

Storage Technology Trends



Foreword

- Data carries information during the transmission on networks. What is the relationship between information and data? What is the function of data storage? This course describes the definitions of information and data in the computer field, their relationship, as well as the concept, development history, and development trend of data storage.

Objectives

On completion of this course, you will be able to understand:

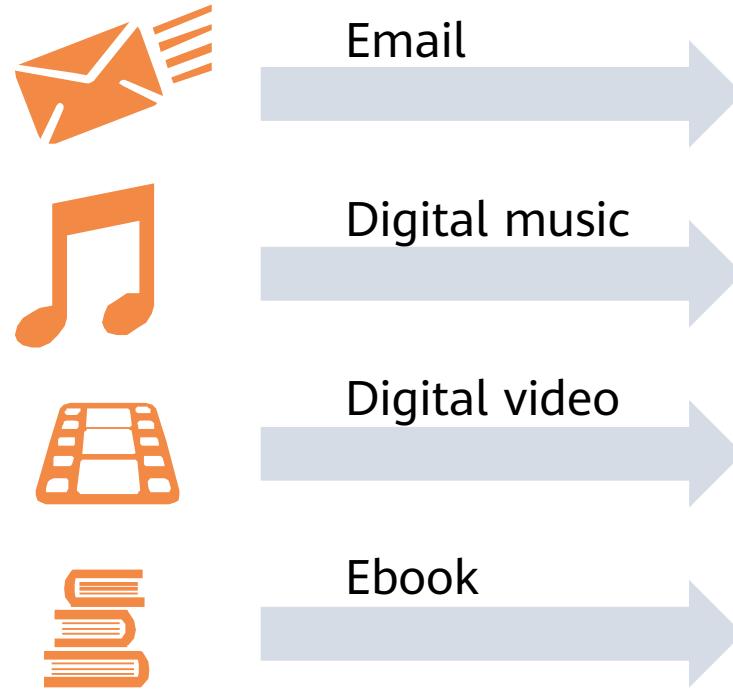
- Definitions of information and data
- Concept of data storage
- Development history of data storage
- Development trend of data storage products

Contents

- 1. Data and Information**
2. Data Storage
3. Development of Storage Technologies
4. Development Trend of Storage Products

What is Data

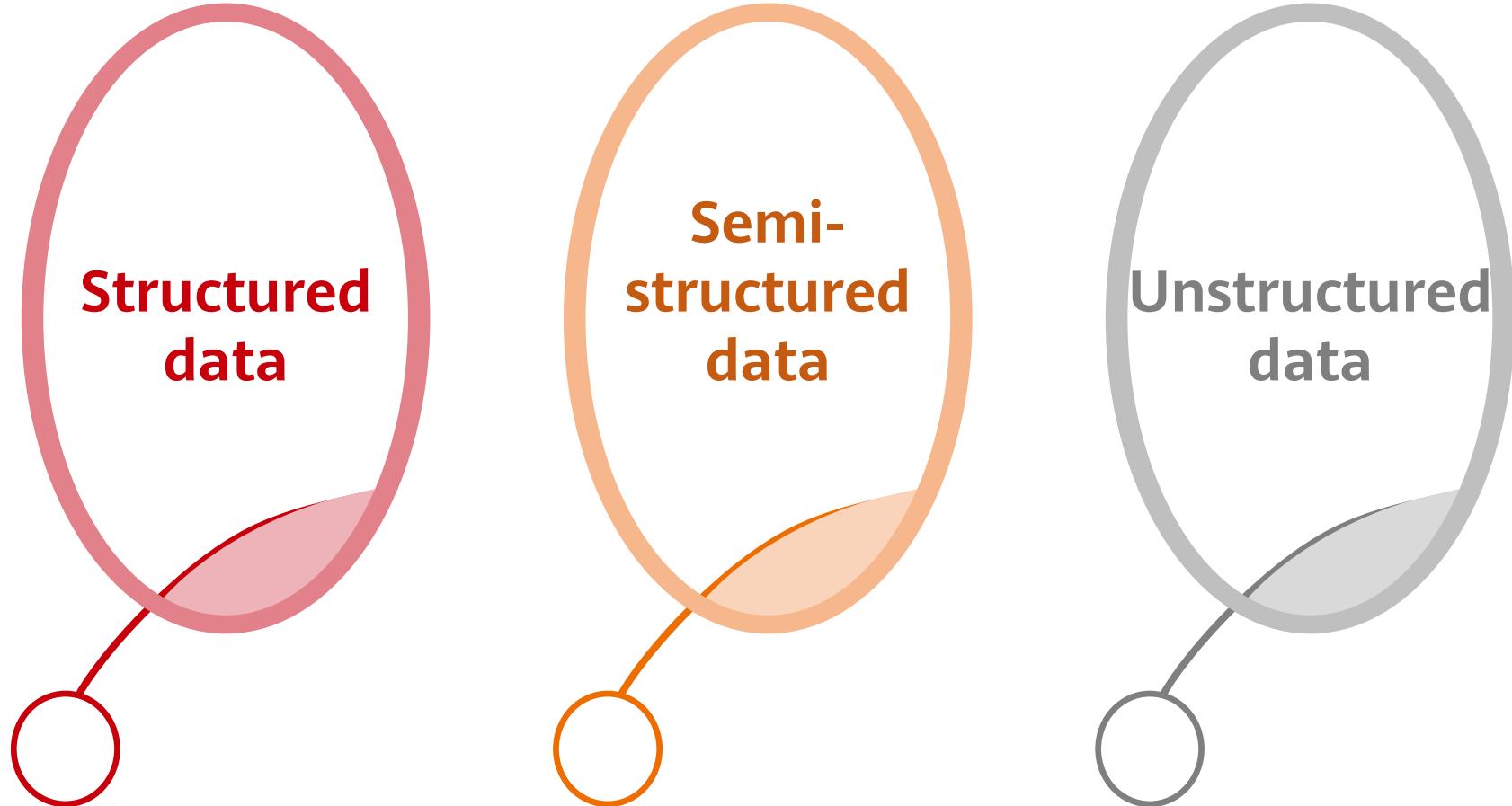
- SNIA (Storage Networking Industry Association) defines data as the digital representation of anything in any form.



