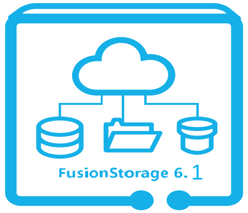
HUAWEI

Huawei FusionStorage Technical Presentation

LEADING NEW ICT

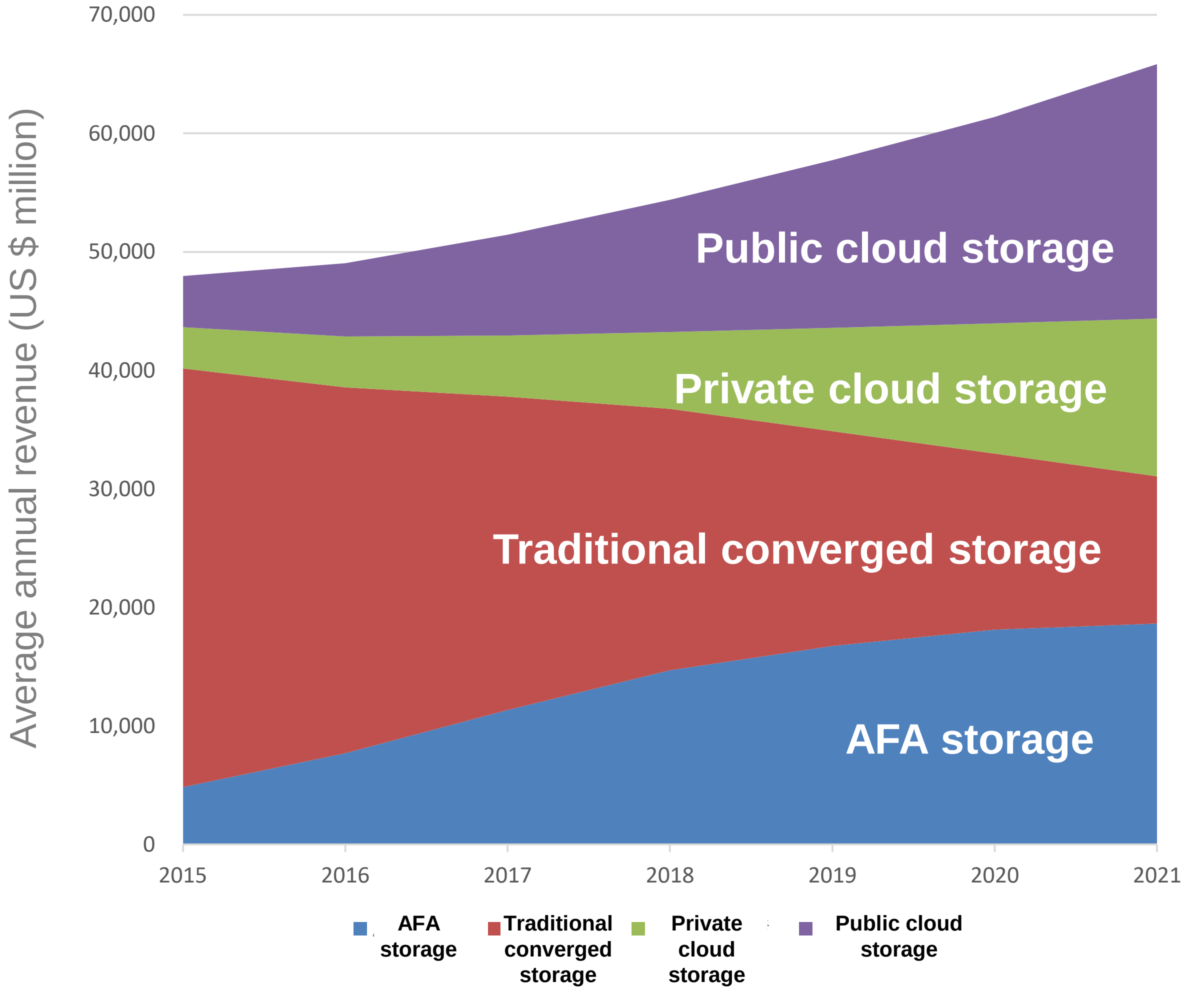


Cloud-based Storage — an Irresistible Trend

Trend

Global Storage Market Development: Cloudification Is a Dominant Trend

LEADING NEW ICT

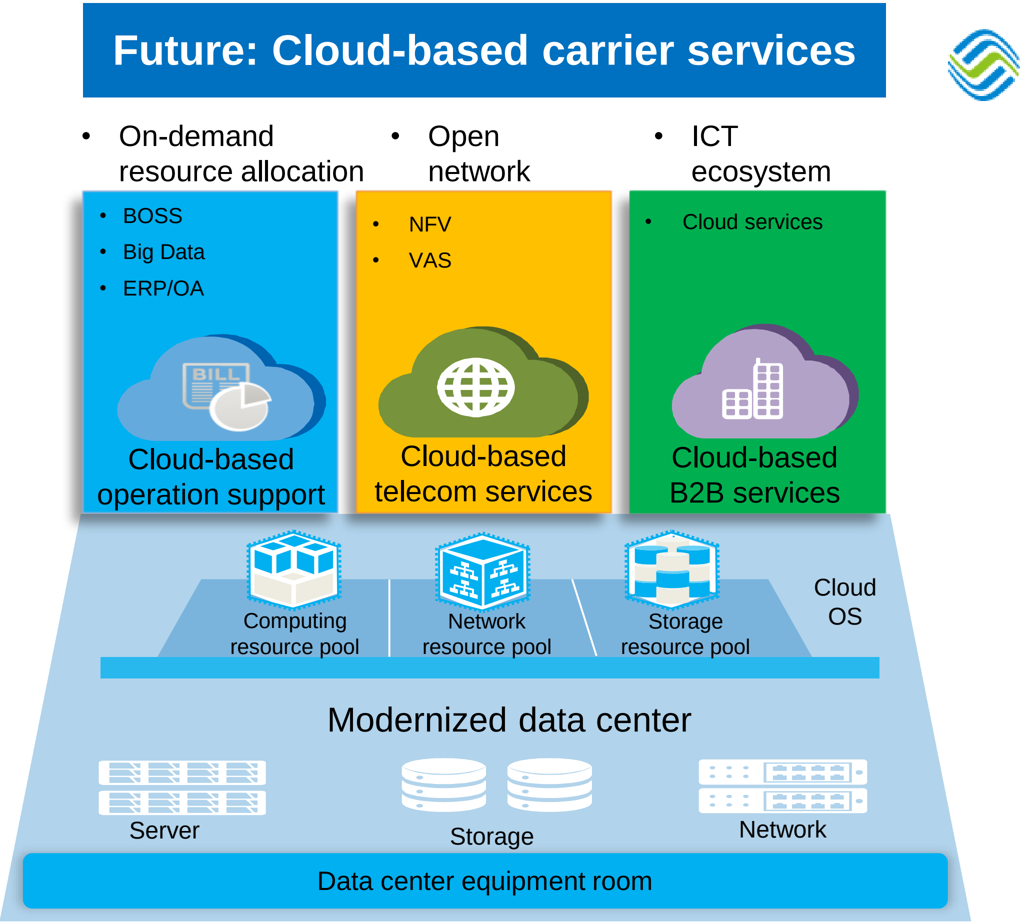


Data source: Gartner, Wikibon, IDC

- **Distributed cloud storage will grow rapidly in the next 10 years. It is estimated that it will account for 70% of the storage market in 2027.**
- Cloud storage devices change from dedicated devices to **universal devices**. They simplify management and reduce TCO, and support large-scale linear expansion.
- Emerging services (such as video, big data, and digital services) are cloudified. With the rapid growth of the cloud storage market, traditional data gradually evolves from offline to online.

HUAWEI

Carrier: IT Infrastructure Cloudified, and Elastic Distributed Storage Pool Preferred



