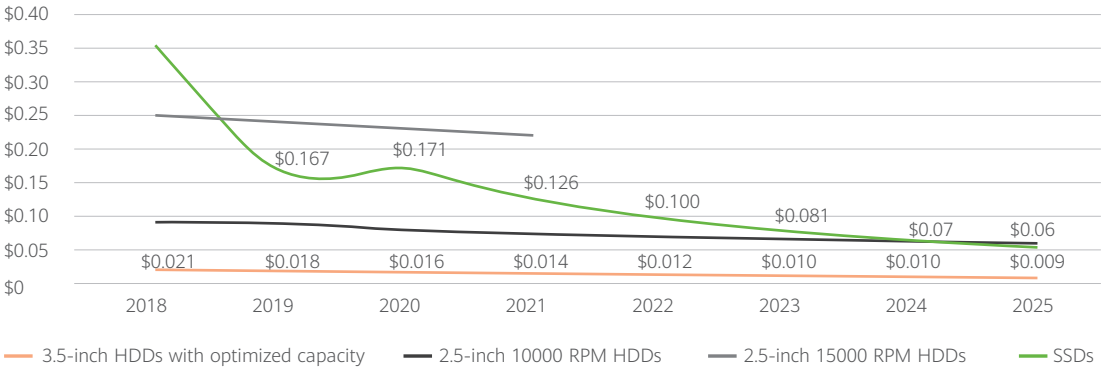

Figure 5: SSD price trend per unit capacity predicted by IDC

Worldwide drive for all-flash storage

The ratio of all-flash storage in developed countries is on average over 45%, included 56.3% in the US, 54.7% in Australia, 54.3% in Netherlands, and 50% in Sweden.

But now, countries that traditionally have low all-flash usage ratio are adopting more and more all-flash storage. In 2021, the annual growth rate of the all-flash storage market in Indonesia reached 54%, while that in Poland, China, and Mexico reached 34%, 24%, and 14%, respectively. This trend is expected to continue in the future.

Suggestions

Plan all-flash storage tailored to enterprise current and future data volumes and requirements

Digital transformation is the cause of huge data growth and service pressure. When making their storage construction plans, enterprises should evaluate the current and future IT system requirements in advance, to ensure premium experience and as these can cause them to fall out in the new round of all-flash storage competition among vendors.

All-flash storage is an excellent choice for enterprises' technical teams to improve performance. By working with storage providers to evaluate the future data volumes and service pressure trends, enterprises can formulate all-flash storage strategies, and analyze benefits and O&M cost changes after the strategies are implemented.

Seize opportunities to replace legacy storage with all-flash models

HDDs are common in many enterprise storage environments, but now most of them are coming to the end of the warranty period. Enterprises undergoing digital transformation are in urgent need of better, more performant storage devices. This is an excellent opportunity for storage enterprises to promote all-flash storage. The zero-interruption service migration is a key selling point of mainstream vendors, and a must-have for enterprises.

All-flash storage: Excellent performance, rock-solid reliability, and large capacity for much lower CAPEX and OPEX

Data is the core of digital transformation. Data center consolidation helps gather data scattered on many devices to minimal devices, to reduce purchase, maintenance, management, and power consumption costs, and facilitate data value mining and enable service growth.

All-flash storage systems that run on large-capacity SSDs and dedupe and compression technologies enable equivalent to 5 to 10 HDD storage systems. This huge benefit makes it an obvious choice to migrate services from several HDD storage to a single all-flash storage, to supercharge the efficiency of data centers.

Outlook IV

Ransomware Protection

Data storage is the last line of defense against ransomware

When ransomware strikes, it steals and encrypts valuable data that can only be decrypted by paying a ransom. According to the Global Cybersecurity Outlook 2022 released by World Economic Forum, ransomware is the biggest concern for cyber business leaders.

Trend Analysis

Ransomware targets large enterprises and infrastructure, with enterprise data at sharply-increasing risks

In 2021, the US oil pipeline company Colonial Pipeline was the target of a ransomware attack, causing the company to halt all operations and pay a ransom of US$4.4 million. In the same year, the American insurance corporation CNA Financial experienced encryption of 15,000 devices, costing the company US$40 million to re-access the data. In April 2022, ransomware was the cause of 1.4 TB data leakage from multiple Toyota suppliers, forcing Toyota to cut its production capacity that year by 500,000 vehicles.

These are just some of the many examples of how hackers are targeting large enterprises and infrastructure. In fact, ransomware attacks are now low-cost and simple. Specifically, ransomware attacks can now be customized and commercialized, made into available products for malicious users through memberships and subscriptions. This has pushed the threat of ransomware everywhere. A typical ransomware attack steals and deletes all data copies stored in your local and even disaster recovery (DR) centers, which not only means data is hard to recover but also causes leakage of private and confidential data. Then the enterprises are blackmailed with not just ransom, but risk damage to brand image, loss of business opportunities, legal proceedings, and labor costs. This collateral damage can be huge to a company - in some cases, these collateral losses are as high as 23 times that of the initial ransom. A Cybereason report shows 49% of enterprises that have paid the ransom from an attack only retrieve part or none of the lost data, while 80% of enterprises that paid the ransom are targeted by ransomware a second time.

Storage: Part of the ransomware protection process and the last line of defense for data security

Traditional causes of data security risks include natural disasters and system hardware faults, such as fire, flood, and disk damage. These threats can be easily handled with DR solution and disk wear detection technology. However,

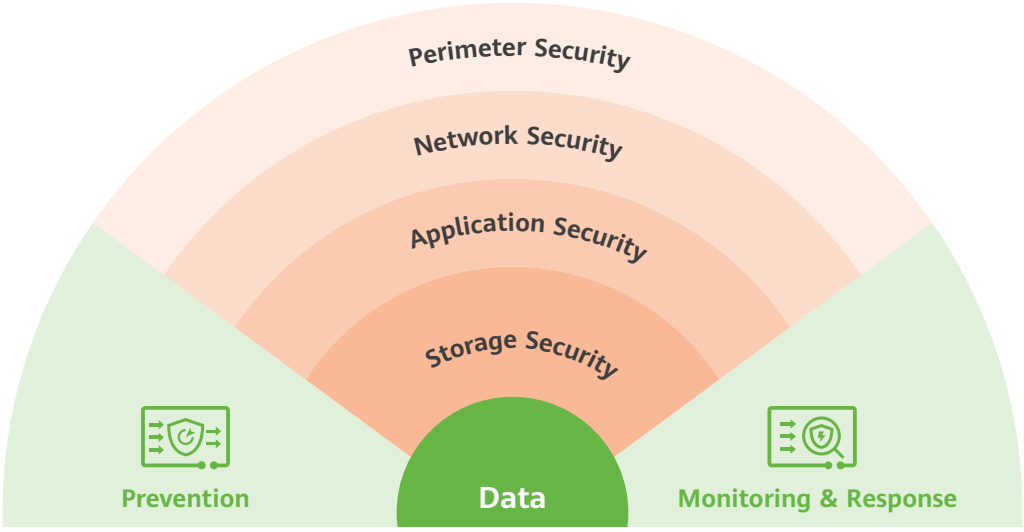

Figure 6: Comprehensive data protection and defense framework

currently, the number of human-caused damages represented by ransomware attacks keeps increasing and causes enormous economic losses. This necessitates the construction of comprehensive data security protection covering both network and storage.(Figure 6: Comprehensive data security protection)