Format in which data is stored

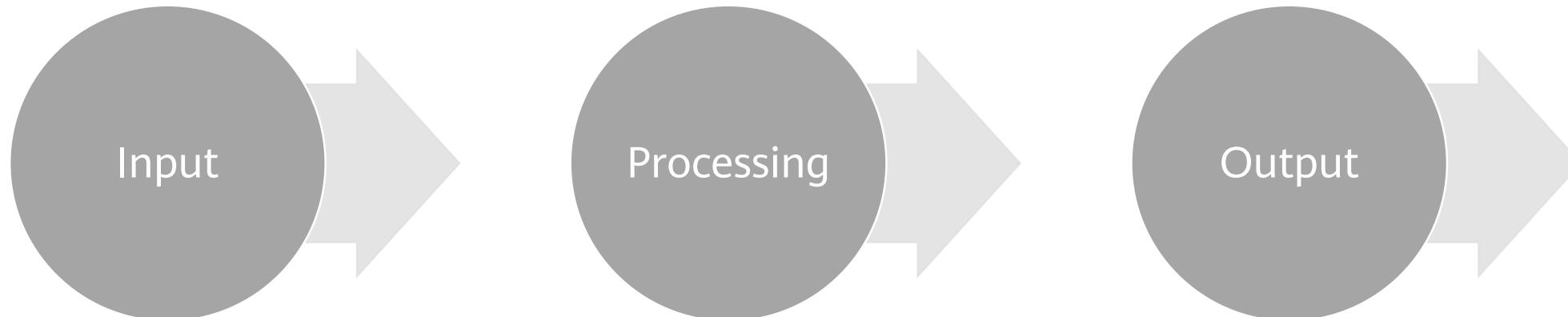
```
0101001010100010000011  
1100011100010001110001  
1100000111101010100101  
0101001010100101001010  
1001010101001010100010  
1010010101001010101010  
01010101010010100010  
0101001010101010100101  
0101001010101010100101  
01010100100101001000  
1010101001001010010010
```

Data Types



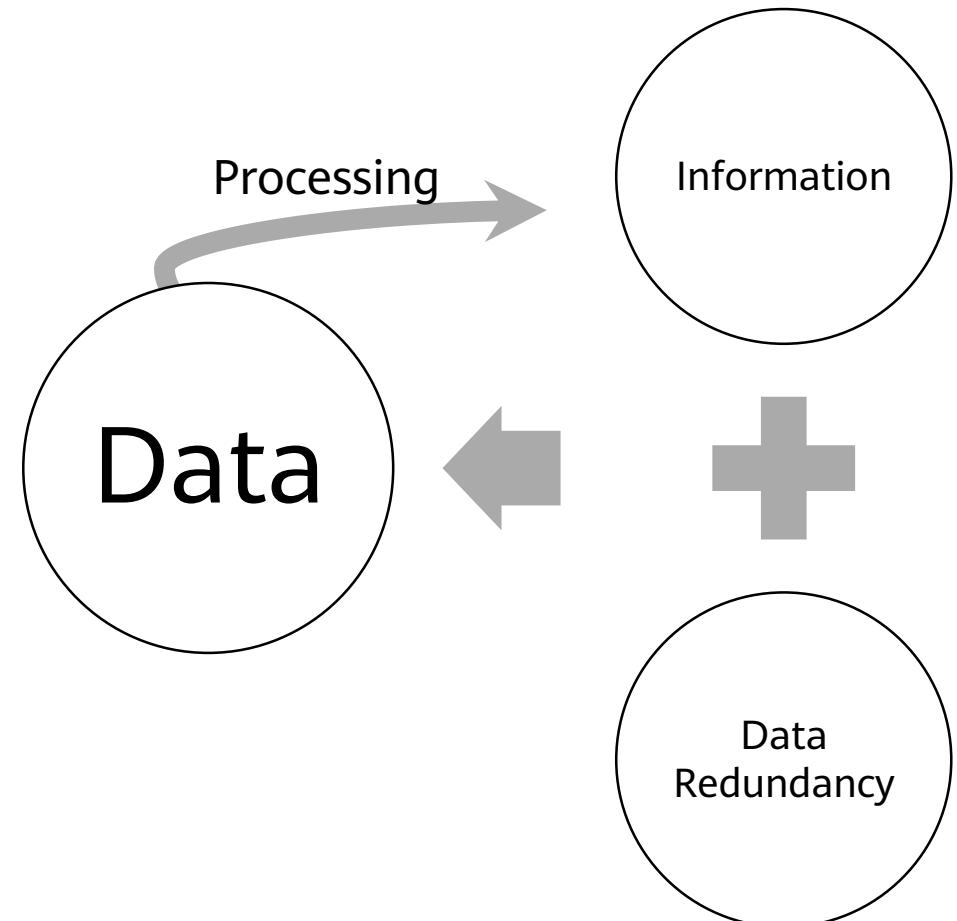
Data Processing Cycle

- Data processing is the reorganization or reordering of data by humans or machines to increase their specific value. A data processing cycle includes three basic steps: input, processing, and output.