Future: Cloud-based carrier services

- On-demand resource allocation
- Open network
- ICT ecosystem

Cloud-based operation support
- BOSS
- Big Data
- ERP/OA

Cloud-based telecom services
- NFV
- VAS

Cloud-based B2B services
- Cloud services

Cloud OS

Computing resource pool

Network resource pool

Storage resource pool

Modernized data center

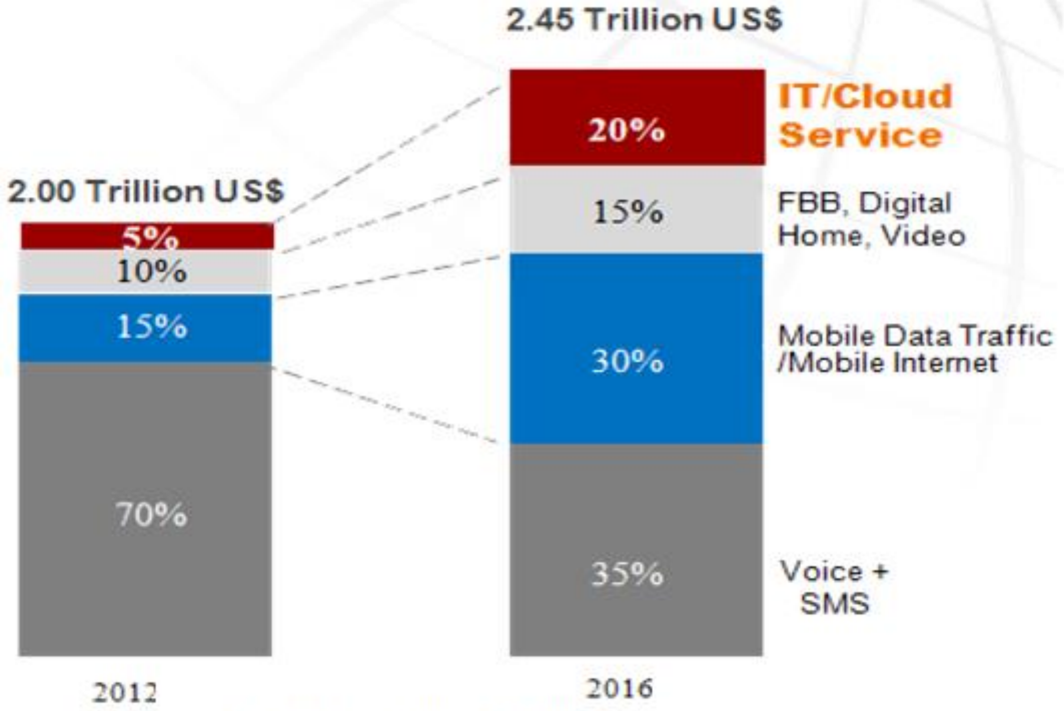
Server

Storage

Network

Data center equipment room

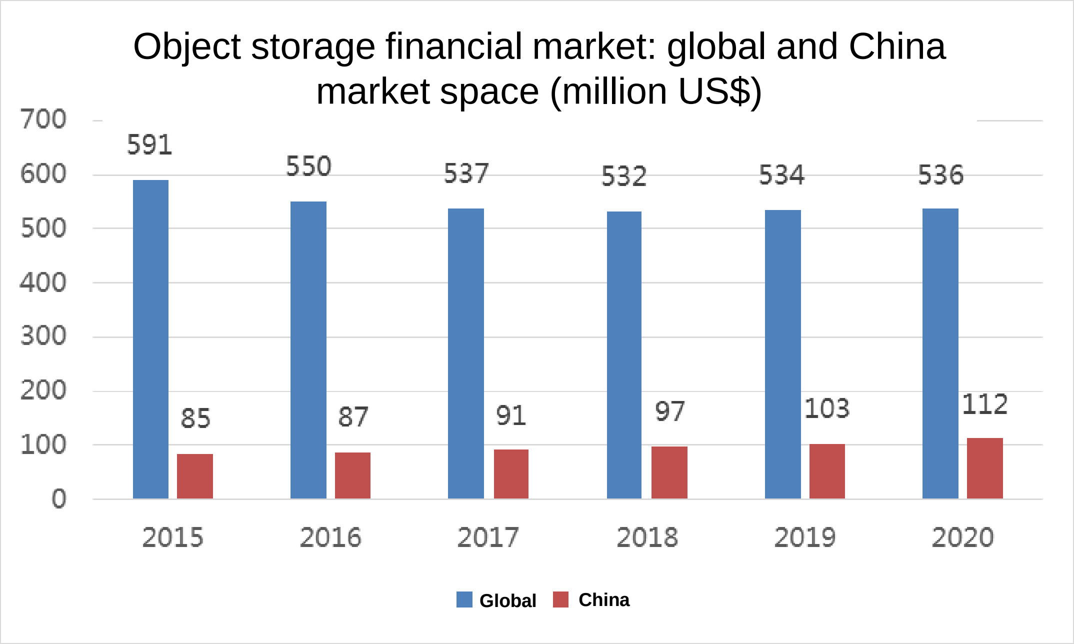
中国移动 China Mobile · vodafone · Telefonica · telenor · 中国电信 CHINA TELECOM

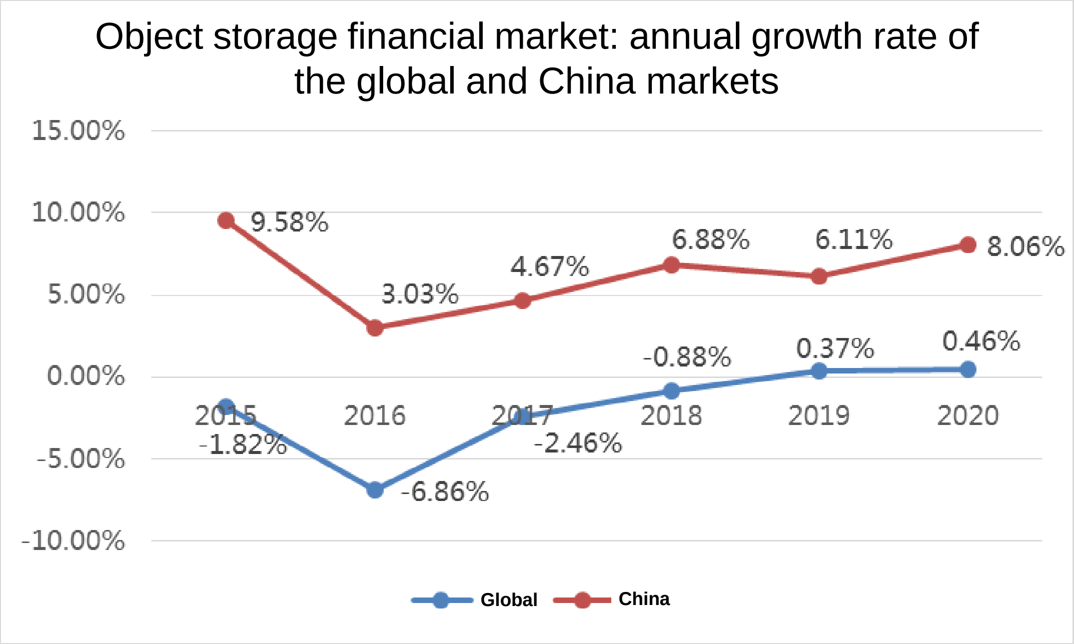


Source: Gartner and Ovum Analysis

- The core of cloud-based IT infrastructure is computing, network, and storage resource pooling as well as flexible provisioning.
- To cope with uncertainties in B2B services and meet requirements of traditional internal ICT on cost efficiency, distributed storage with flexible expansion and large capacity is preferred.

Finance: Object Storage Space Will Reach US $500 Million in 2020, with the Major Growth in China

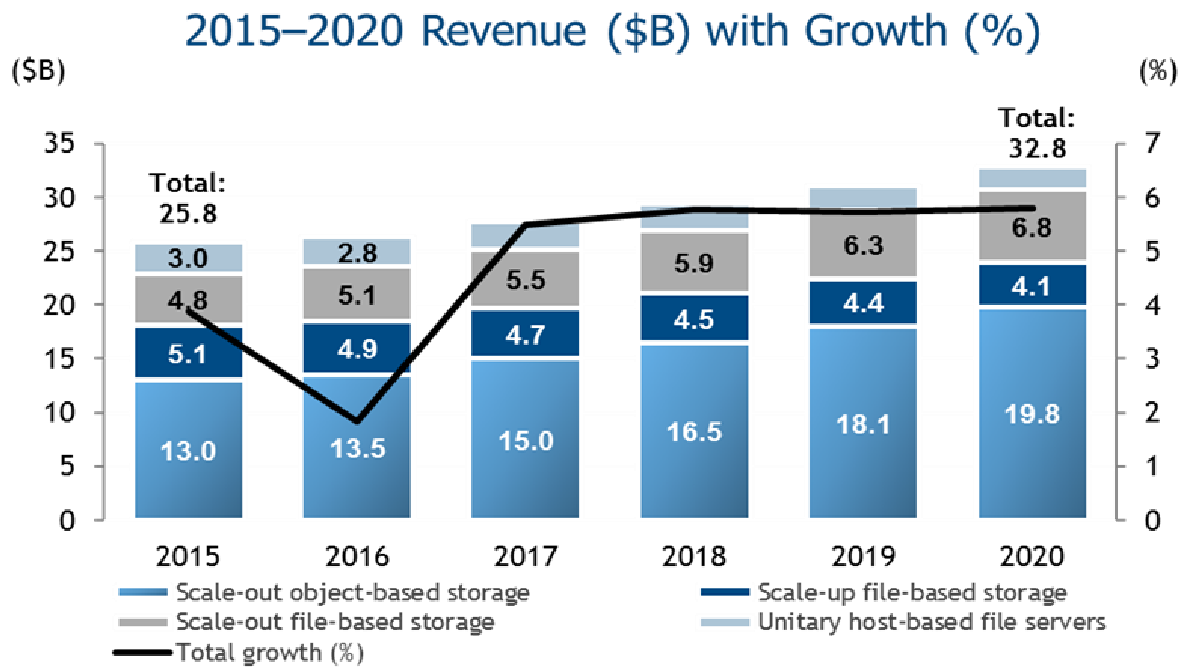


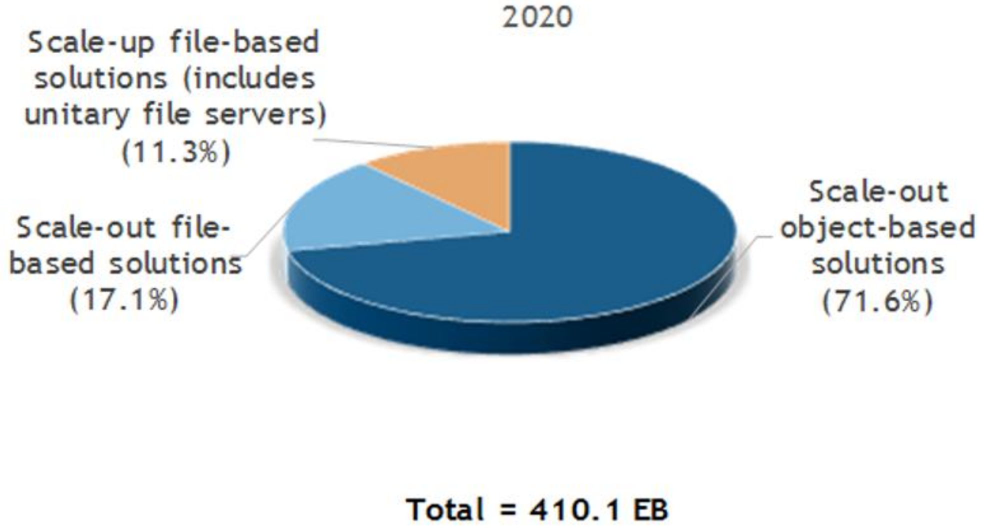


From Gartner: Forecast: Enterprise IT Spending by Vertical Industry Market, Worldwide, 2014-2020, 4Q16 Update, and financial SA interview

- **By 2020, the estimated object storage space (finance) in the world is about US $536 million, with US $112 million in China.**
- **The object storage space of the global financial industry (bank, insurance, and securities) remains stable and will increase slightly from 2019.**
- **Object storage grows majorly in China compared with the global trend.**

Distributed Object Storage Space Enlarges and Grows Rapidly





Selected Segment Growth Rate

- ▲ Scale-out object-based storage CAGR 8.7%
- ▼ Scale-up file-based storage CAGR -4.1%
- ▲ Scale-out file-based storage CAGR 7.3%
- ▼ Unitary host-based file servers CAGR -6.2%