Ransomware exploits zero-day vulnerabilities (system or device vulnerability that has been disclosed but is not yet patched), phishing emails, and physical attacks to imbed your system with ransomware. The network functions, while designed to prevent, block, scan, and eliminate ransomware, are rendered useless if your system is infected with a virus. In our modern digital age, data storage needs to do more than just storing data – it needs to protect data. Specifically, storage uses technologies such as pattern recognition and machine learning to identify ransomware, and uses data security features such as ransomware detection, secure snapshot, data isolation, and data recovery to provide logical and physical protection for data. As the final stop of data, it is critical to build storage security protection capabilities. (Figure 7: Data storage ransomware protection overview)

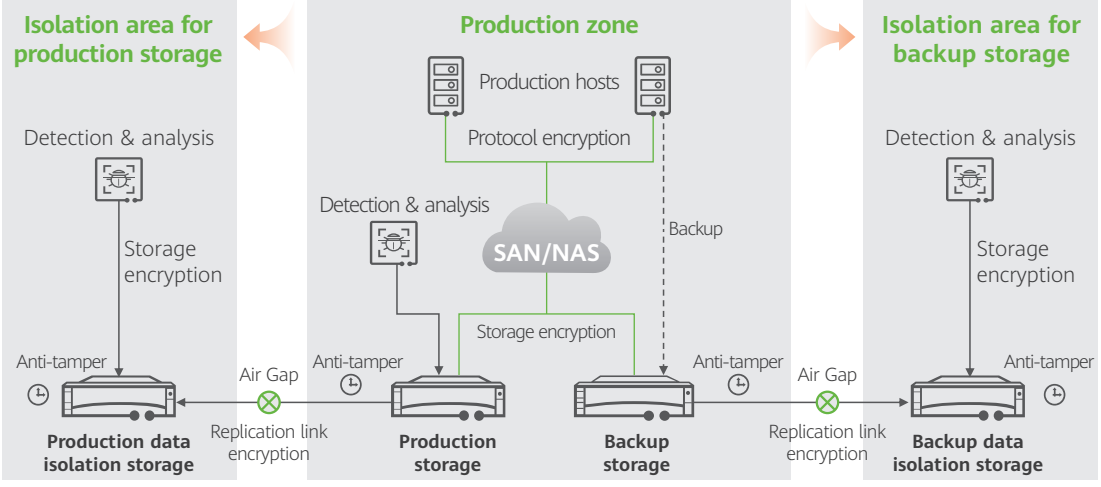

Figure 7: Data storage ransomware protection overview

Ransomware protection with primary storage: After data enters the production storage, a safe zone is created inside the storage to prevent data from being tampered with or deleted through the secure snapshot and Write Once Read Many (WORM) features of the storage. An independent physically-isolated zone is also created, combining with air gap technology to automatically disconnect replication links and replicate data to the isolation zone for enhanced protection.

Ransomware protection with backup storage: Similar to primary storage, the encryption, secure snapshot, and WORM features of the backup storage ensure the data in the storage system is clean. An isolation zone is also established to ensure data security, allowing operators to quickly restore secure data and services in the event of an attack.

It is critical that storage accurately detects ransomware. The leading detection strategy in the industry is: A baseline model is established based on historical data to check for abnormalities in the changed feature values of the metadata of copies; abnormal copies are further compared to determine the file size change, entropy value, and similarity; the machine learning model is used to determine whether file changes are caused by ransomware encryption and mark them accordingly.

Suggestions

Combine resources of the storage and data security teams, for a comprehensive data security protection system

Typically, enterprises data security teams comprise network experts who are responsible for imposing strict security policies on network security devices (such as firewalls), to protect high-risk ports and reduce exposure to threats. But, even mainstream solutions alone are insufficient for handling ransomware attacks. One option is to include storage experts to the data security team to establish a comprehensive data security protection system. It cannot be understated the importance of storage experts and storage protection layer in a system. Data protection measures such as secure snapshots and data isolation technologies prevent data from being tampered with, while detection and analysis technology accurately and quickly identifying ransomware, helping recover data from ransomware attacks as soon as possible.

Protect the last line of defense with a comprehensive ransomware protection storage solution

To build a comprehensive ransomware protection storage solution, enterprises need to: prioritize protection for key data assets based on their application service level agreement (SLA) requirements; implement anti-tampering and offline protection for data copies on production and backup storage, retaining a clean data copy for restoration; select high-performance all-flash as the production and backup storage, as it can quickly recover services and reduce service downtime loss in the event of ransomware attacks.

Outlook V

Digital Resilience

As data is becoming the core asset for enterprises, digital resilience is a major metric for any enterprise resilience framework

In the digital age, data is a core asset for enterprises, who must build a highly resilient system around it. That is, enterprises must prioritize digital resilience, to build sustainable long-term supply chains, legal compliance, financial transparency and health, and operational continuity.

Trend Analysis

Digital resilience is key to building leadership

Digital transformation empowers enterprises to carry information in their data. The acquisition, storage, transmission, exchange, and processing of digital information depend on high-reliability and -stability infrastructure. Data infrastructure is the foundation of digital information, which in turn is the upper-layer application of data infrastructure. The integrity, confidentiality, and availability of digital infrastructure play a key role in economic and social informatization and digitalization. If critical IT infrastructure is damaged or data is breached, related economic and social entities will be impacted and digital economy systems will become unstable. IDC predicts that 65% of global GDP will be digitized by 2022. This makes digital resilience more important than ever, with over 80% of enterprises at medium and high risk.

Natural disasters and manmade mistakes make digital resilience essential

Now, data has become a new means of production, and the lifeblood of organizations. However, the IT system tends to be vulnerable under unprotected conditions. System faults, natural disasters, and human accidents can cause system breakdowns, the losses from which are often beyond imagination.

In March 2021, a fire destroyed four major data centers operated by OVH, Europe's largest cloud infrastructure operator, paralyzing a large number of customer websites, and rendering some customer data lost and unrecoverable. In October 2021, the Tokyo Stock Exchange, the world's third largest stock exchange, was suspended for one day due to storage device failures. The lack of emergency protection and DR solutions even had an impact on global stock exchanges.

Backup and DR measures are essential in any IT system. If a core IT system fails, service systems of governments and enterprises will be greatly affected. Statistics indicate that 60% of businesses who suffer losses due to the lack of service redundancy and data backup plans will

become insolvent within two to three years, with 50% of the daily turnover lost in the first two days. Needless to say, there is urgent need to improve digital resilience through backup and DR technologies.

Data protection technologies for new services are booming

Breakthroughs in emerging tech, such as 5G, mobile Internet, and IoT have been the catalyst for industry developments, such as those in online financial transactions, smart manufacturing, e-governments, and digital culture and tourism. Full-lifecycle data protection faces new challenges in terms of the protection level, object, performance, and scale.

Protection level: Varying levels of protection depending on the assets scope. For example, a bank with assets worth under CNY 100 billion can adopt an intra-city DR solution for its IT systems, whereas those with assets exceeding CNY100 billion are better served with the geo-redundant 3DC DR solution.

Protection object: Both structured and unstructured data need protection. For example, big data is helping in production decision-making for enterprises, and mass unstructured data on which it depends also needs DR protection.

Protection performance: Cold data in the DR center is converted to warm data to maximize value. For example, the number of CDRs generated by a million 5G users increases by 7-fold, requiring higher protection performance.