What is Information

- Information is processed, structured, or rendered in a given context to make it meaningful and useful.
- Information is processed data, including data with context, relevance, and purpose. It also involves the manipulation of raw data.



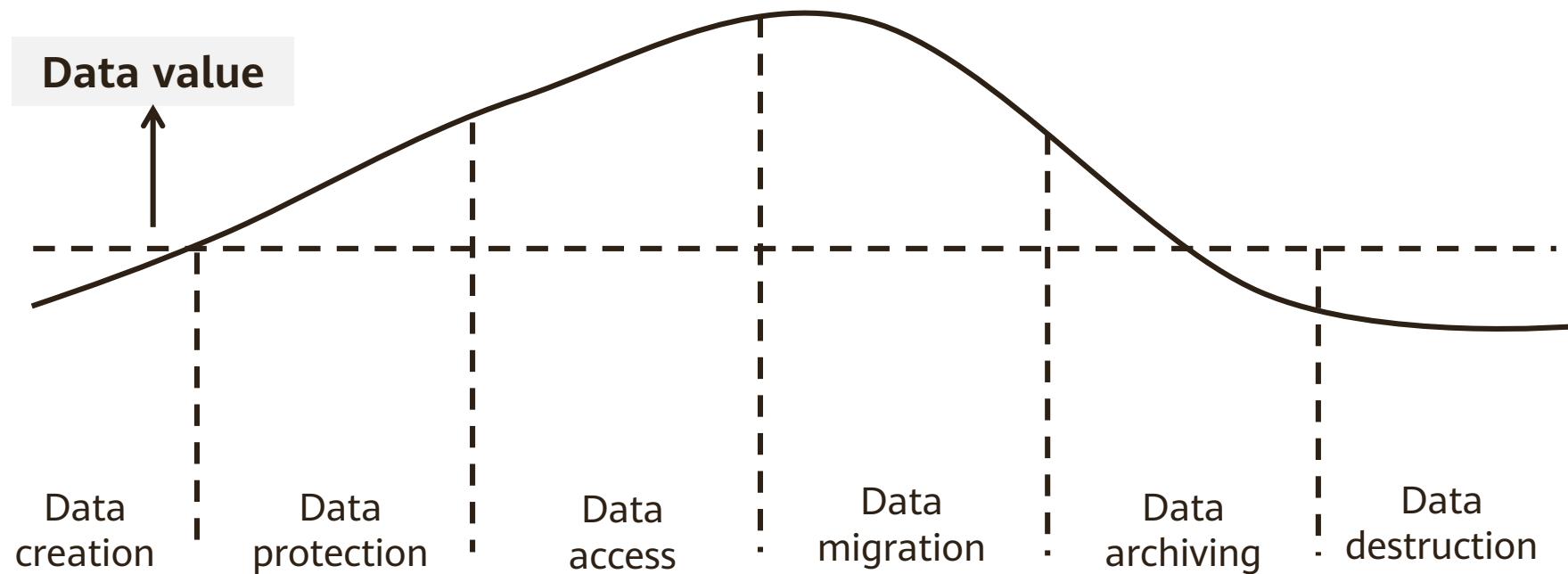
Data vs. Information

- After being processed, data can be converted into information.
- In order to be stored and transmitted in IT systems, information needs to be processed as data.

Item	Data	Information
Feature	Raw and meaningless, with no specific purpose	Valuable and logical
Essence	Original materials	Processed data
Dependence	Data never depends on information	Information depends on data
Example	Meteorological data or satellite image data	Weather forecasts

Information Lifecycle Management

- Information lifecycle management (ILM) refers to a set of management theories and methods from the stage in which the information is generated and initially stored to the stage where the information is obsoletely deleted.



Contents

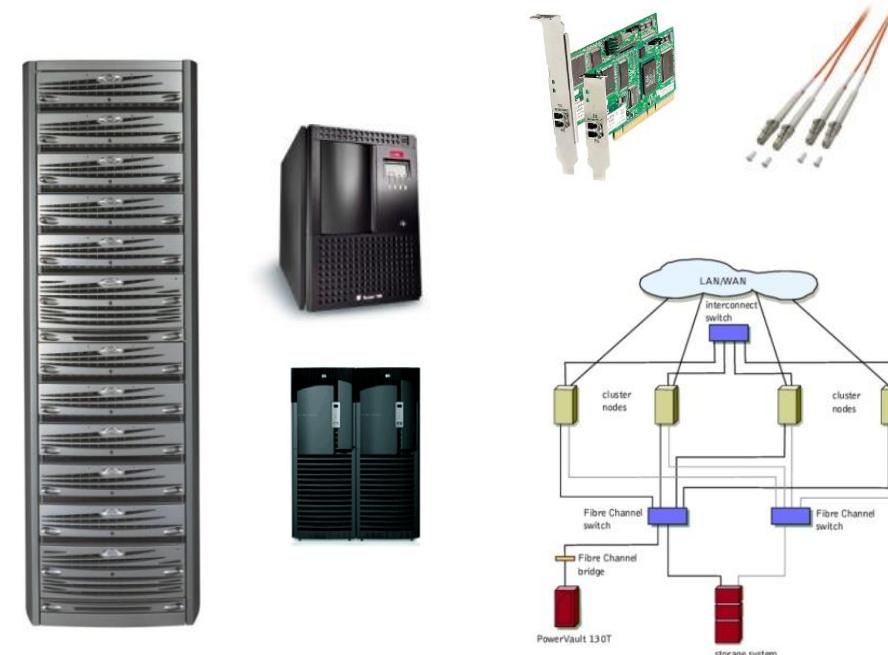
1. Data and Information
- 2. Data Storage**
3. Development of Storage Technologies
4. Development Trend of Storage Products