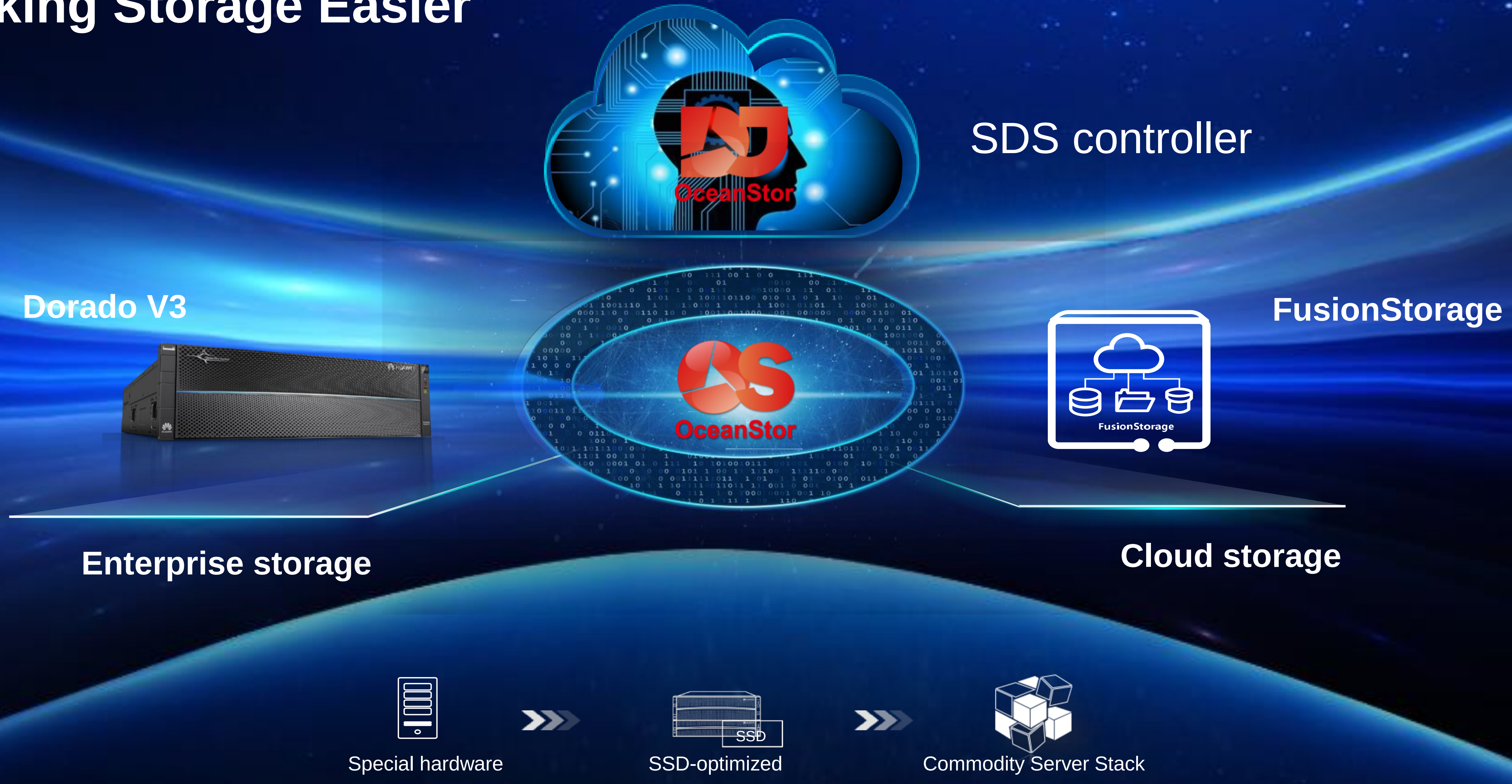
Total Market CAGR 4.9%

- **Gartner predicts that the year-on-year growth of non-structured data is over 40%. By 2021, over 80% of enterprise non-structured data will be stored in scale-out file systems and object storage systems, while the current proportion is 30%.**
- **IDC predicts that the market value of object storage in 2020 is US $19.8 billion, CAGR is 8.7%, capacity reaches 293.6 EB, and CAGR is 30.7%.**

From IDC: World wide for file and object based storage_forecast 2016 to 2020

From Gartner: 2017 Strategic Roadmap for Storage

Huawei Storage Strategy: On-Demand Data Service, Making Storage Easier



SDS controller

Dorado V3

FusionStorage

Enterprise storage

Cloud storage

Special hardware

SSD-optimized

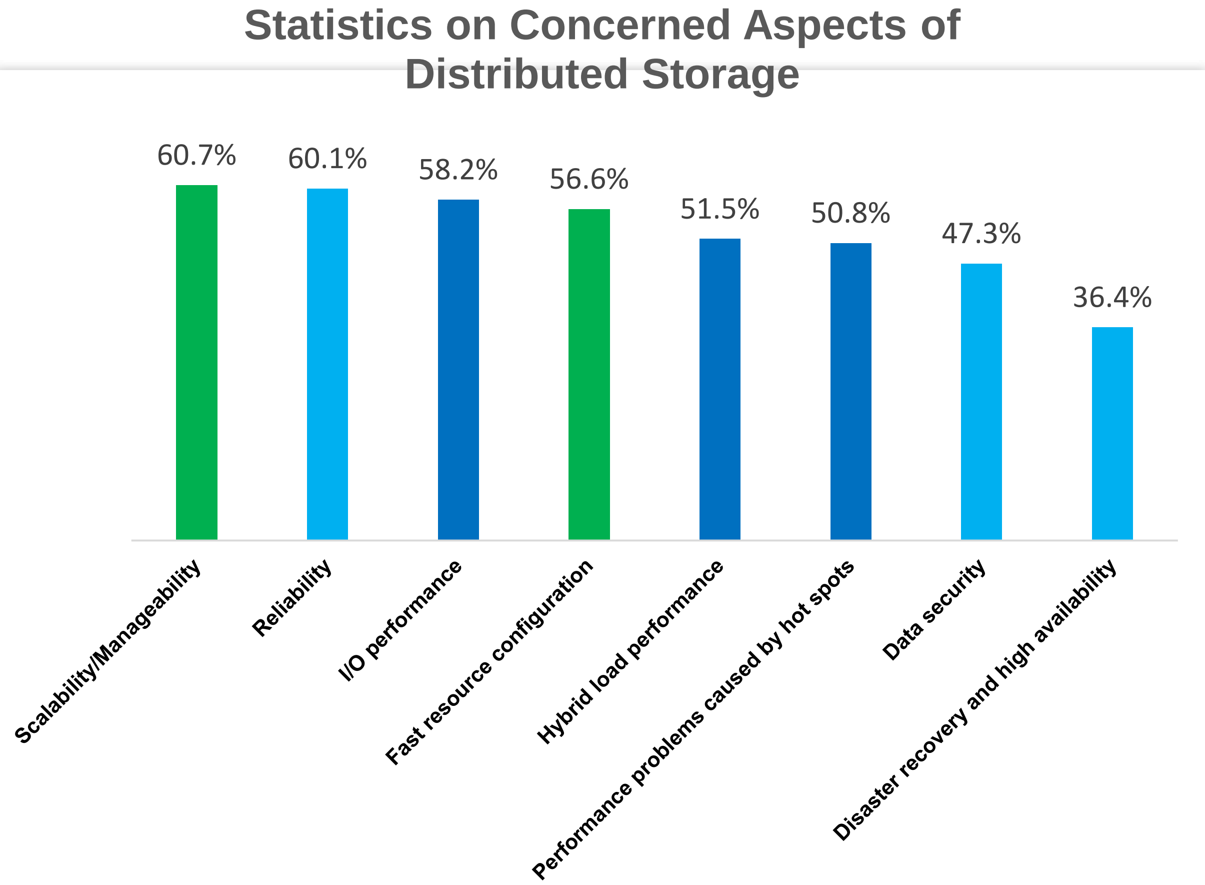
Commodity Server Stack

Storage hardware evolution

FusionStorage 6.1

Huawei FusionStorage Cloud Storage Overview

Key Issues to Be Solved in Distributed Storage



Data source:
130 questionnaires and 50 bidding documents from global carriers, governments, financial institutions, and large enterprises

Three aspects concerned by users

- **Reliability**
- **Performance**
- **Elastic expansion**

Huawei FusionStorage All-Distributed Cloud Storage

Huawei FusionStorage

Most reliable distributed cloud storage

Positioning: FusionStorage is Huawei's next-generation distributed cloud storage. It supports distributed block, object, and file storage services, provides 99.9999% high availability and tens of millions of IOPS, and is the best storage system for the overall cloudification of key enterprise services.

Customer benefit 1: optimal reliability

99.9999% high availability

- ✓ Serving as the largest distributed storage active-active cluster in the industry, it provides solution-level 99.9999% high availability and supports enterprise-level cloud-based key service.
- ✓ **Level 4 high reliability guarantee**: Disk-level, node-level, cabinet-level (block), and data center-level flexible deployment, meeting the high availability requirements of different services.
- ✓ Fast system self-healing: Unique subhealth detection and self-recovery, with the reestablishment speed being smaller than 15 minutes/TB.

Customer benefit 2: outstanding performance

4.5 million SPC-1 V3 IOPS @ < 1 ms

- ✓ It is the first storage that supports NVMe SSD and 4.5 million SPC-1 V3 IOPS@ < 1 ms performance. A single system can meet the performance requirements of millions of VMs.
- ✓ **With distributed EC algorithm**, the storage space usage reaches 90%, which is twice as much as that in multi-copy mode, saving 60% of purchasing costs and achieving high reliability and high efficiency.

Customer benefit 3: elastic and on-demand expansion

Linear expansion of performance and capacity

- ✓ System optimal expansion: The distributed architecture is easy to expand up to 4096 nodes, tens of millions of IOPS, and EB-level capacity to support the expansion of cloud services.
- ✓ On-demand deployment of storage resources: Distributed block, object, and file storage services are supported.