Protection scale: The amount of data to be protected has increased from PBs to EBs. For example, an L4 autonomous car can generate 60 TB of data per day, 50 times that of an L2 car.

In this background, there is huge emphasis on zero service interruption or data loss. For service continuity, data protection solution vendors offer integrated SAN/NAS active-active solutions to protect both structured and unstructured data; for data retention in new application ecosystems, retrieval and backup of mass unstructured data are used to protect mass small files.

Suggestions

Achieve the goal of "Three Zeros, Two Always" for secure, trustworthy data

Different workloads face different data security challenges and threats. To improve digital resilience, the goal of "Three Zeros, Two Always" for data security and trustworthiness must be achieved:

Zero data loss: Customers' data must not be damaged or lost due to software and hardware problems. Even in the event of an abnormality, damaged data, or misoperation, the system can be recovered and continue without major disruption.

Zero data leakage: Data must not be accessed or obtained without authorization during storage, transmission, and processing.

Zero data tampering: Data will be protected from unauthorized modification or damage during storage, transmission, and processing, and even in an unlikely infiltration, tampered data can be recovered upon detection.

Always-on services: The fulfillment of customer service specifications ensures storage services not be, but if storage services are interrupted due to customer environment problems or malicious attacks, the services will be recovered within the customer's tolerance period after the environment recovers or malicious attack is resolved.

Always compliant access: Data storage, transmission, and processing must comply with laws and regulations such as General Data Protection Regulation (GDPR).

Enhance DR security for core data to ensure service continuity

Enterprises should enhance protection grades to comprehensively upgrade service continuity. For enterprises that lack sufficient DR and data backup measures, a comprehensive protection framework is essential for full DR of mission-critical services and full backup for data.

To improve service continuity and data reliability, the following preparations should be met: examine enterprise services, check for the lack of DR for key services, and help enterprises build redundant data centers. If DR systems are already set up, upgrade the local active-passive DR solution to an active-active solution to ensure zero loss of key data, and develop intra-city, active-active DR architecture to a multi-site, multi-center protection solution to facilitate multi-copy, cross-region recovery. Enterprise data should be backed up by level and class, while local backup can be upgraded to local + remote data backup for core data. Anonymized and unimportant data can be backed up to online public clouds or low-cost storage media.

Another best practice is to regularly perform DR drills, recovery verification tests, and data application to ensure the health and effectiveness of the DR system.

Improve enterprise resilience and DR capabilities for new production platforms and applications

As enterprises invest more resources in innovative technologies, it is recommended that elastic architectures and agile iteration capabilities be built for emerging production platforms and applications such as distributed databases, big data, AI, and containers. Efficient incremental protection, compliance with laws and regulations, and mass retrieval should be offered for unstructured hot data.

Because of high service complexity and insufficient solution measures in emerging ecosystems, enterprises and storage vendors should work together to develop scenario-specific solutions and data protection technologies that improve the integrity of enterprise data protection.

Outlook VI

AI-Powered Storage

From management to products, AI powers autonomous-driving storage throughout data lifecycle

More and more enterprises introduce AI for IT Operations (AIOps) to handle huge data volumes, improve efficiency, and facilitate automated O&M. Now, AI technologies in the storage field are no longer limited to monitoring and O&M of devices, but are integrated into storage products.

Trend Analysis

Enterprises – Use AI to improve O&M automation of storage systems

The explosive growth of data volumes in data centers (DCs) has created new challenges to storage management, such as fault location and risk identification and as such, existing O&M methods are no longer sufficient. According to Gartner, by 2023, 40% of I&O teams in large enterprises will use AI-augmented automation. Enterprises are expected to invest more in AI tech to automate storage O&M in DCs, improving resource management and O&M efficiency at reduced manpower.

Enterprises and storage vendors – Jointly develop a 3-layer AI architecture (Cloud-Center-Device AI)

To produce high accuracy and reliability, AI training requires a large amount of data for accumulation and model optimization. To meet this, enterprises are using storage vendors' AI management tools to build a 3-layer AI architecture to centrally manage storage devices, simplify infrastructure O&M, and improve efficiency.(Figure 8: 3-layer AI architecture)

Device AI: Software and hardware resources on devices are automated, with recommendation of device configuration items, auto-detection faults and slow disks, acquiring device data for cloud training, and running AI model updates from the cloud via online updates or offline imports.

Center AI: Dedicated software can implement unified management of multiple devices in a DC, as well as resource pooling, standardization, and automation of storage devices. The software is deployed in a private DC and therefore isolated from the extranet for stringent data security controls.

Cloud AI: Powerful cloud-based computing resources are used to train AI models using the training data uploaded from storage devices; optimized AI models are distributed on demand to DC management software and devices; cloud

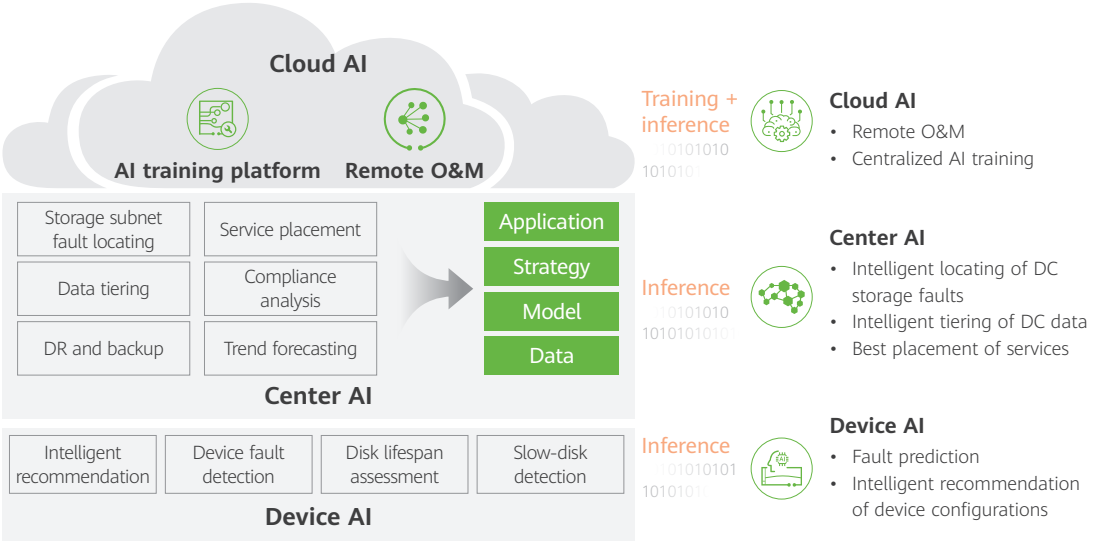

Figure 8: 3-layer AI architecture

management software can implement remote intelligent O&M of storage devices, despite its weaker capabilities than DC management software. For security purposes, remote O&M prohibits device modifications.

Storage vendors – Build intelligent storage products to optimize device efficiency and reliability

To fit diversified storage requirements of different applications, storage vendors are integrating AI into storage products to enhance device performance and reliability. Dell EMC storage systems use built-in intelligent tuning and data reduction algorithms for self-optimized storage provisioning and optimal data reduction ratios; NetApp systems can intelligently optimize hardware resources scheduling to accelerate data access; and Huawei storage intelligently allocates hardware resources to accelerate data reads and writes, while intelligently adjusting data reduction algorithms based on data types, to increase data compression rates and lower storage cost per unit of data.