What is Data Storage

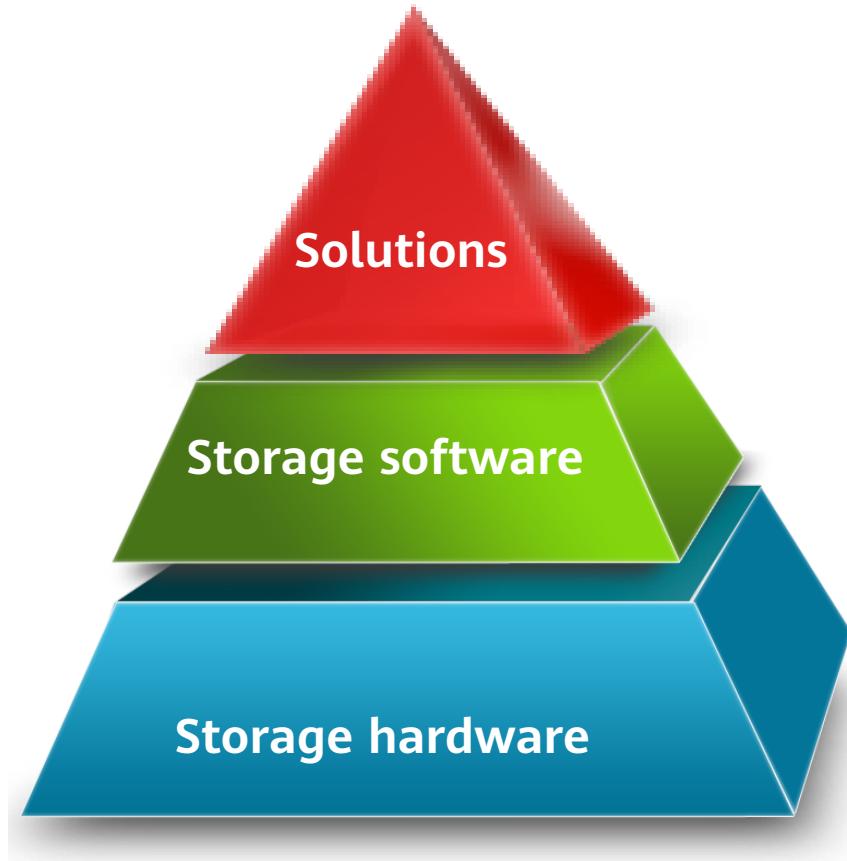
Storage in a narrow sense



Storage in a broad sense



Data Storage System

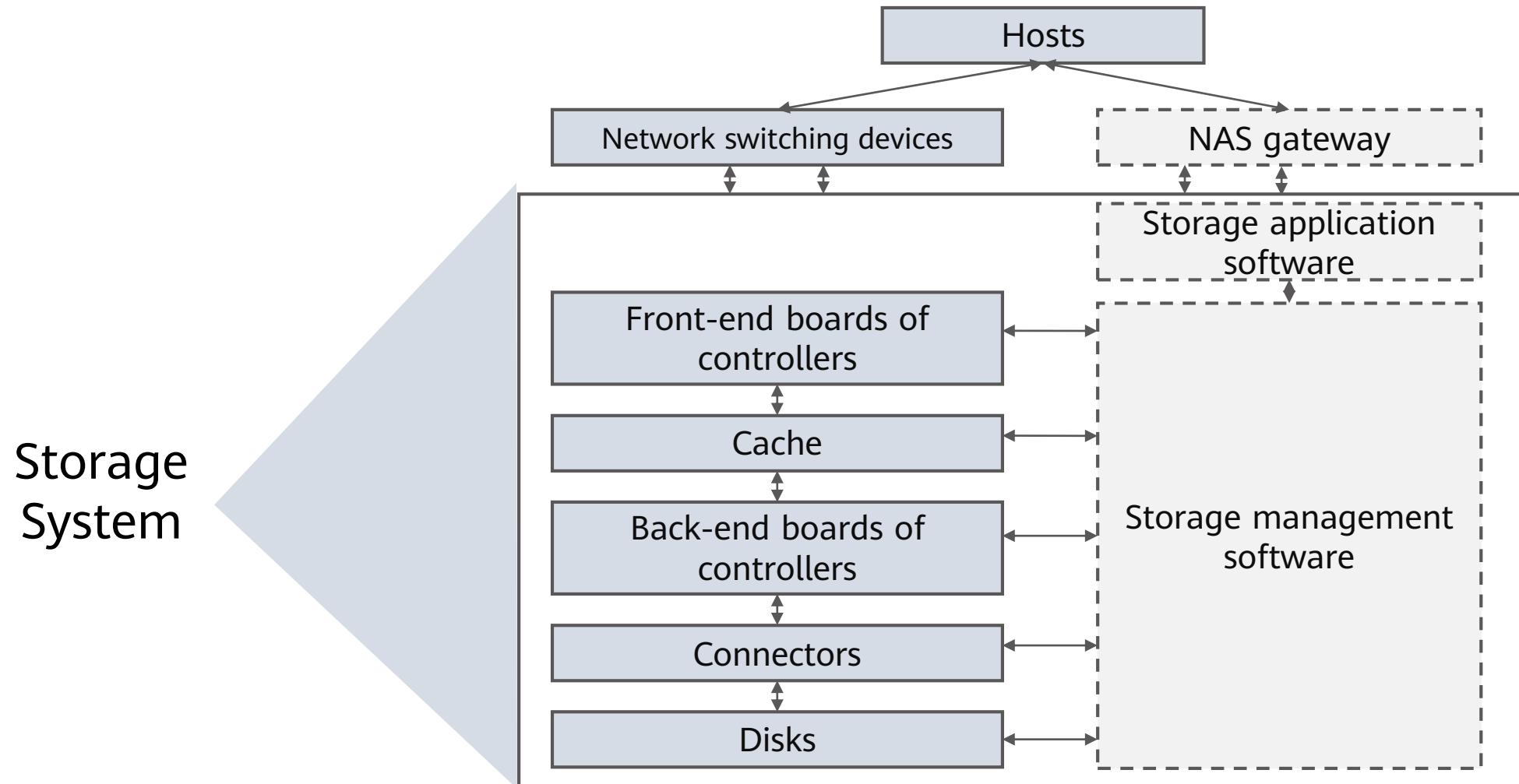


-
- Disaster recovery (DR) solutions
 - Backup solutions
-

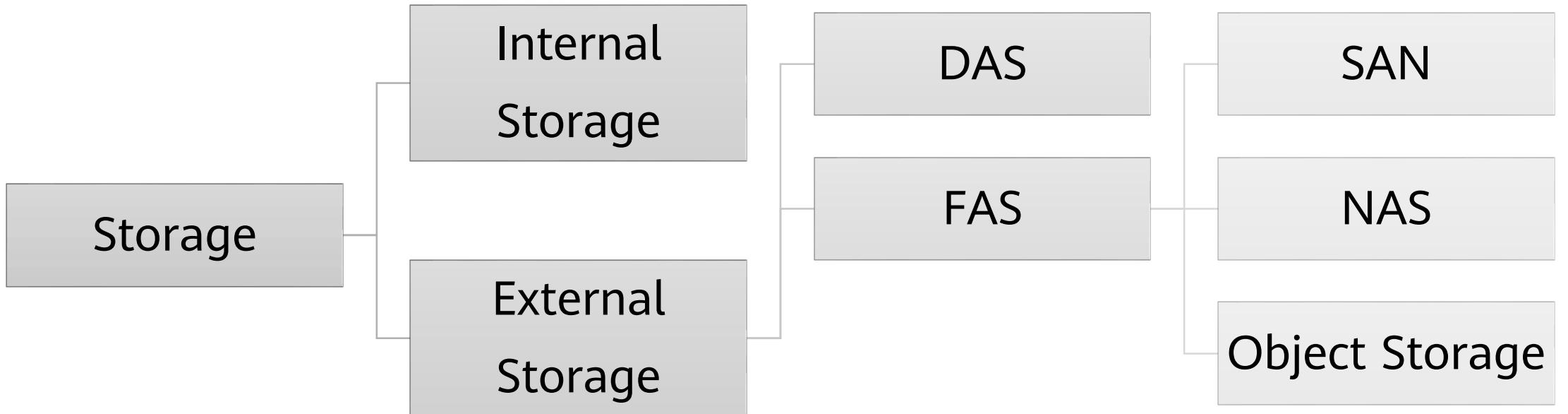
- Storage management software
 - Snapshot and mirroring software
 - Backup software
 - Multipathing software
-

- **Storage devices**
 - Disk array
 - Tape library
 - Virtual tape library
 - ...
- **Connection elements**
 - HBA cards
 - Switches
 - Cables
 - ...