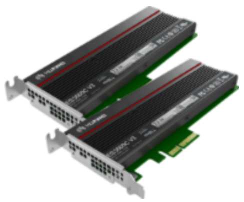
Note: It took the leading place in releasing the SPC-1 V3 test result on 2017-06-08. With the performance of 4.5 million SPC-1 V3 IOPS@0.787ms, it creates a world record at a new benchmark of SPC-1 V3.

http://www.storageperformance.org/results/results_spc1_v3/spc1_v3_active#a31007

Component-Level Reliability Assurance: Comprehensive Detection and Intelligent Rectification

Disk

Hardware fault and slow disk
Excessive SMART information
UNC error



SSD card

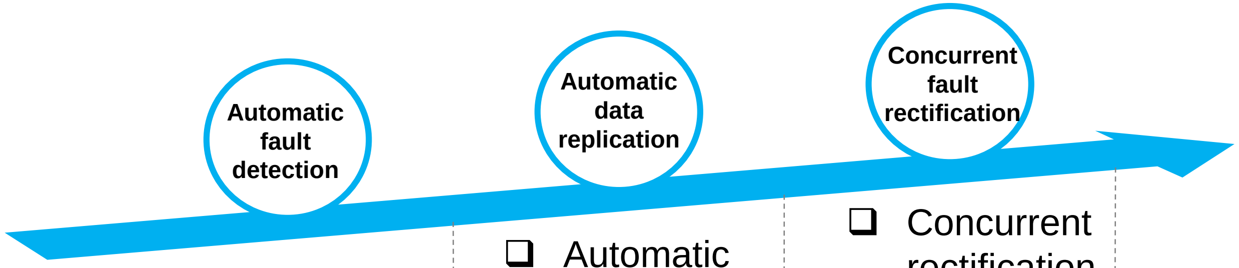
High temperature alarm and high bad block rate
DDR/FLASH access failure
Capacitor fault or SEU fault

Server

CPU fault detection
Intelligent CPU speed control
Too many memory ECCs

Network

Network adapter fault (packet loss or rate reduction)
Network link fault (rate reduction)
Switch fault (packet loss or port subhealth)


Automatic fault detection
Automatic data replication
Concurrent fault rectification

- □ Data damage
- □ Disk damage
- □ Node damage

- □ Automatic replication
- □ QoS
- □ Dispersed storage

- □ Concurrent rectification
- □ Equalized performance
- □ Less transmission

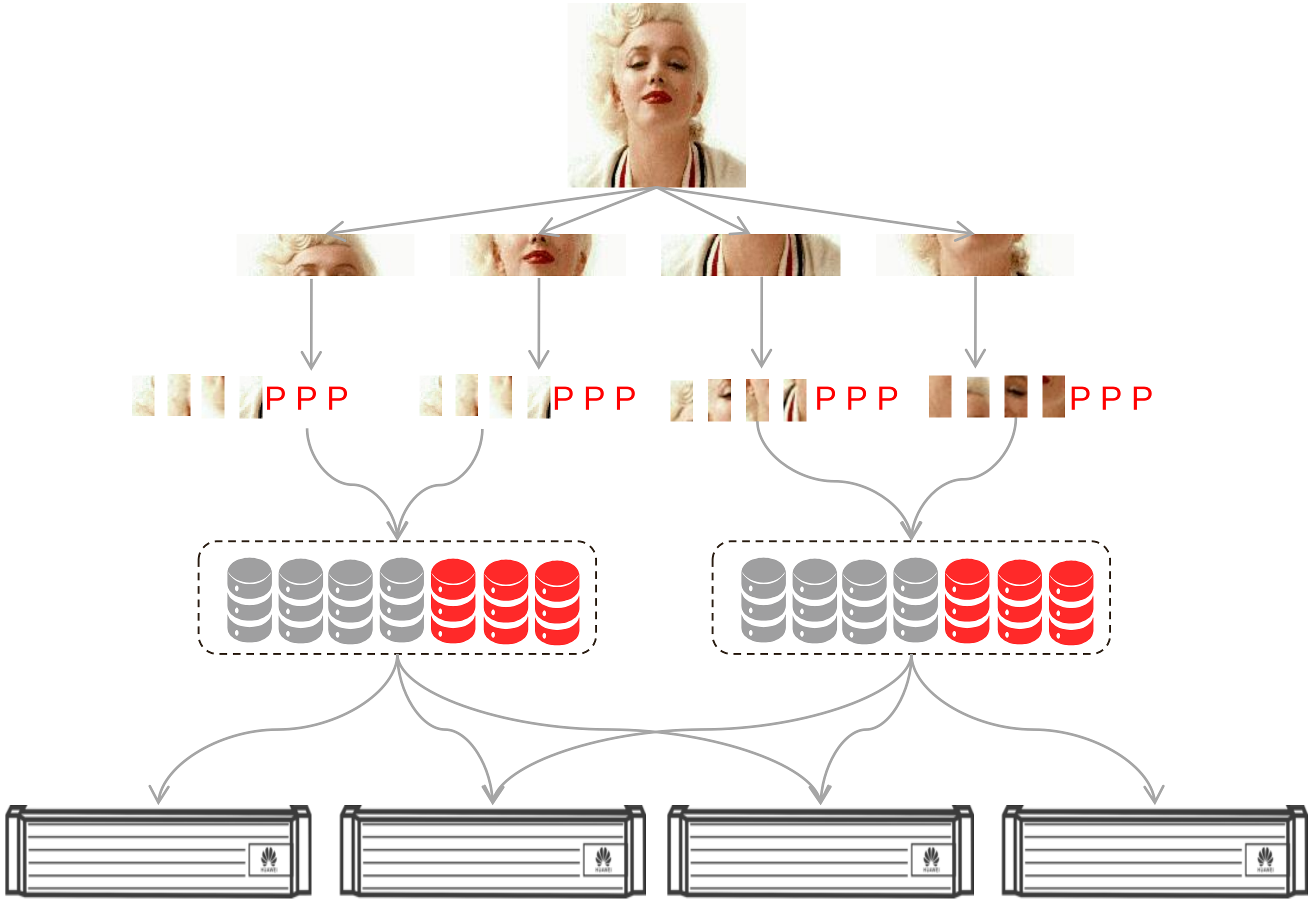
Operations performed concurrently on multiple nodes with fast rectification

QoS is set, not affecting service performance.

DHT algorithm filters out invalid data.

It takes only 15 minutes to restore 1 TB of data, leading the industry.

EC Improves Space Utilization



Files are in disorder and divided into multiple chunks, which are evenly distributed.

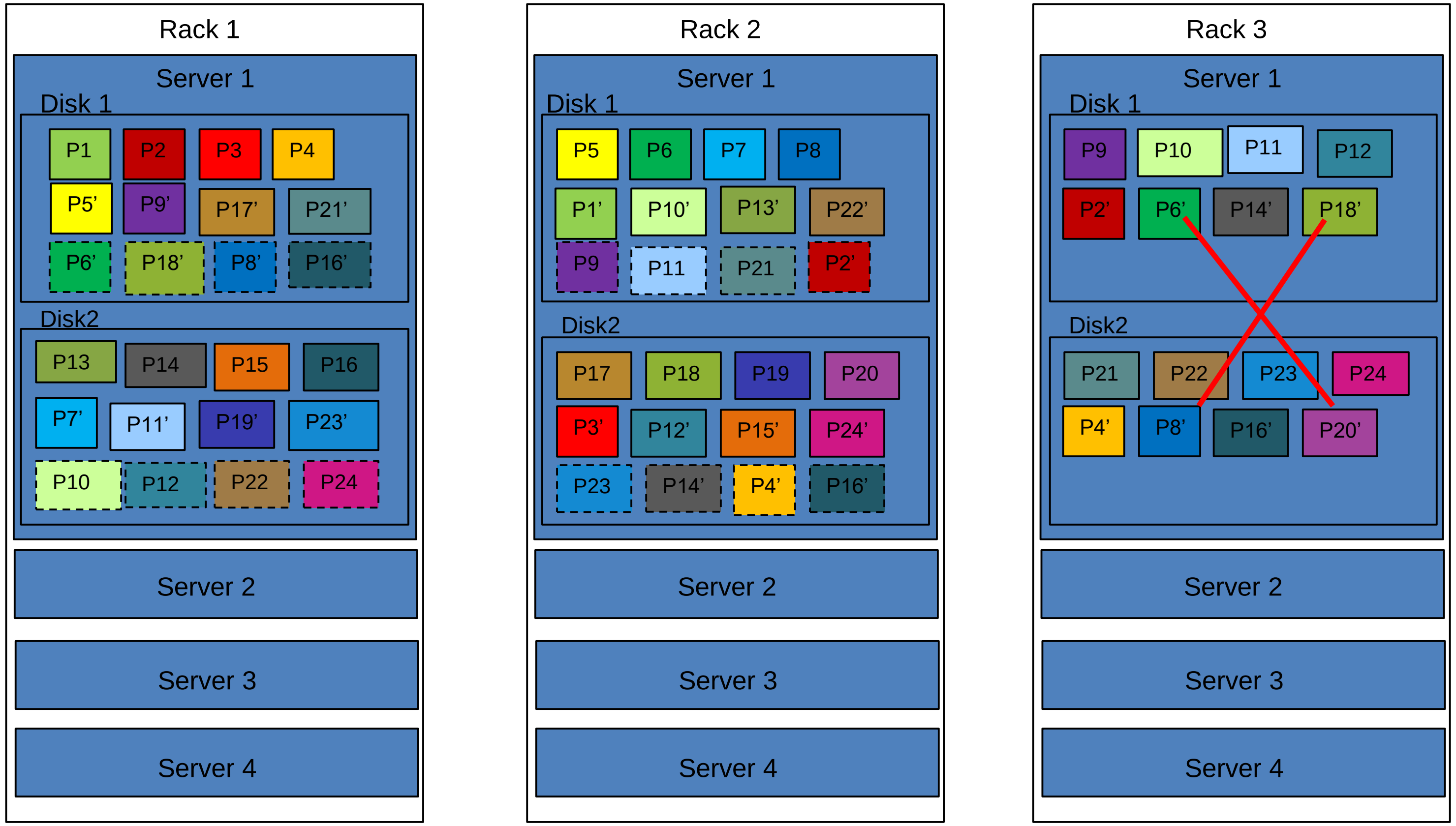
N+M (M ≤ 3) strong redundancy is established based on the erasure code.

Flexible configuration policies, providing more capacity, security, and performance policies in different application scenarios.

SSD medium is used as the EC cache. The small-block I/O write performance is about 40% higher than that of the client.

Intelligent algorithm control. Small-block I/Os use the SSD CACHE as the EC strip cache, and large-block I/Os are directly divided into segments for storage using EC.