In traditional storage, algorithms and data are coupled, and multiple fixed algorithms are distributed at the cache and scheduling layers, and

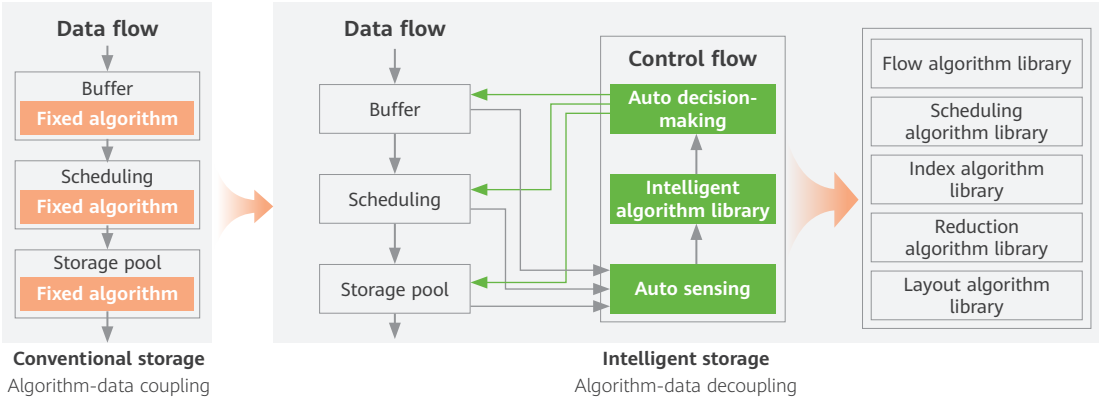

Figure 9: Algorithm-data decoupling with intelligent storage

in the storage pools of storage devices. However, algorithm parameters need to be manually adjusted to ensure the access efficiency of different types of data. By contrast, intelligent storage incorporates architectural innovations by decoupling algorithms from data. A self-learning and adaptive algorithm library is built to make autonomous decisions on layout, scheduling, and reduction of different data types, ensuring efficient and flexible access in diverse data applications.(Figure 9: Algorithm-data decoupling with intelligent storage)

Suggestions

Develop new evaluation elements for storage AI management software

To accelerate enterprise digital transformation, both storage vendors and enterprises must consider how to integrate AI management software into enterprise production and management services. It is recommended that enterprises establish clear evaluation factors and standards for the AI management software provided by storage vendors. This will drive storage vendors to upgrade AI management software to enterprises' most concerned core values. The evaluation elements should cover the following dimensions:

Responsibility scope: AI is not developed to replace humans, but to assist and strengthen human abilities and contributions by learning and transcending how human beings perceive and respond to the world. It is recommended that enterprises develop the responsibility scope of AI within which storage vendors upgrade and expand AI capabilities to guarantee that storage AI management is under the control of enterprises.

Technical specifications: AI algorithms depend on learning and training. Model understanding and training data volumes determine the error rate of AI inference results. It is recommended that enterprises develop service-specific, quantifiable AI technical specifications that can be proven by storage vendors, while storage vendors that do not meet AI specification requirements will not be adopted by enterprises.

Capability extension: The evaluation criteria for AI management software should extend from independent capabilities to E2E closed-loop designs. For example, storage disk fault prediction should focus on the closed-loop capabilities of storage management software, such as identification, prewarning, proactive isolation, replacement, and data rebalancing.

Upgrade enterprise tech stacks for storage AI

Due to the large-scale deployment of AI in storage devices and management software, enterprise infrastructure management teams need to systematically plan data-centric AI capabilities. Enterprise digital transformation starts with retraining internal staff to ensure tech stacks transforms from solely storage management to E2E automation and enterprise AI capability building.

Equally, internal infrastructure teams can use intelligent management software for storage resources to fit service requirements, and ensure service agility through standardized and service-oriented resource management. Similarly, they need to stay up-to-date with AI trends; explore management intelligence in intelligent storage infrastructure; and use AI to mine data value and inform business decisions.

Outlook VII

Storage Business Models

A flexible storage business model to help handle explosive data volumes and economic uncertainties

Following years of increasing economic uncertainties and huge data growth, enterprises in the global economy are looking for more flexible storage investment models to reduce their capital expenditure (CapEx). Against this background, industry influencers have introduced a new business model based on consumption.

Trend Analysis

Available storage capacity: Making storage more accessible

Current levels of investment in storage are too low for enterprises to meet the ever-increasing data volumes. Breakthroughs in technology partially solve this problem because they reduce costs. Considering the most essential requirement of data storage is capacity, most enterprises will prioritize and procure large-capacity storage offering premium levels of security and efficiency. Thanks to data redundancy protection and reduction technologies, enterprises now can buy storage by available capacity, instead of the traditional method of purchasing by raw capacity.

Available capacity, also called effective capacity, refers to the storage capacity in a storage device that can be used for enterprise storage, accurately determining device capacity. This model helps lower the costs of deploying storage devices, and reduce the total cost of operation (TCO). This model is the result of a series of data reduction technologies such as data dedupe and compression, and represents the capacity that can be written into data by users' hosts. The data reduction algorithms increase the available capacity of users, allowing more data to be stored in the same space, and further reduce storage costs. Moreover, this model allows enterprises to pick and mix products to meet their data storage requirements, simplifying storage planning and procurement processes.

This new model is developed differently in different countries. While commonly available in Europe, North America, and Japan, developing countries such as China are still exploring it. But, there is optimism the available capacity mode can further help developing countries move to a digital future.

STaaS: A counter of storage vendors against the consumption-based business models of the public clouds

There has been an obvious shift on how

enterprises are preparing their business. While traditionally, this would be implemented by purchasing assets, features, and functions, now they are emphasizing results. Specifically, prioritizing solutions to transfer low-value, repetitive tasks to suppliers and partners.

Generally, enterprises want flexible, low-cost, and low-risk IT facilities. Ideally, by leveraging the low costs, high-scalability advantages of multiple vendors, to avoid being locked in by a single vendor. But another issue is that of data sovereignty, which requires enterprises to consider data residency and supervision in the destination country when expanding market overseas.

Storage as a Service, or STaaS, allows enterprises to replace their own internal storage infrastructure with hybrid cloud IT options that provide centralized lifecycle management for hardware and support from suppliers. This eliminates difficulties such as outdated infrastructure and architectures, and ensures reliable services through continuous monitoring and optimization of data environments. Currently, vendors such as Dell Apex, HPE GreenLake, Hitachi Vantara, IBM, NetApp Keystone and Pure, all provide STaaS to help customers explore new business opportunities.

STaaS can support enterprises in their digital goals because it uses innovative infrastructure technologies, is quick and easy to set up, and allows enterprises to coordinate technology adoption and IT operations. Stimulated by the pay-per-use consumption strategy in public clouds, hybrid cloud IT have begun offering an internal STaaS operation deployment mode. At the beginning of 2020, no major storage vendor provided STaaS, but by the end of 2021, almost every vendor provided the basic block STaaS to compete with ultra-large cloud vendors. Gartner predicts that roughly 15% of enterprise storage capacity will be deployed as hybrid cloud IT STaaS by 2025, increasing to 50% by 2030.

One main reason impacting the take-up rate of hybrid cloud IT STaaS is that the data size of a single enterprise is quite small compared with that of a single public cloud. So, despite the stable annual investment, the price of STaaS per unit capacity is higher than that of the traditional CapEx mode over the course of the contract.