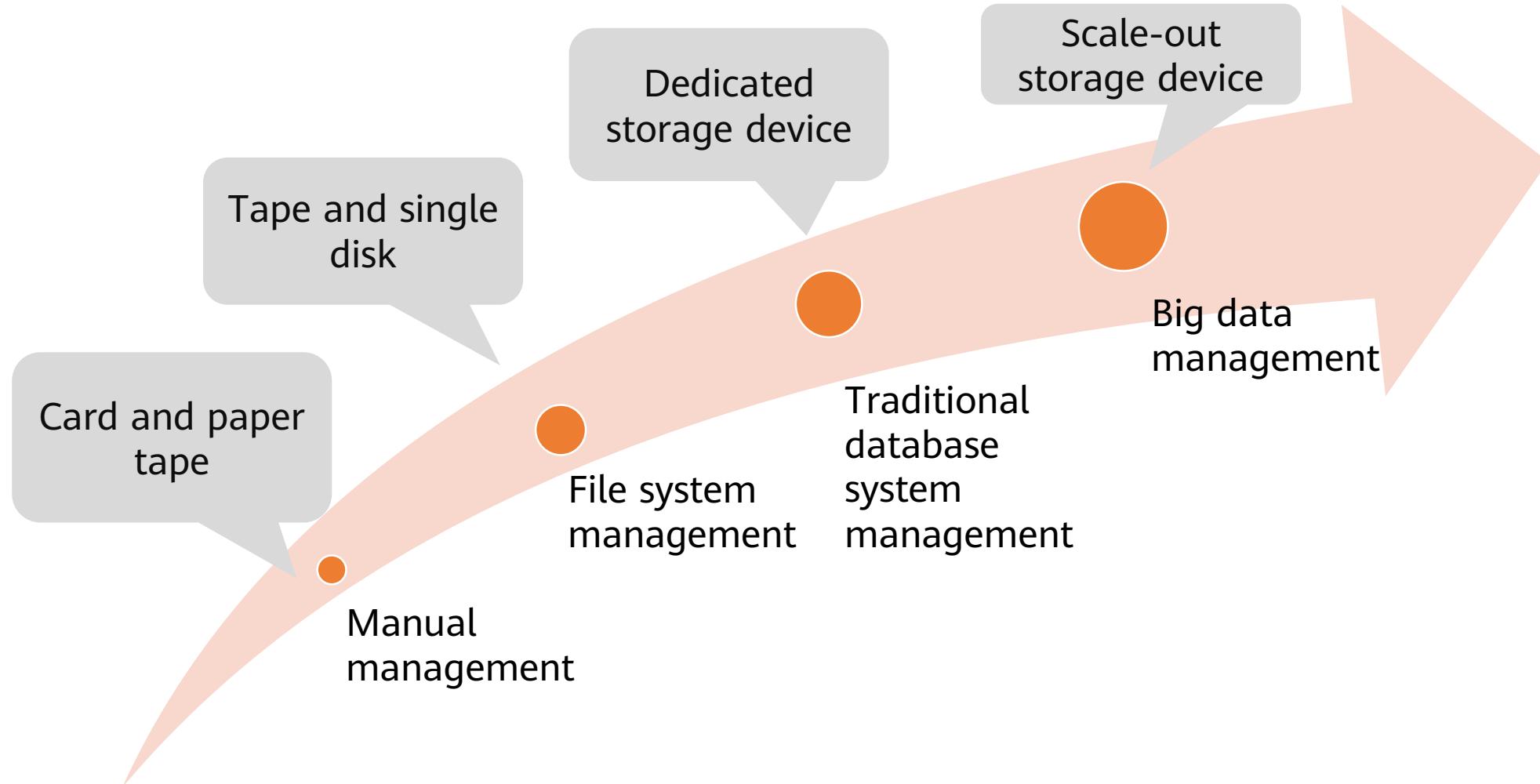
Physical Structure of Storage



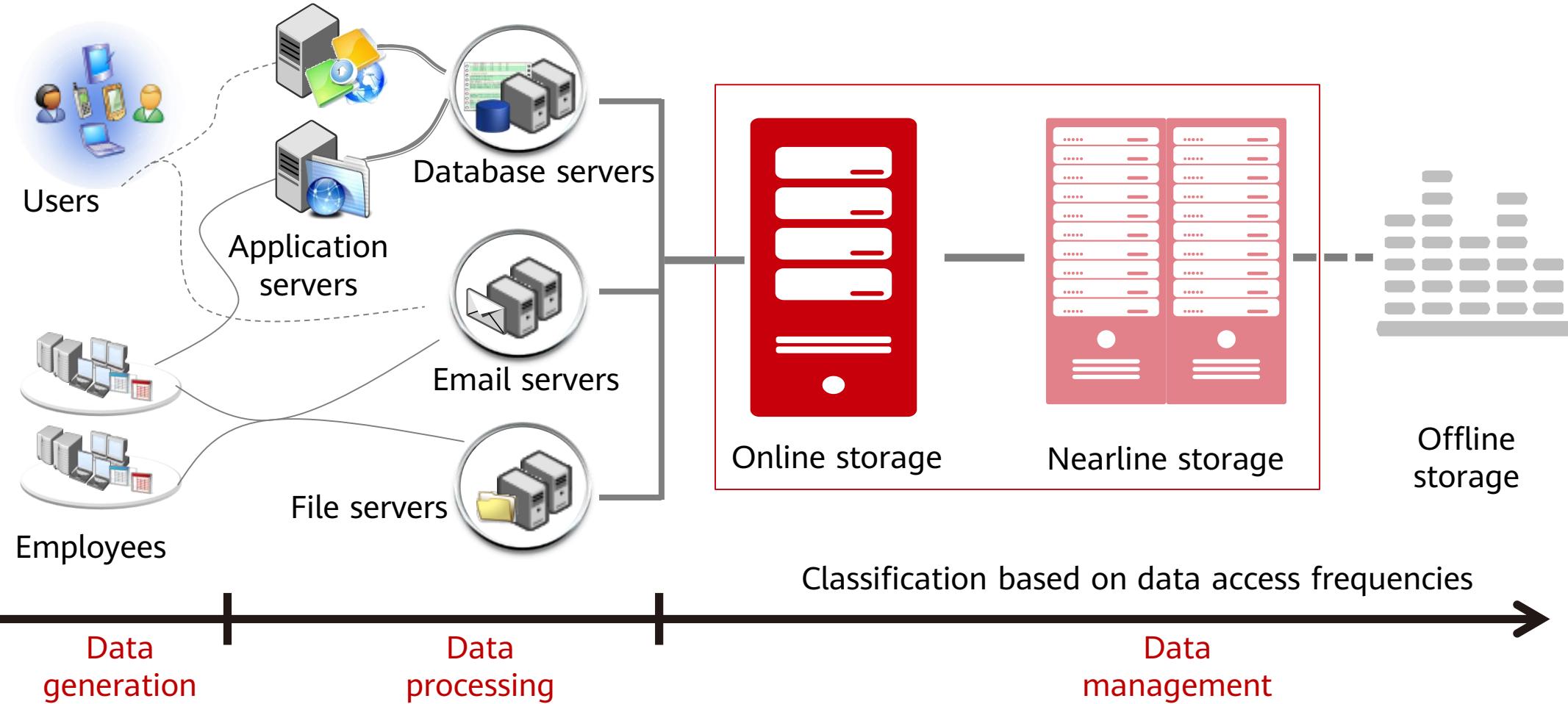
Data Storage Types



Evolution of Data Management Technologies



Data Storage Application



Contents

1. Data and Information

2. Data Storage

3. Development of Storage Technologies

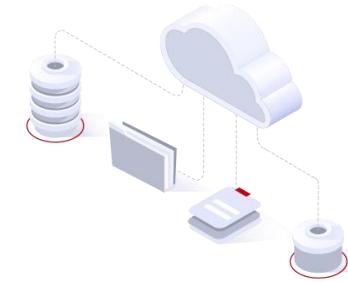
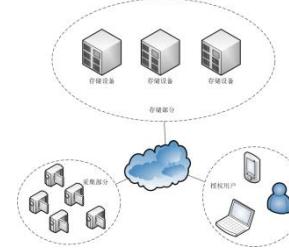
- Storage Architecture