Robust Cluster Reliability: Cabinet-Level Security



Rack 1

Server 1

Disk 1

P1 P2 P3 P4

P5’ P9’ P17’ P21’

P6’ P18’ P8’ P16’

Disk2

P13 P14 P15 P16

P7’ P11’ P19’ P23’

P10 P12 P22 P24

Server 2

Server 3

Server 4

Rack 2

Server 1

Disk 1

P5 P6 P7 P8

P1’ P10’ P13’ P22’

P9 P11 P21 P2’

Disk2

P17 P18 P19 P20

P3’ P12’ P15’ P24’

P23 P14’ P4’ P16’

Server 2

Server 3

Server 4

Rack 3

Server 1

Disk 1

P9 P10 P11 P12

P2’ P6’ P14’ P18’

Disk2

P21 P22 P23 P24

P4’ P8’ P16’ P20’

Server 2

Server 3

Server 4

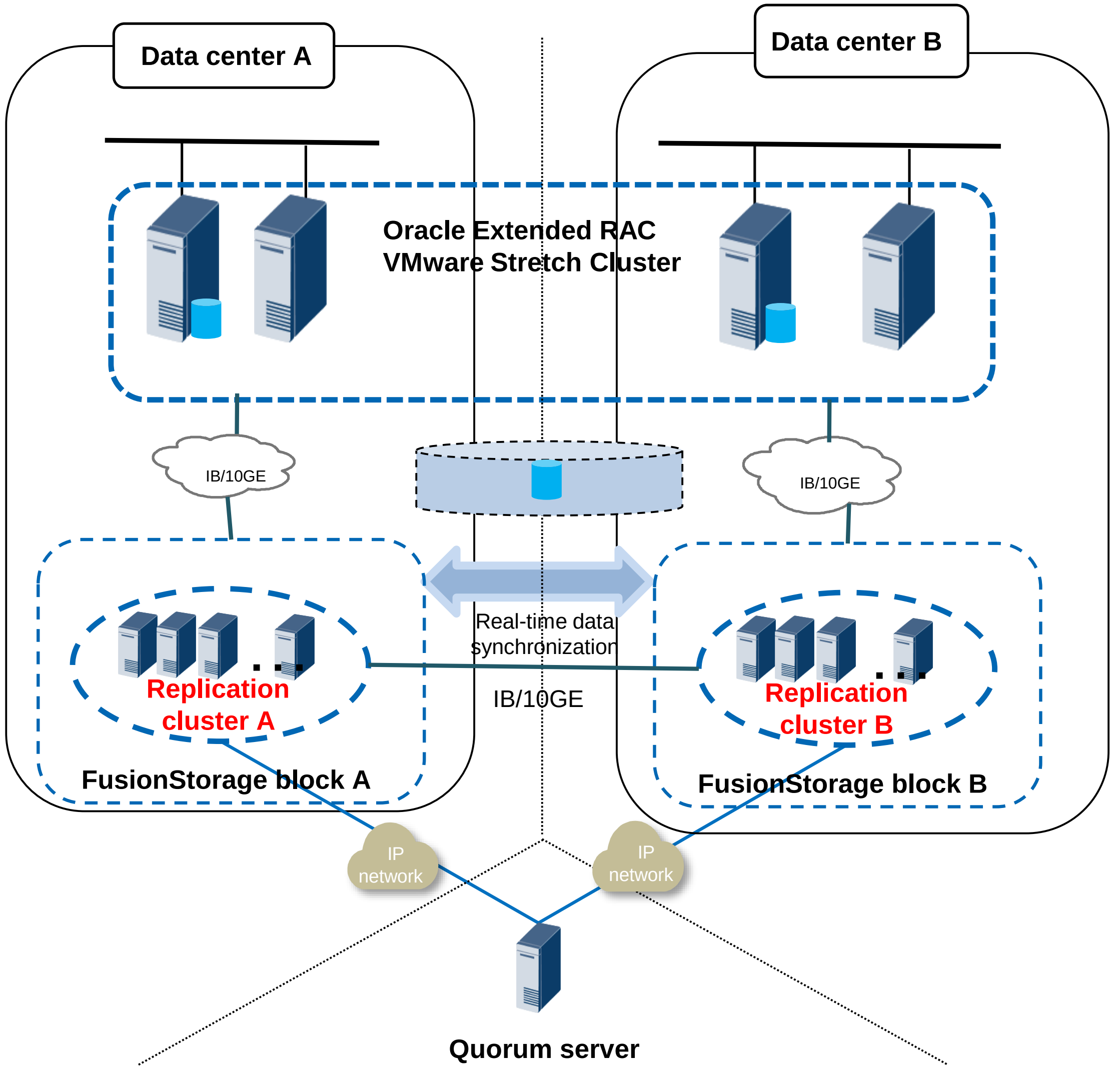
Various security levels: disk-level, server-level, and cabinet-level

Multi-copy mechanism. Data of the active and standby copies is distributed in different cabinets, and two cabinets are allowed to become faulty at the same time.

The system automatically detects faults and automatically triggers data reconstruction.

Automatically adjust the start time of reconstruction based on the data security level, reducing unnecessary reconstruction time.

Active-Active Block Storage, Implementing Solution-Level 99.9999% Reliability



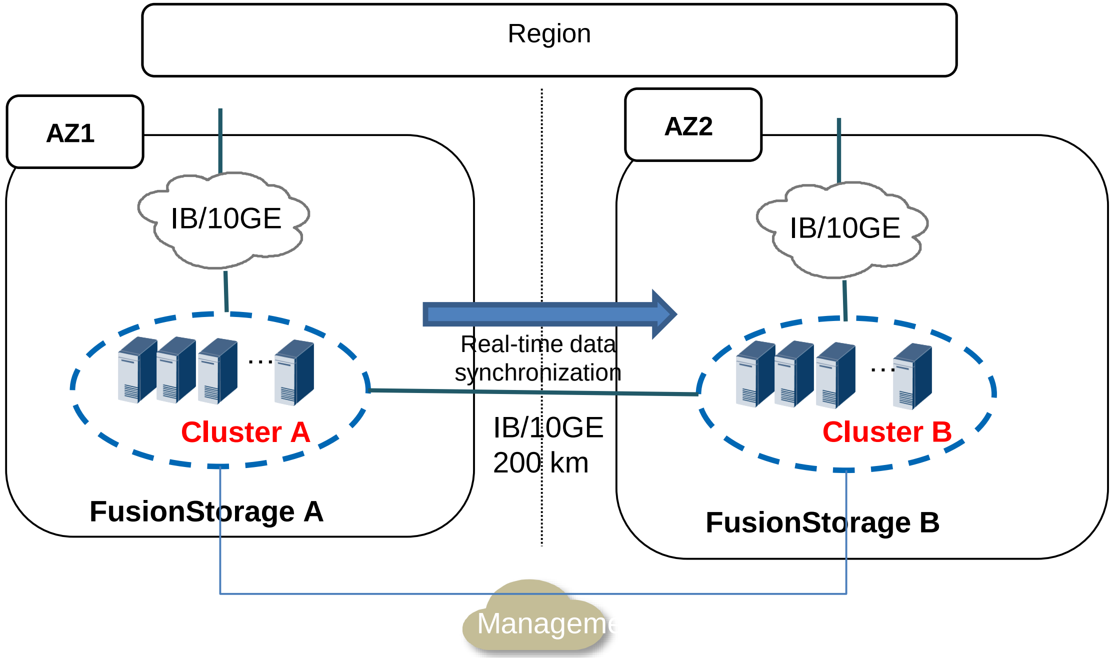
Cloud storage cross-AZ high availability capability building

Superb reliability: All A-A, RTO ≈ 0, RPO = 0, and dual-arbitration mode

Outstanding performance: A single computing node can provide 80,000 active-active read and write performance.

Large scale: Supports 4,864,000 active-active volumes.

Synchronous Object Storage Replication, Implementing Data Disaster Recovery at a Distance of 200 km



Basic principles

- One distributed object storage system is respectively deployed in AZ1 and AZ2 in synchronous replication mode. If AZ1 is faulty, no data is lost.

FusionStorage object storage synchronous replication

- Data uploaded by a user is copied to AZ2 to ensure data consistency between the two AZs.
- When AZ2 is faulty, synchronous replication becomes asynchronous replication. After the fault of AZ2 is rectified, the data difference between AZ1 and AZ2 is automatically resolved.
- You can configure and expand the cluster as required.
- RPO = 0, RTO ≈ 0

What You See Is What You Get with Third-Party Performance Authentication

Rank	Product	Vendor	IOPS
1	FusionStorage	Huawei	4,500,392
3	DS8888	IBM	1,500,187
5	FusionStor SF6000	FusionStack	801,083

SPC-1 certification 4.5 million IOPS@ < 1 ms latency, ranking first in the industry

Integration of software and hardware, achieving ultimate performance of a single node

Maximized performance of a single node

- Supports RoCE network adapters to simplify transmission paths and achieve optimal latency.
- Optimizes PCIe SSD and NVMe SSD of high-performance components.
- Integrates software and hardware to optimize the matched hardware.
- An all flash storage does not need independent cache.

Current situation of peer vendors

- Only 10GE is supported. The network latency is long and congestion easily occurs.
- Supports SSD, but does not optimize high-performance hardware such as NVMe.
- Provides pure software products and there is no hardware perception.
- An all flash storage still requires separate SSD cache, which wastes disk capabilities.