Suggestions

One benefit of adopting cloud-based, service-oriented IT infrastructure was to change from purchasing devices to subscribing to services. This enables customers to more flexibly use IT resources, to better meet service development requirements. But in doing so, there are these considerations:

- IT resources should be purchased based on your actual service requirements and future strategies. Compare the CapEx and operational expenditure (OpEx) of the entire hardware lifecycle to determine the right product. While STaaS offers stable costs every year, its price per unit of capacity throughout entire contract period is higher than traditional CapEx models.

- Storage devices or services should be built on the latest data reduction technologies and new business models, to slash TCO and tailor IT facilities to business development needs.

Outlook VIII

Multi-Cloud Architecture

As multicloud becomes the new normal, the IT architecture that centrally stores and shares data and deploys applications in multiple clouds will be the optimal choice

Continuous digital transformation encourages rapid upgrades in enterprise IT architectures. Several industries including finance, carrier, healthcare, and Internet have already started multi-cloud deployment and gained insights into best practices.

Trend Analysis

In cloud evolution, having a combination of multiple public and private clouds has become the best choice

Cloud transformation has been accepted by a wide range of industries. It helps large enterprises take IT construction to a whole new level. It also enables small- and medium-sized enterprises to implement digital management at lower costs, instead of being overwhelmed by complex IT architecture construction.

As cloud evolution practices evolve, enterprises are keen to find the best construction mode. Initially, many enterprises preferred a single-cloud platform. However, as more types of services are deployed, the architecture of the single-cloud platform faces an increasing number of problems. On the one hand, different cloud vendors have their own strengths. On the other hand, the use of a single cloud provider increases enterprises' concerns about data security. Therefore, going multi-cloud, which includes multiple public and private clouds, has become an important strategy for enterprises to consider in cloud evolution. Enterprises can deploy public clouds or build their own data centers to meet workload requirements.

The public cloud appeals to IT decision makers because it offers lower costs and reduced risks, as well as good elasticity. Innovative and emerging services can be migrated from on-premises systems to public clouds in order to take full advantage of the benefits offered by public cloud vendors. In addition, local public clouds meet destination countries' data residency and regulatory requirements, helping multinational enterprises further expand their overseas business. However, the use of multiple public clouds exposes enterprise data to public environments, which may not meet data management, control, and privacy requirements. By contrast, enterprise-built data centers allow for independent data management, control, and operation, prevent core services from being locked in by vendors, and fully use the capabilities of enterprise-built

platforms, helping enterprises take the initiative in cloud construction. Therefore, a combination of multiple public and private clouds has become the best choice for enterprises in cloud evolution.

Different public and private cloud vendors have their own advantages in IaaS, PaaS, and SaaS. To meet workload requirements, many enterprises are trying to deploy different applications on suitable cloud platforms in order to procure an optimal combination of IT technologies. In addition, global economic uncertainty threatens enterprises' single-supplier strategy. The strategy of using multiple cloud suppliers can help enterprises avoid vendor lock-in and build a more robust IT platform. Statistics show that 89% of global enterprises have strategic plans for multi-cloud deployment. Enterprises are actively embracing multi-cloud construction and relying on cloud management services to improve the multi-cloud experience and efficiency.

Enterprises are making efforts to remove bottlenecks in data sharing, data mobility, and unified data management across clouds

After an enterprise goes multi-cloud, data is scattered on different cloud platforms. Applications cannot invoke data across clouds. As a result, multiple data silos are created, preventing the enterprise from constructing a data-centric digital architecture. Therefore, some enterprises are seeking a new multi-cloud data architecture that centrally stores and shares data and deploys applications in multiple clouds. To share data more efficiently, acceleration engines for diverse applications are also being built on the storage side to accelerate cloud-based application deployment.

Based on application types and data security needs, enterprises attempt to deploy different

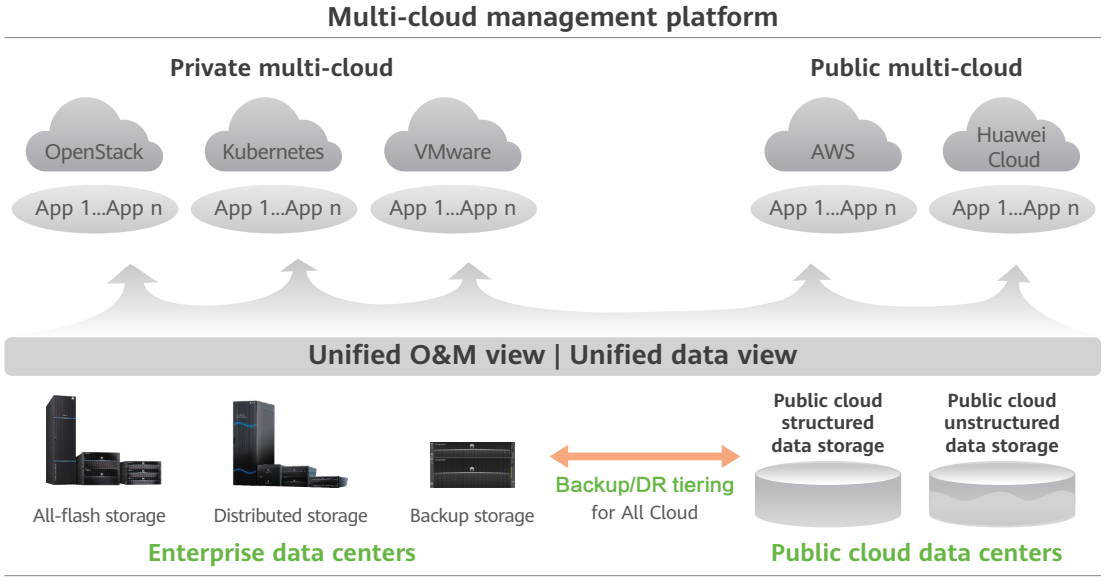

Figure 10: Enterprise multi-cloud IT architecture

applications and data in enterprise data centers or public clouds. This requires enterprises to implement unified data management and mobility across clouds for optimal efficiency in data storage and management. Unified data management refers to unified storage resource provisioning and O&M via a holistic data view. Data mobility includes global data scheduling and hybrid-cloud DR, backup, and storage tiering. Storage vendors are also actively deploying professional storage, which is widely used in data centers, on public cloud platforms by means of software-hardware integration or pure software, helping enterprises with smooth cross-cloud evolution.(Figure 10: Enterprise multi-cloud IT architecture)

To continue to further digital transformation, enterprises going multi-cloud will need to continue exploring ways to maximize resource sharing across clouds and data centers, implement unified data management, and benefit from an optimal combination of cloud vendors.

Suggestions

Migrate innovative services that have uncertainties and emerging services like OA to public clouds, while retaining core services in their on-premises data centers

Innovative and emerging services require IT systems to deliver elastic scaling and on-demand resource application and release. Public clouds provide benefits like low costs and reduced risks in addition to good elasticity. Enterprises are also encouraged to trial public cloud services in fields like DR, backup, and cold data retention to benefit from the elasticity and large-scale sharing advantages of public clouds.

To maintain the competitive edge of core services, enterprises must develop strong capabilities in IT platform R&D. However, public clouds only

provide standard IT services. Therefore, retaining core services in enterprises' on-premises data centers can truly stimulate IT R&D and innovation, implement independent data control and operation, and prevent cloud vendor lock-in, helping enterprises take the initiative in building digital competitiveness.