- Storage Media

- Interface Protocols

4. Development Trend of Storage Products

History of Storage Architecture Development



1950s

- Traditional storage

1980s

- External storage

1990s

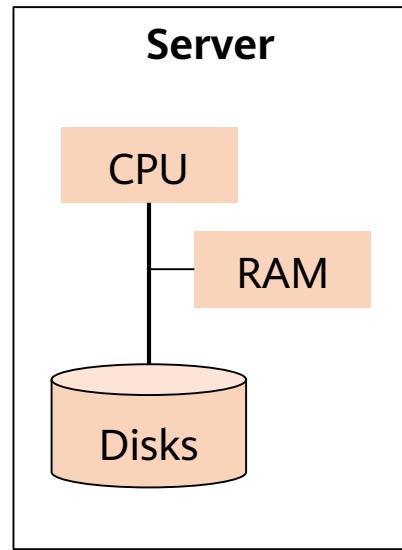
- Storage network

2000s

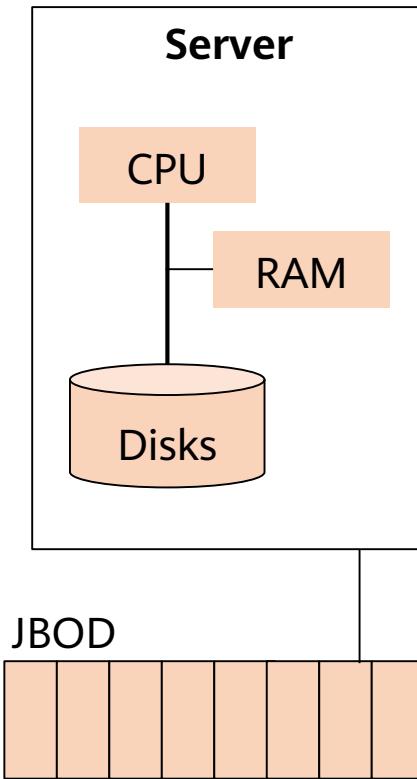