Industry-leading performance expansion

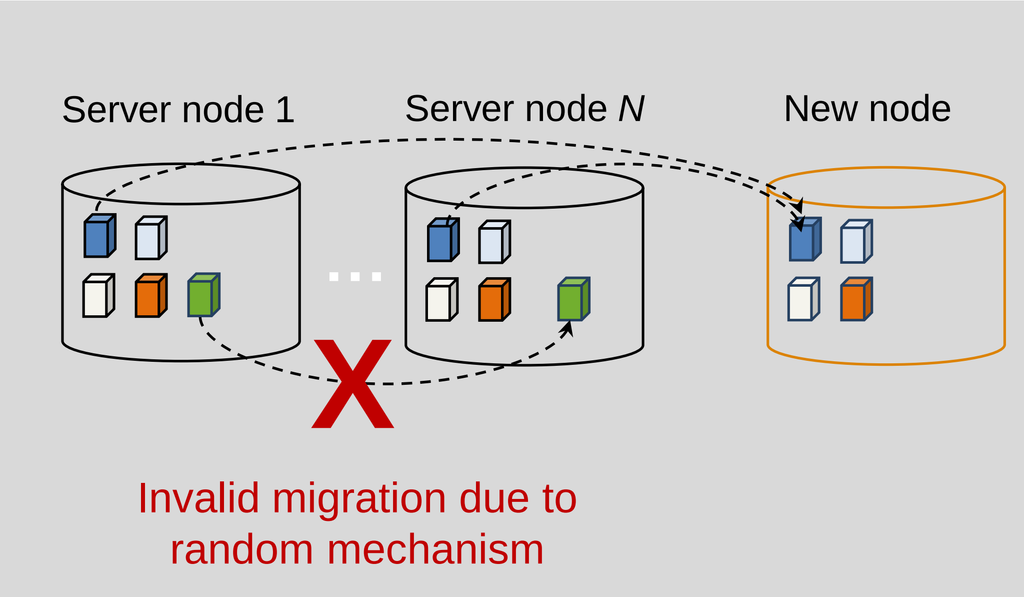
Performance expanded linearly with an increase of node numbers

- The DHT algorithm is used to evenly distribute data. Service pressure is distributed to multiple nodes to fully utilize all disk capabilities.
- Scalable to 4096 storage nodes
- Stateless cluster, no metadata bottleneck, automatic load balancing inside storage, no hot spot, and no bottleneck.

Current situation of peer vendors

- Data of each volume is not scattered and stored in a limited number of disk groups. Therefore, the performance of all disks cannot be maximized. You need to manually strip the data.
- Scalable to 64 nodes
- Hotspot data is prone to become a bottleneck.

Multi-Copy Balanced Partition, Saving Bandwidth for Precise Data Migration



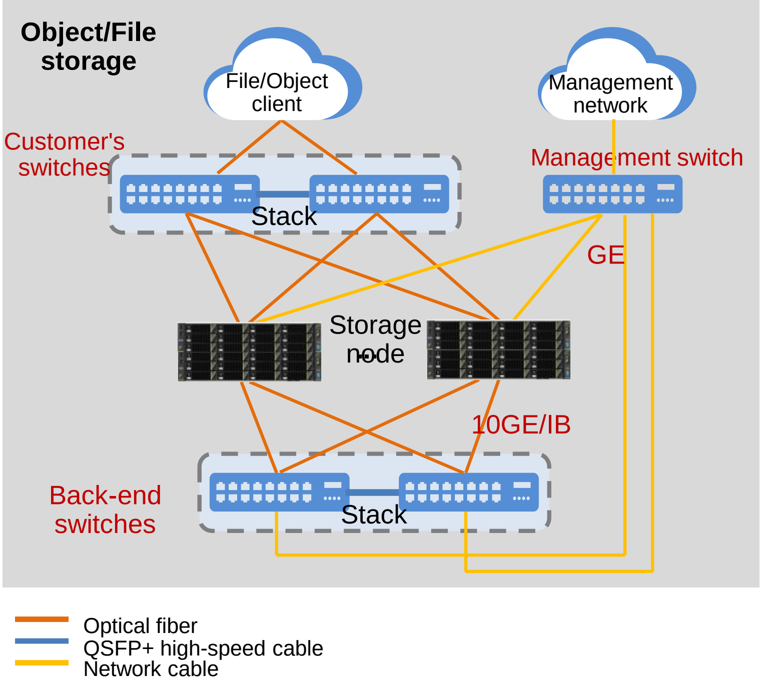
Invalid migration due to random mechanism

The random mechanism of other algorithms (such as CRUSH) causes invalid migration. As the system keeps updating (through capacity expansion and fault recovery), the invalid migration rate is uncontrollable, causing a migration storm.

Intelligent filtering, eliminating migration storms

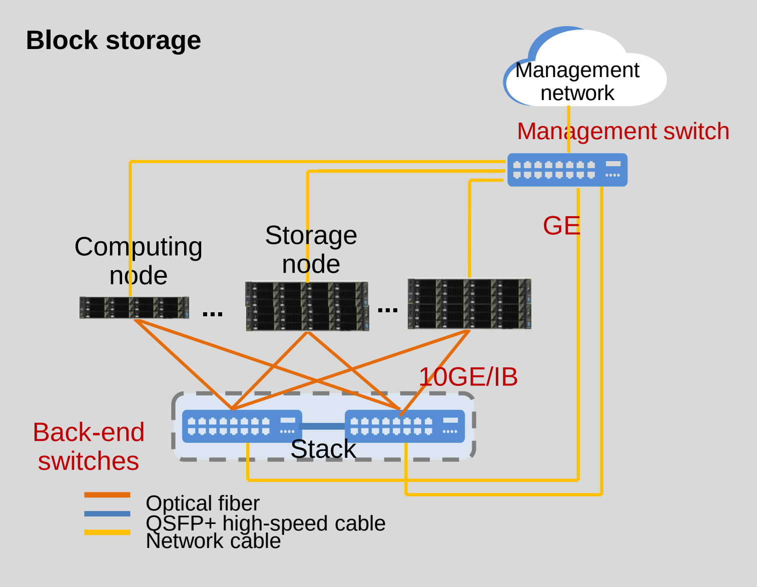
FusionStorage adopts multi-copy balanced partitions: Huawei-developed algorithms ensure balance and automatically filter out invalid migration, improving migration efficiency and reducing impact on services during capacity expansion.

Various Networking Types and On-Demand Deployment



Front-end and back-end networks are separated. Security isolation between the service network and storage network is supported.

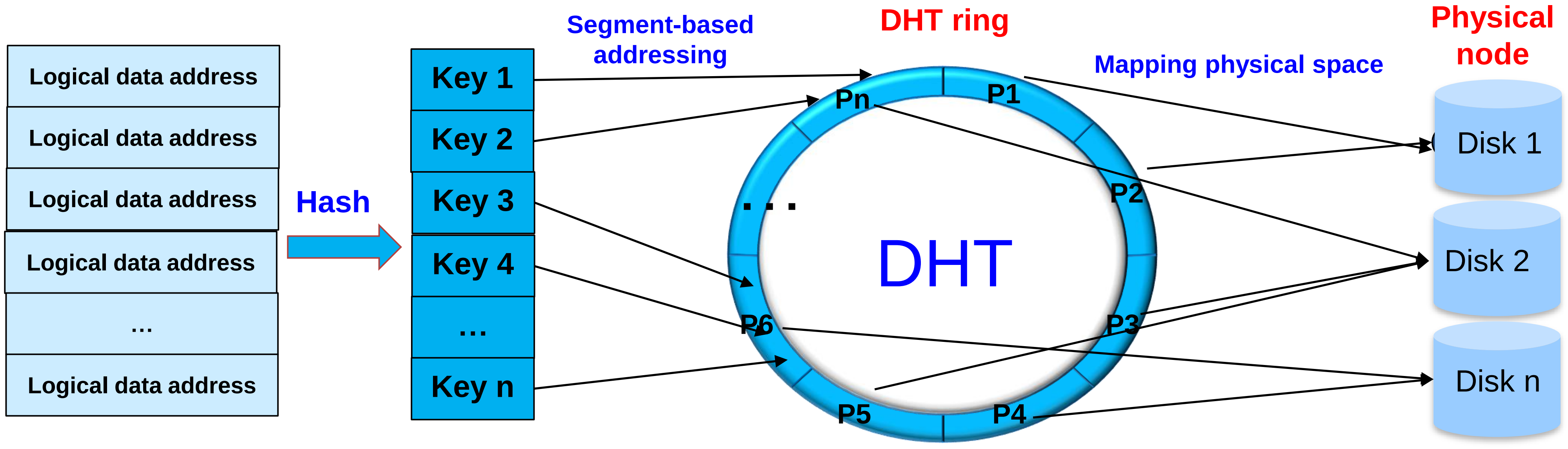
A block storage computing node (block client) and a storage node (server) share the storage and management networks, providing a shortest network access path. Unified management of the client and server is supported.



Various networking types and on-demand deployment

FusionStorage provides various networking types, including GE, 10GE, 25GE RoCE, and 40/56 Gbit/s InfiniBand, meeting diversified performance and cost requirements.

Distributed Hash Addressing, Eliminating Hot Disk Problems and Implementing Large-Scale Expansion



Technical characteristics

DHT ring technology: DHT (Distributed Hash Table) is a ring space formed by 2^{32} super virtual nodes.

Partition: DHT ring is divided into N equal parts. Each part is a partition.

Physical node: One physical node is a disk, which is mapped to a partition.

Benefits

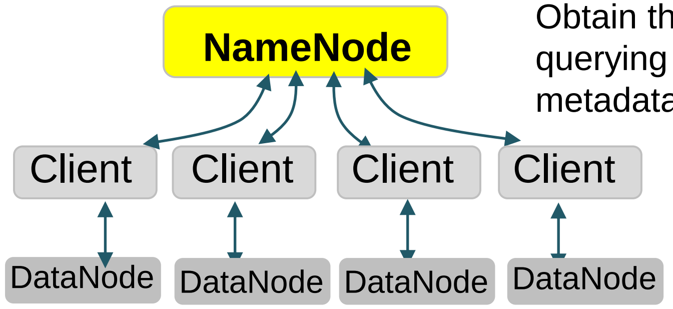
Outstanding performance: Storage data is evenly distributed on all disks through the DHT ring. All disks use data read and write to eliminate the read/write bottleneck caused by hot disks.

High data reliability: The partition allocation algorithm can be flexibly configured to prevent duplicate data from being stored on the same disk, board, or cabinet.

Fast horizontal expansion: When a new physical node is added, only some data (partition) needs to be moved and load balancing is achieved.

I/O Path Shortened and I/O Efficiency Improved by 33%

Industry solution: metadata



Obtain the target by querying the centralized metadata node.

1

Metadata nodes become a performance bottleneck
Limited scalability

Client

1. File mapping object
2. Object mapping PG
3. PG mapping OSD

Server

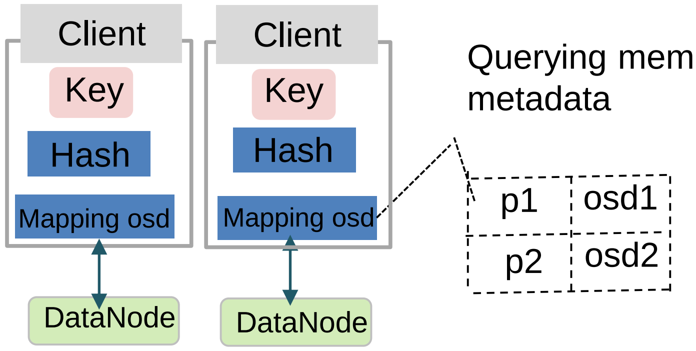
1. Message distribution
2. Write Bluestore
3. Integration of metadata in the background

2

Long I/O path and large background consumption

FusionStorage: distributed Hash

Obtain the target through internal computing.