In multi-cloud construction, enterprises are advised to use the IT architecture that centrally stores and shares data and deploys applications in multiple clouds, and plan a unified data management platform across clouds to maximize data sharing

When selecting storage to build a unified data management platform across clouds, consider the following points:

- Leverage the technical advantages of multiple cloud vendors for faster application development and select an IT architecture with application- storage decoupling for data sharing across clouds.
- Promote storage vendors to deploy professional storage on public cloud platforms by means of software-hardware integration or pure software, thereby helping with smooth cross-cloud evolution.
- Plan a unified data management platform across clouds to implement a holistic data view, global data scheduling, and hybrid-cloud DR/backup/tiering, so that data is stored by service level on suitable storage tiers with efficient data mobility.

Enterprises and storage vendors actively collaborate on innovation projects based on the latest cloud technologies like containers

As databases, AI-powered analytics, middleware, and content services develop, enterprises need to work on shortening their applications' time to market (TTM) and reconstructing applications into microservices. Containers are the best carrier of the microservice architecture and have been adopted by most industries, especially the finance and Internet industries.

Containers are evolving from stateless to stateful applications. Therefore, persistent storage is needed to carry containers. In addition, storage systems need to work with container management platforms to enable large-scale provisioning of containerized applications. Disaster recovery of containerized applications also requires close cooperation between container management platforms and storage systems. Therefore, enterprises and storage vendors should actively embrace the latest cloud technologies (such as containers) and carry out joint innovation projects to fully unlock the value of multiple clouds.

Outlook IX

Energy Saving

Green Data Storage: A Must for Data Centers to Reach Net-Zero Carbon Emissions

Organizations everywhere are looking to meet carbon peak and neutrality goals, and this starts with data centers. To build sustainable data centers, new innovative ways much be adopted to reduce the energy consumption of IT equipment, particularly, storage devices, in addition to lowering power usage effectiveness (PUE).

Trend Analysis

Energy-efficient data centers: The key to going carbon neutral

To date, 136 countries that are responsible for 88% of global emissions have pledged to reach carbon neutrality. Large economies such as China, the United States, the European Union, and Japan have developed clear plans and enacted legislation to facilitate the goal.

One crucial aspect is the energy efficiency of data centers. By 2025, it is expected that the total energy consumption of data centers will account for 4.5% of the world's total energy consumption, up from less than 1% in 2010. To build low-carbon data centers, countries around the world have released programs to guide the development and optimize data center operations. Such programs include China's Three-year Action Plan for the Development of New Data Centers (2021-2023) (the Action Plan for short), U.S.' Data Center Optimization Initiative (DCOI), and Japan's Green Growth Strategy. In addition, the green data center market share is expected to increase by USD 76.59 billion from 2020 to 2025, and the market's growth momentum will accelerate at a CAGR of 19.48%. To further this goal, large IT organizations such as Microsoft, Google, and China Mobile have already invested in building green data centers.

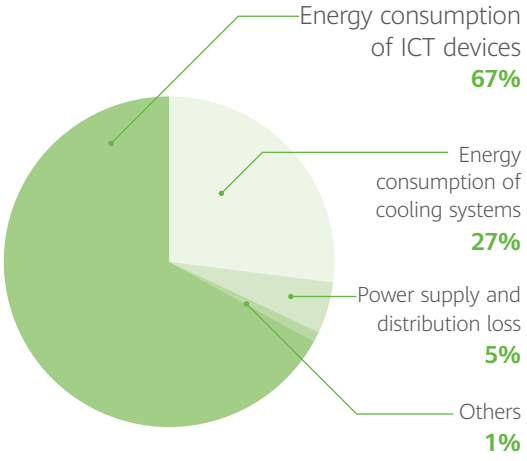

Figure 11: Proportion of power consumed by each component in a data center with a PUE of 1.5

Sustainable data centers: Lower PUE + green data storage

Policies related to green data centers have specific requirements on PUE. For example, the Action Plan requires that new large data centers produce a PUE of 1.3 or under, and the DCOI asks a PUE of 1.5 for existing data centers, and 1.4 or lower for new data centers.

However, reduced PUE is just one step on the road to the low-carbon future. More importantly is the issue of lowering energy consumption of IT facilities. According to a report from the State Information Center of China, ICT equipment consumes 67% of total power supply in a data center, at an average PUE of 1.5. (Figure 11: Proportion of power consumed by each component in a data center with a PUE of 1.5)

Further, storage devices are expected to be the main electricity-drawing IT components. If we consider that the total amount of data created globally is projected to grow to more than 180 zettabytes by 2025 (three times the amount in 2020), the storage of never-before-seen data volumes will cause power consumption levels to skyrocket. For example, the annual power consumption of 1 terabyte data storage in a data center is 300 kWh; however, this will surge to 300,000 kWh when storing 1 petabyte data, which is equivalent to 235.5 tons of carbon emitted. Without an effective green strategy, 2030 levels of carbon emissions caused by storage will easily exceed the total global carbon emissions recorded in 2019.

Fortunately, this issue is now no longer being ignored. Shanghai Pudong Development Bank and Industrial Bank have placed energy efficiency at the center of their storage deployment, while China Academy of Information and Communications Technology (CAICT) has set up a zero-carbon computing power co-construction plan to evaluate low-carbon data center products and solutions, and certify green storage and other sustainable IT facilities.

Technological innovation in storage energy saving: catalyst for low-carbon development of data centers

Under increasing pressure from storage energy consumption, storage vendors are making efforts to turn data centers green via technological innovation, which in turn contributes to sustainable development.

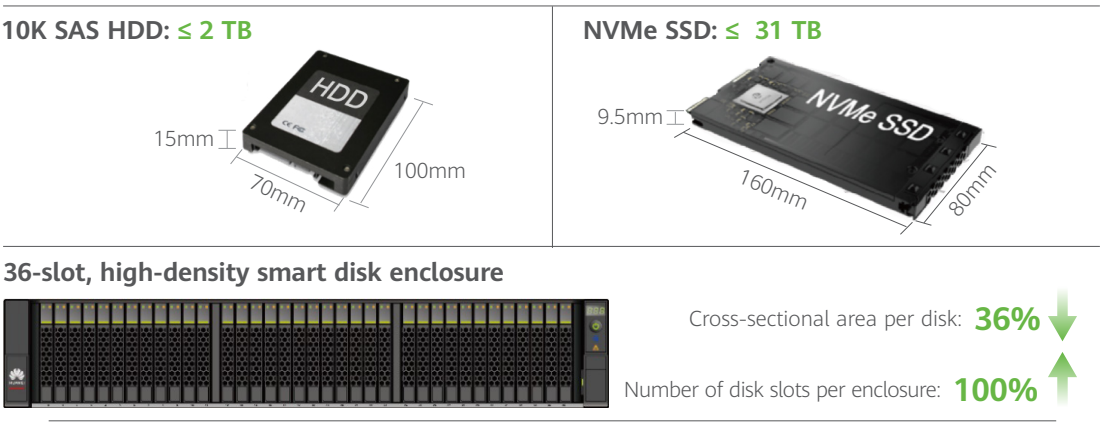

Figure 12: High-density storage designs

1.Develop storage products with high-density designs, system convergence, and data reduction.

- **High-density designs:** A storage product equipped with large-capacity SSDs and high-density disk enclosures can store the same amount of data with less energy, meaning lower power consumption per unit capacity. (Figure 12: High-density storage designs)
- **System convergence:** Multi-protocol convergence and silo convergence enable one-for-all storage and improve resource utilization. One storage system supports multiple protocols like file, object, and HDFS to meet diversified requirements and integrate multiple types of storage. In addition, converged resource pools implement resource pooling to improve resource utilization.
- **Data reduction:** Deduplication and compression algorithms greatly reduce the amount of stored data without information distortion, helping data centers use less power.

2.Develop new-generation storage products that are powered by large-capacity persistent memory.

Research shows that moving data along the lengthy path between a processor and memory contributes to more than 63% of a device's power consumption. Currently, storage vendors are developing storage products equipped with large-capacity persistent memory to shorten the data transfer distance and reduce transfer times. This helps slash power consumption.

3.Promote energy saving through storage lifecycle management.

In the storage manufacturing phase, manufacturing plants widely use photovoltaic power generation, zero wave soldering, paperless labeling, and renewable materials such as aluminum and tin. In the storage use phase, intelligent O&M based on AIOps enables on-demand use of storage resources. In addition, a proper recycling system can be established to ensure that at the end of the storage product lifecycle, e-waste is handled in an environmentally friendly manner for optimal recycling and minimal environmental impacts.

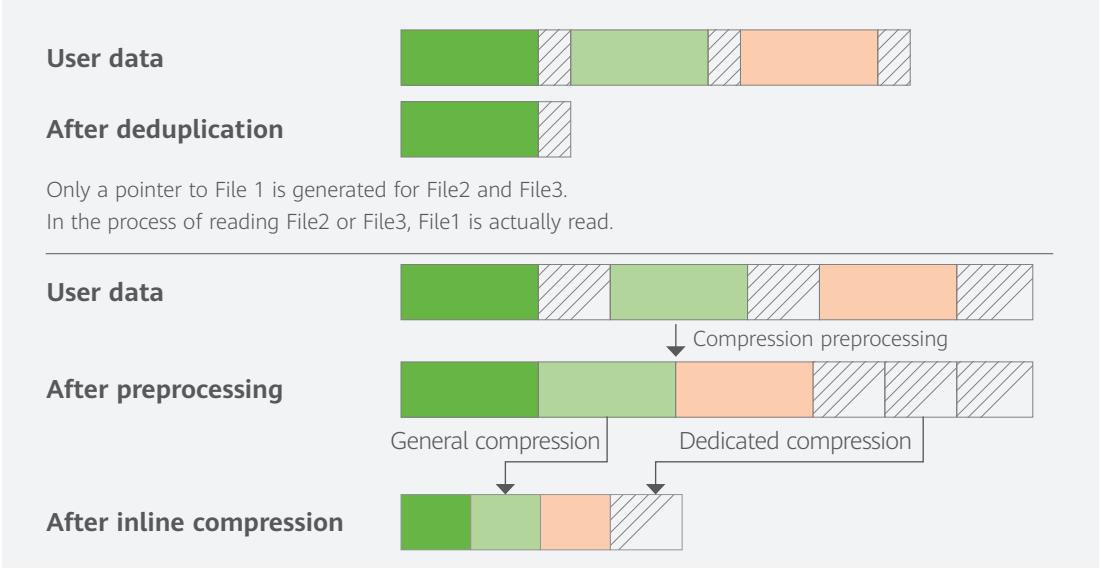

Figure 13: Deduplication and compression technologies

Suggestions

Strike a balance between storage performance and energy saving

According to McKinsey & Company, although up to 61% of consumers are willing to pay a premium for low-carbon products, more are concerned about products' direct benefits (health and nutrition). We can conclude that customer experience is still the top consideration in storage construction. Enterprises should strike a balance between storage performance and energy saving to form a positive cycle of business and environmental protection, thus building the foundation for green storage.

Actively promote storage vendors to innovate for lower power consumption

Enterprises are encouraged to proactively deploy storage products powered by energy-saving technologies and push storage vendors to innovate, for example, in hardware density at component and device levels for higher density and better heat dissipation. New-generation storage products can be equipped with large-capacity persistent memory to reduce energy consumption caused by data transfers.

Building a Data-Centric, Trustworthy Storage Foundation for Diverse Applications

In order to effectively protect valuable data assets, professional storage devices must be fast and reliable. Over the past few years, Huawei OceanStor Dorado all-flash storage has become popular across multiple industries thanks to its high performance and reliability. More than 40 of the world's top 100 banks have chosen Huawei OceanStor Dorado to go digital.

We are entering the yottabyte data era, where more data applications are being released and more data is being produced than ever before. Against this backdrop, Huawei data storage will build a data-centric, trustworthy storage foundation to drive digital transformation.

Huawei's OceanStor Dorado all-flash storage enhances NAS capabilities and provides industry-leading SAN features. Both NAS and SAN are built on the FlashLink algorithms and SmartMatrix architecture. FlashLink algorithms and end-to-end NVMe deliver a latency as low as 0.05 milliseconds, and the SmartMatrix architecture enables high tolerance to hardware failures and only 3 seconds of service interruption every year. OceanStor Dorado provides the industry's only integrated SAN and NAS active-active solution.

OceanStor Pacific distributed storage houses 120 disks per 5 U chassis. Such a high-density design combines with the 22+2 high-ratio erasure coding (EC) algorithm to deliver a 20% higher disk density per unit space than similar products. In addition, with huge breakthroughs in technical architecture, including data flows adaptive to large and small I/O, converged indexing for unstructured data, ultra-high-density hardware, and EC algorithms, a single storage unit can make data analytics 30% more efficient by supporting hybrid workloads across high-performance computing (HPC), big data analytics, and AI computing.

OceanProtect backup storage provides up to 155 TB/h backup bandwidth and 172 TB/h recovery bandwidth, beating the next-best player by 3x and 5x, respectively. OceanProtect's advanced data reduction capability allows you to work with more data for less. Innovative technologies achieve a premium data reduction ratio of up to 72:1 — 20% better than the industry's next best product. To protect against ransomware, Huawei has developed a ransomware protection storage solution that combines primary storage and backup storage. We use machine learning

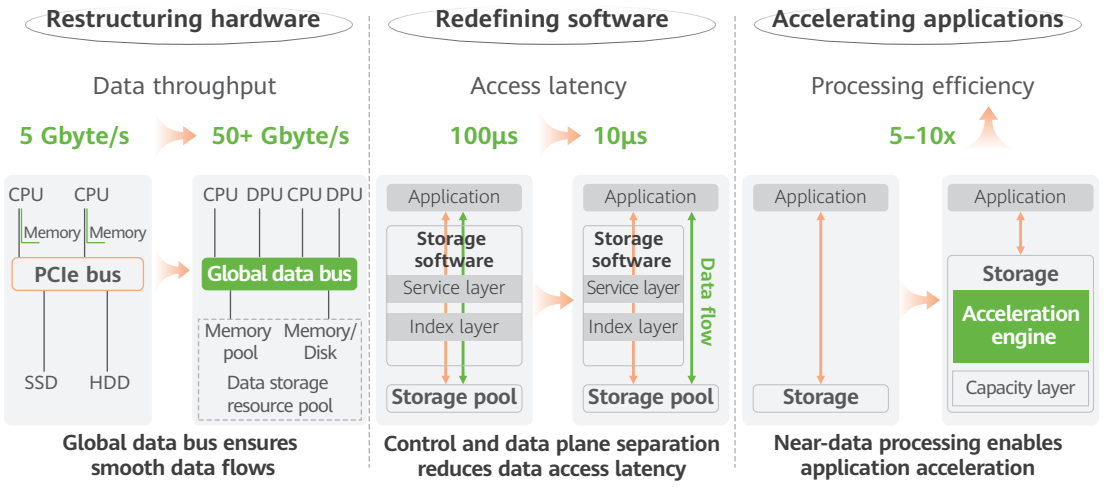

Figure 14: Reshaping storage architectures with a data-centric approach

models to detect ransomware, with a detection rate of 99%.