Querying memory metadata

p1	osd1
p2	osd2

Parallel hash calculation and local memory metadata, no performance bottleneck
Any horizontal expansion

Client

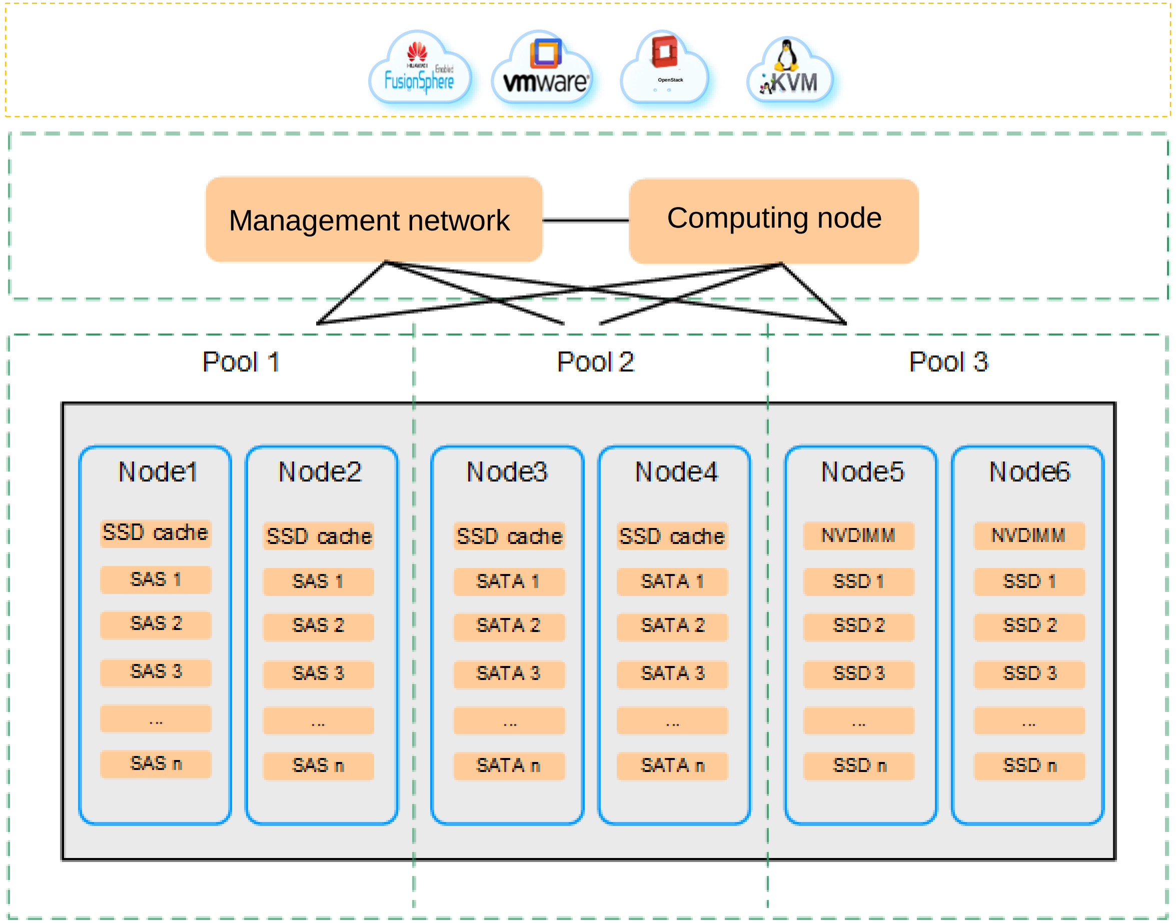
1. Hash calculation partition
2. Partition mapping OSD

Server

1. Position of the mapped disks
2. Write cache

With four E2E steps, the I/O path is shortened without background consumption.

Multi-Pool Architecture, Meeting Resource Requirements of Different Applications



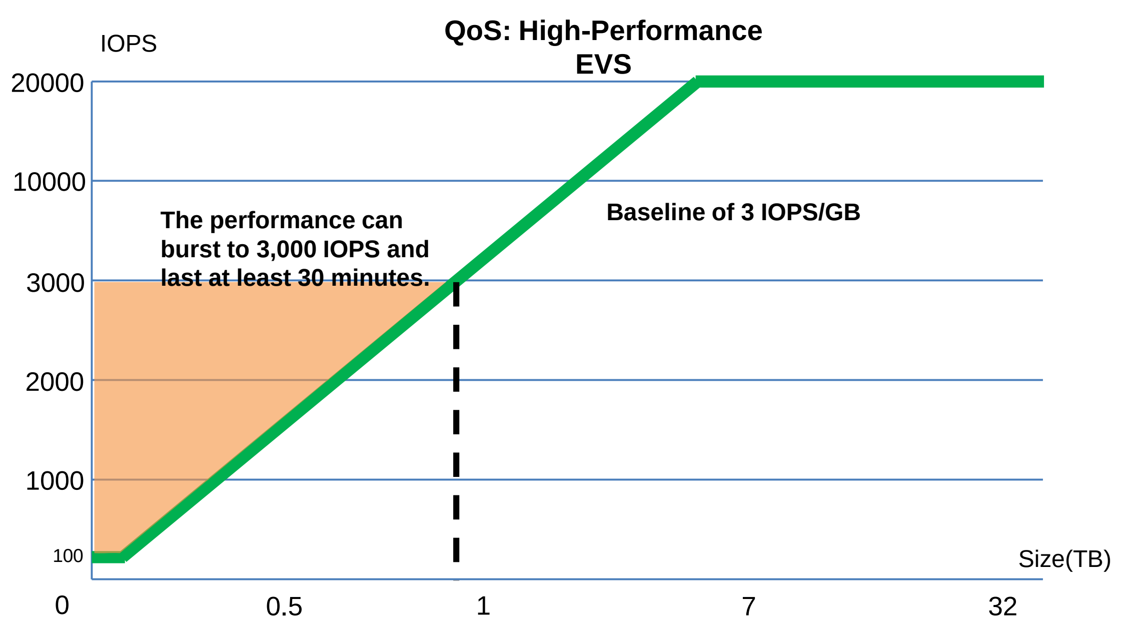
Multiple pools deployed on demand to match applications and resources in an optimal manner

- Customized storage policy: Supports on-demand configuration of main storage, cache, and redundancy policies based on resource pools to meet the performance and cost requirements of different services.
- High reliability: Resources in each resource pool are isolated, and faults do not affect each other.
- Ultimate expansion: Supports a maximum of 128 sub-resource pools and a maximum of 4096 servers in the entire system, meeting future cloud service expansion requirements.

QoS: Supports Intelligent Flow Control and Application Tiering

Voice of Customers

- Customers require that **IOPS and bandwidth resources in storage pools be properly allocated to** applications with different priorities.
- Require enough IOPS and bandwidth resources for core services.
- Require performance burst for a short period of time when the remaining capacity of disks is less than a certain number.



1. The performance starts from 100 IOPS and improves as the capacity increases. The maximum performance can reach 20,000 IOPS.

2. The performance of the volume whose size is smaller than 1 TB can burst to 3,000 IOPS. The duration becomes longer as the volume size increases.

3. The performance of the volume whose size is larger than 1 TB is more than baseline 3,000 IOPS. Therefore, the burst capability is not required.

FusionStorage QoS

- Provides the **refined I/O control** capability which enables the storage system to provide applications of different priorities with differentiated services.
- The **burst** capability enables small-capacity EVSs to obtain high performance within a short period of time, meeting burst and short-time high performance requirements.

Application Scenario

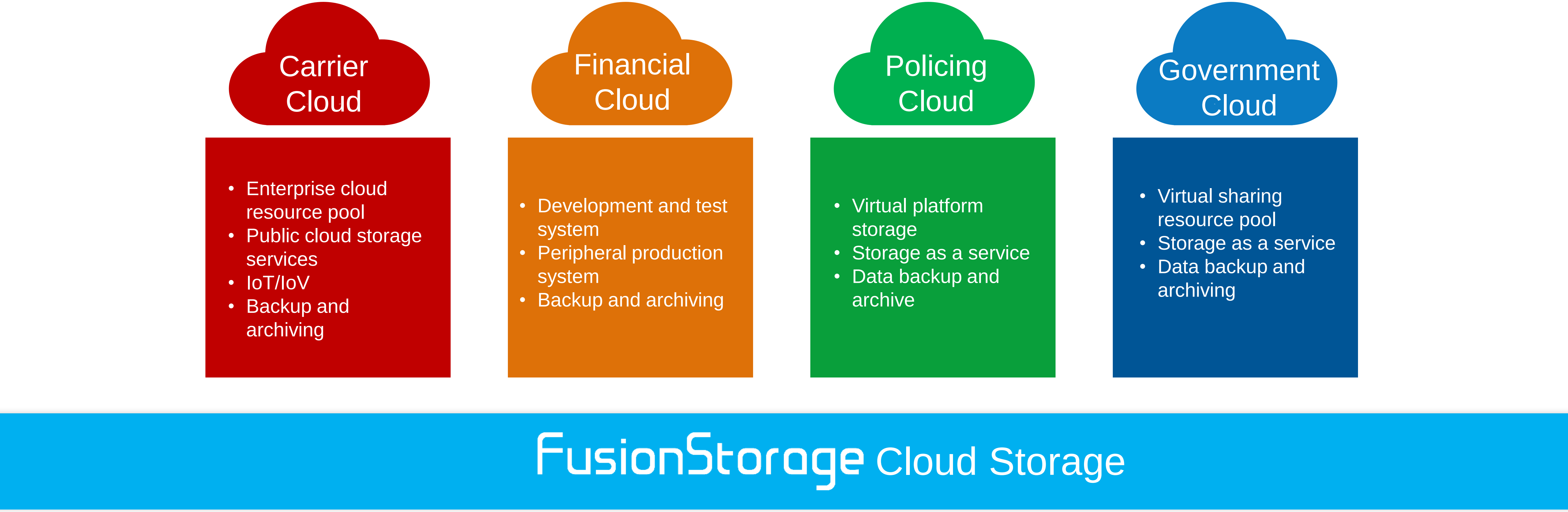
- **Application tiering:** Applications of different priorities can be configured with different types of EVS. QoS controls the allocation of IOPS and bandwidth resources to maximize usage of storage pool resources and avoid core services being affected by other services.
- **VM startup:** The burst capability provides small-capacity system volumes with performance burst for a short period of time, greatly shortening the startup time of VMs.