FusionCube pre-integrates a distributed storage engine, virtualization software, and cloud management software to support on-demand resource allocation and linear expansion.

Data Management Engine (DME) frees O&M personnel from complicated tasks. On average, DME increases resource efficiency by over 20% and shortens service TTM by 10 times, giving personnel more time and resources to focus on what matters most. In addition, DME provides unified management GUIs and standard APIs to centrally manage Huawei and third-party storage devices. Its open APIs, scripts, and plug-ins allow the DME to connect with mainstream cloud management platforms and IT Service Management (ITSM).

Datacenter Virtualization Solution (DCS) combines virtualization software and Information and Communications Technology (ICT) hardware to build lightweight, flexible, diversified, and open data center infrastructure for enterprises. DCS Supports a minimum of two nodes with flexible capacity expansion, supporting large-scale applications. DCS can also add container, DR, backup, and unified management software on demand, reducing costs.

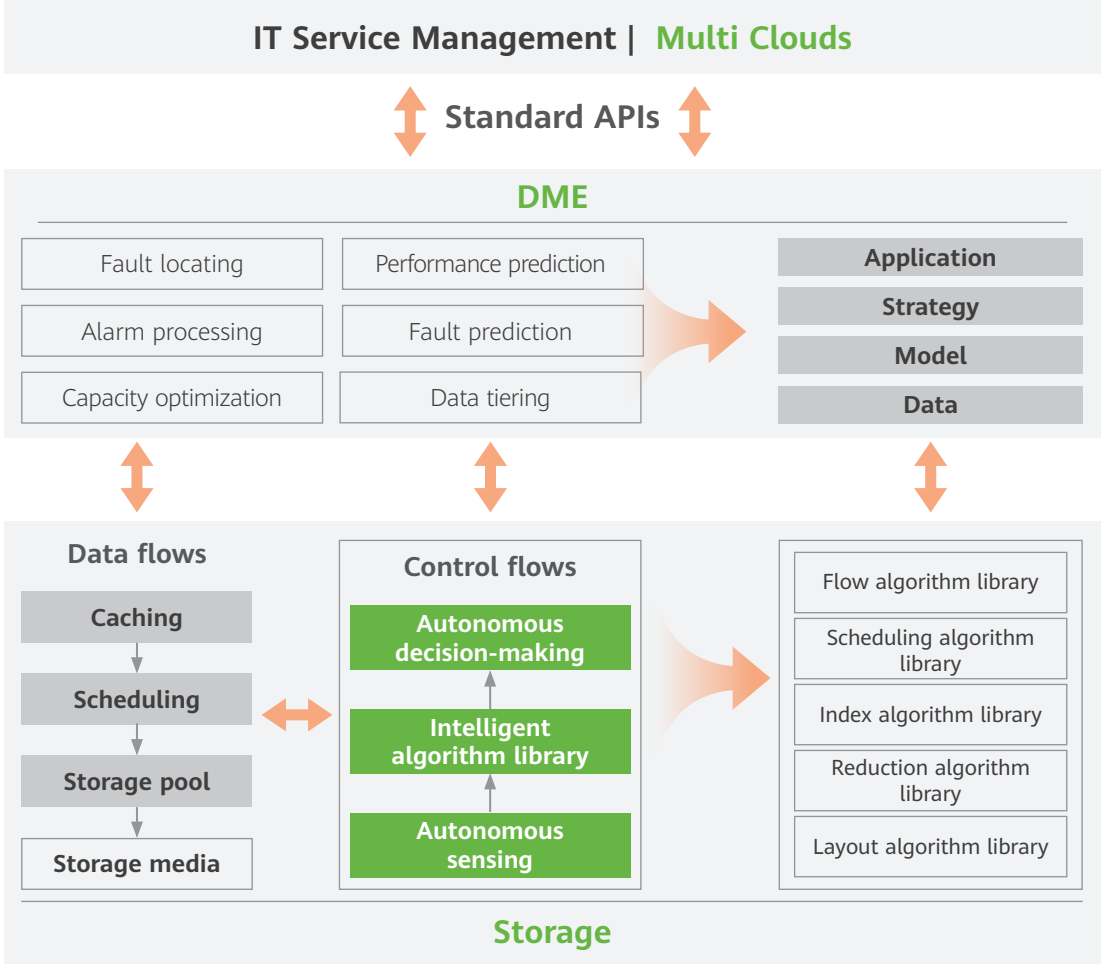

Figure 15: IntelligentStorage: AI-native data storage

Innovation never stops. In the future, storage architecture innovations mainly focus on reshaping hardware and software architectures through a data-centric approach and using various data application acceleration engines. Together, these can improve storage performance by 10 times.

- Currently, data flows across two buses: the Double Data Rate (DDR) memory bus and the PCIe system bus. A global data bus with high throughput allows us to integrate these two buses, greatly improving data flow efficiency.(Figure 15: IntelligentStorage: AI-native data storage)

- In the past, data was indexed and cached at each layer, from application and computing to network and storage. In the future, control and data plane separation will greatly reduce data access latency between applications and the storage pool.

- For all applications, including service logic and data processing, we can place processing power near the storage system based on application features. This will enable near-data processing and accelerate applications.

As data storage develops, enterprises aim to provide a consistent experience both on and off the cloud and enable real-time data processing and automatic management. To meet these requirements, we propose the IntelligentStorage solution, which embeds AI into data storage. Aiming to enable autonomous and intelligent data storage, IntelligentStorage enables us to:

- Support autonomous decision making in scenarios like the layout, scheduling, and reduction of different types of data within storage devices.

- Intelligently allocate CPU and memory resources to overcome local resource bottlenecks.

- Enable automatic root cause analysis for unknown faults across all scenarios.

- Enable data flow and application deployment in a multi-cloud environment.

Appendix

1. **Net Zero Tracker**
   https://zerotracker.net/https://zerotracker.net/
2. **Energy consumption and emission mitigation prediction based on data center traffic and PUE for global data centers**
   https://www.sciencedirect.com/science/article/pii/S2096511720300761
3. **On Global Electricity Usage of Communication Technology: Trends to 2030**
   https://www.mdpi.com/2078-1547/6/1/117/htm?utm_source=morning_brew
4. **Green Data Center Market by End-user and Geography - Forecast and Analysis 2021-2025**
   https://www.technavio.com/report/green-data-center-market-industry-analysis
5. http://www.sic.gov.cn/archiver/SIC/UpFile/Files/Default/20220527100509602351.pdf
6. **Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025**
   https://www.statista.com/statistics/871513/worldwide-data-created/
7. **How To Identify And Break Down Tech Silos In IT**
   https://www.advsyscon.com/blog/break-down-silos-in-it/
8. **https://www.mckinsey.com.cn/**

HUAWEI TECHNOLOGIES CO., LTD.
Huawei Industrial Base
Bantian Longgang
Shenzhen 518129, P.R. China
Tel: +86-755-28780808
www.huawei.com