FusionStorage 6.1

FusionStorage Cloud Storage

Typical Cases

Typical Scenario: Converged Storage of Industry Cloud Resource Pool



Carrier Cloud

- Enterprise cloud resource pool
- Public cloud storage services
- IoT/IoV
- Backup and archiving

Financial Cloud

- Development and test system
- Peripheral production system
- Backup and archiving

Policing Cloud

- Virtual platform storage
- Storage as a service
- Data backup and archive

Government Cloud

- Virtual sharing resource pool
- Storage as a service
- Data backup and archiving

FusionStorage Cloud Storage

Globally Applied FusionStorage Has Rich Experience in Large-Scale Deployment

Finance	Energy	Carrier	Public Sector	Education
China Merchants Bank	Sinopec	China Mobile	Supreme People's Court	Saudi Arabia TVTC
ICBC	CNPC	China Telecom	China Customs	Peking University
Shanghai E-Capital Transfer	China HuaNeng Group	T-Systems	Guangdong MSA	Beijing Jiaotong University
Spanish BME	SAP Labs China	Telefonica	Spanish Xanit hospital	Shanghai Maritime University

Currently, FusionStorage has been used in more than 30 countries/regions and has more than 1000 customers.

FusionStorage Helps China Telecom Zhejiang Branch Build PB-Level Distributed Storage Resource Pools

LEADING NEW ICT

Since we used FusionStorage, we have not expanded the capacity of our traditional FC storage anymore. That is, all services are run on FusionStorage. After the test, the performance of FusionStorage was the most advanced and excellent.

Gu Jiong, Chief Cloud Computing Expert of China Telecom, Zhejiang Branch

Challenges

- Poor scalability and low utilization of storage resources, in contrast to computing and network resources which are now provided on demand
- Two-month long purchasing period of non-standard high-end FC storage and requirement for equipment rooms reconstruction
- Unable to simultaneously achieve linear performance and capacity expansion to tackle rapid data growth

Solutions

- Deployed Huawei FusionStorage+x86 rack server with 2 PB capacity at the first phase. More than 200 application platforms were deployed. The current capacity exceeds 7 PB after multiple times of capacity expansion, and the system has been stably running for more than 3 years.
- Used DHT algorithm to enable I/O parallel processing. The measured lower-layer maximum IOPS reached 1.3 million and lower-layer bandwidth reached 120 Gb/s. Huawei stood as the only manufacturer that passed the customer‘s performance test.
- Supported both VMware vSphere and FusionSphere, avoiding modifications on existing computing resource pools or applications.

Customer Benefits

- The new software-defined storage resource pool is scalable and supports on-demand storage resource allocation.
- Two-week construction and expansion periods shortened by the x86 architecture platform
- Easy to scale out and greatly improved performance. Latency: 1.98 ms to 4 ms (medium-load high-end storage latency: 3 to 6 ms)

HUAWEI

FusionStorage Helps CMB Build a Storage Service Platform for Hundreds of Millions of Mobile Phone Users

FusionStorage helps CMB build a cloud storage service platform. The scalable, on-demand, efficient, and open fully distributed cloud storage resource pool replaces the original siloed architecture, shortening rollout time for new services, speeding up cloud transformation, and providing hundreds of millions of active users with efficient internet operation.

Challenges

- Under the pressure of internet finance, massive transactions bring storage systems of banks great challenges. The traditional siloed storage architecture cannot meet massive high-concurrency and high-IOPS services. At the peak hours of WeChat Lucky Money, the number of transactions is more than 1,000 times as much as that processed by CMB Mobile Banking on November 11 (the Shopping Festival of Taobao) and tens of thousands of times as much as that processed by business halls. The number of transactions has reached the upper limit of the original architecture.
- The traditional siloed architecture fails to keep up with the deployment duration and cost requirements of unpredictable hotspot and massive services.

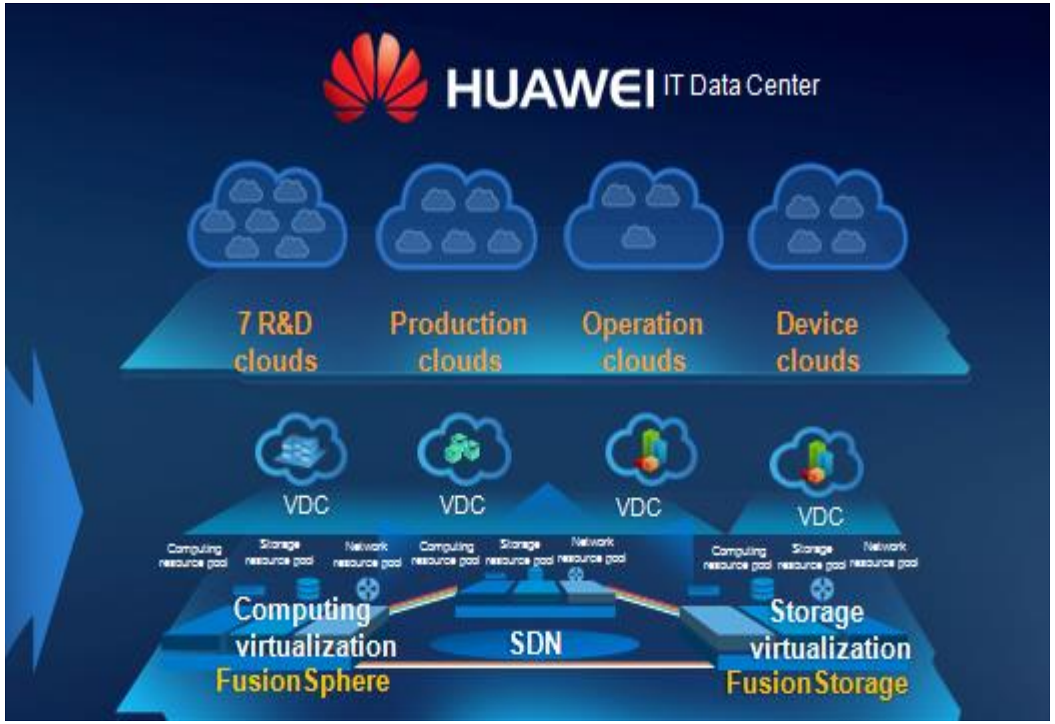
Solutions

- Deployed Huawei FusionStorage+x86 rack servers to provide PB-level block storage service for the development and test system and the channel access system, and provide object storage service for enterprise web disks. Storage resources were provisioned on demand.
- Constructed an easy-to-expand distributed storage resource pool that is compatible with the IaaS platforms from various vendors. VMware vSphere and Huawei FusionSphere without platform binding are supported. The customer has multiple choices.

Customer Benefits

- Fully distributed architecture: delivered 3 PB+ resource pools at the first stage, improved performance by 10 times, and achieved linear expansion of performance and capacity.
- The deployment duration was reduced from several weeks or even months to two weeks. As fully distributed architecture replaced the original siloed architecture, the deployment duration has been greatly reduced, the service provisioning has become more agile, and the customer's cloud transformation has been sped up to deal with massive future services.

FusionStorage Builds the World's Largest Enterprise-Level Cloud Resource Pool



FusionStorage builds the world's largest enterprise-level cloud resource pool – Huawei data center. Four resource pools, including 7 R&D Clouds, production clouds, operation clouds, and device clouds, are built to develop Huawei services.

Challenges

- Applying for the development and test environment takes much time and the performance of the environment is low. Compiling the mobile phone OS takes 68 minutes.
- The number of concurrent operation analysis services is small and TCO is high. Concurrent users that OBIEE supports cannot meet peak-hour requirements.
- Knowledge base of marketing experts requires PB-level capacity. The duration to apply devices for customized demonstration environment requires at least two weeks. The traditional storage cannot meet a large number of personalized requirements.

Solutions

- FusionStorage builds the world's largest enterprise-level cloud resource pool whose block storage capacity exceeds 800 PB, and supports more than one million VMs and dozens of database systems. The system scale increases 10 times per year on average.
- Builds more than 600 PB distributed file and object storage and the terminal service cloud that supports more than 130 million people's access.

Customer Benefits

- The compilation time for Honor 7 OS is shortened from 1 hour to 30 minutes. The compilation time for unified storage is reduced from 150 minutes to 19.4 minutes.
- The Oracle database performance is improved by 3 to 9 times. The report generation time is reduced from 8 hours to 5 hours. The core system has been running properly for over 680 days.
- Services are provisioned in minutes and the environment preparation time is shortened from 12 hours to 20 minutes, improving efficiency by 36 times. 2500 VMs can be started in 5 minutes.

Mexico KIO: FusionStorage Ranks First in 17 Service Overload Performance Tests

KIO network is the largest ISP in Mexico, providing users' key applications with high-availability IT services, including SAP, SAP HANA, disaster recovery, and service continuity. Recently, KIO completed the performance tests on EMC ScaleIO, Ceph Storage, and Huawei FusionStorage. They found that the performance of a single SSD of FusionStorage is 100 times as much as that of Ceph.

Configuration of a storage server node: Huawei RH2288H V3

- CPU: 2 x E5-2680 V3@2.4 GHz, 8 cores
- Memory: 16 x 16 GB
- Disk: 2 x 600 GB SAS
- SSD: 2 x 3.2 TB PCIE SSD card
- NIC: 2 x 10GE, 2 x GE

FusionStorage VS Ceph/ScaleIO

❑ FusionStorage ranks first in 17 service overload performance tests.

❑ Average IOPS of FusionStorage is 1.4 times that of ScaleIO, 15 times that of Ceph while the latency is 9% of Ceph's latency.

❑ FusionStorage integrates into OpenStack cloud platform.

Discussions

Q&A

THANK YOU

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.