- Scale-out storage
- Cloud storage

From Disks to Disk Arrays

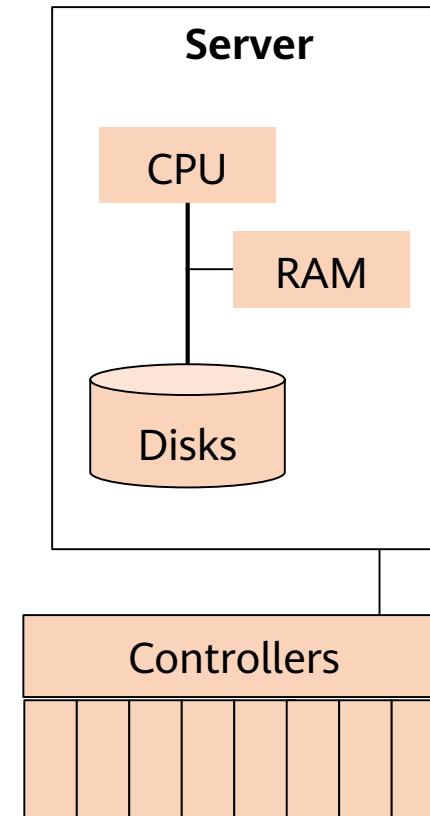
Disks in a server



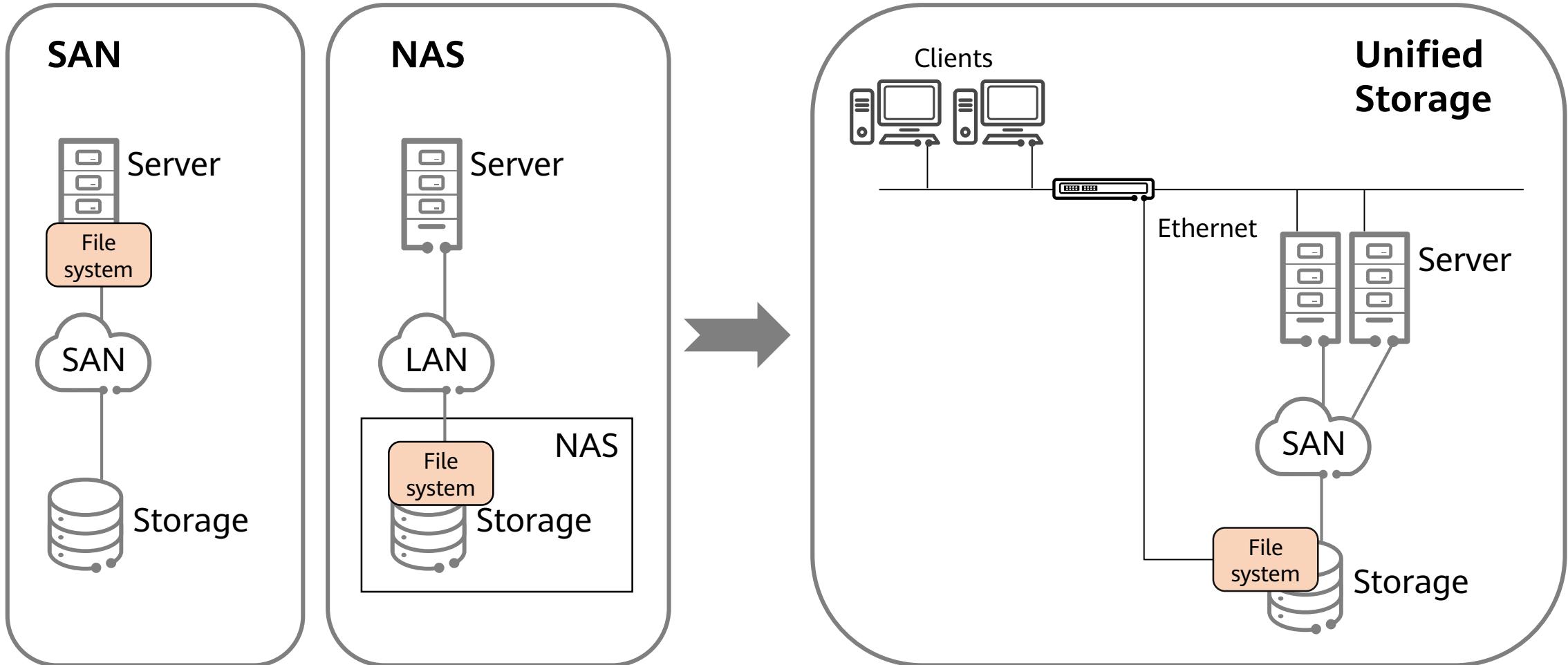
Early external storage



Storage arrays

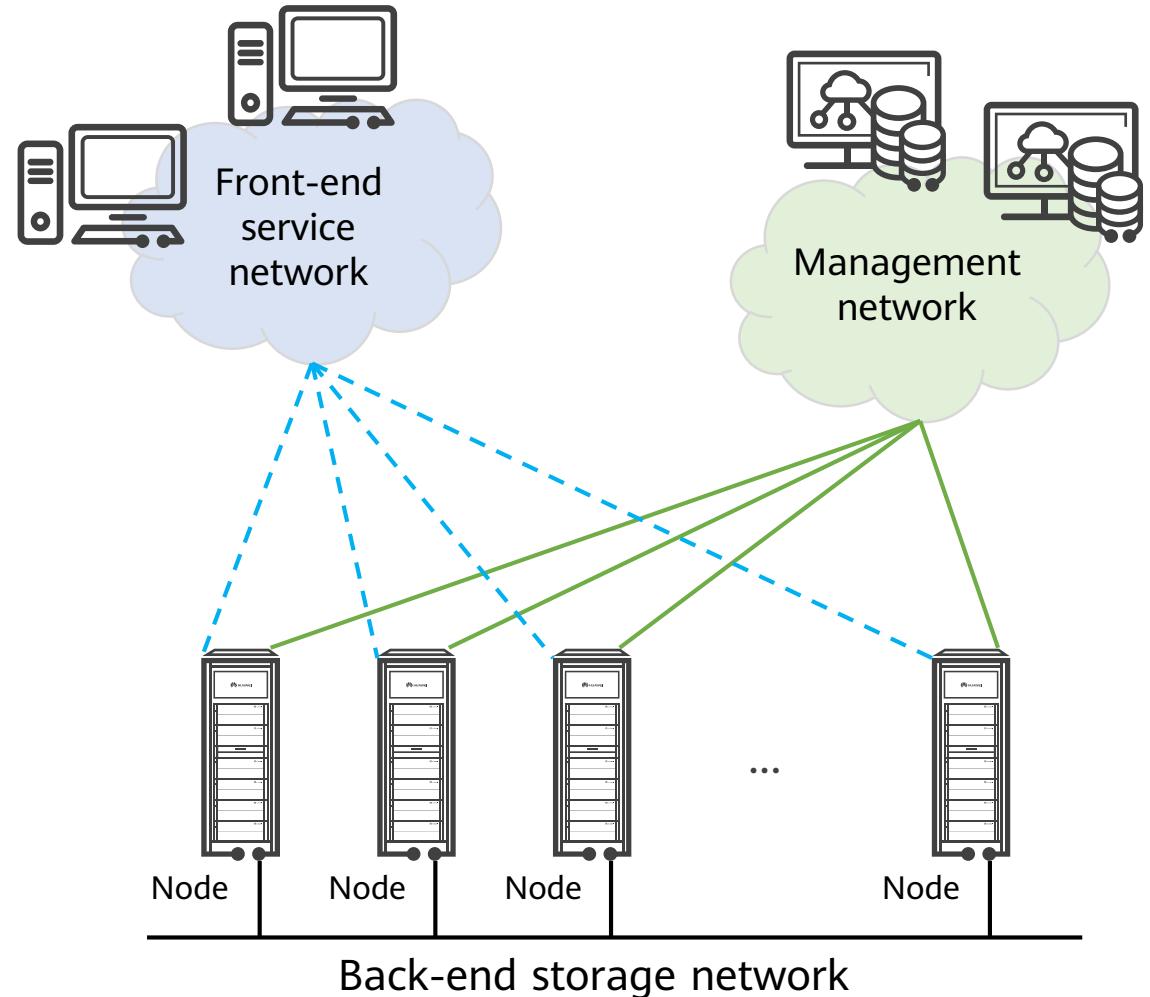


From Separation to Convergence



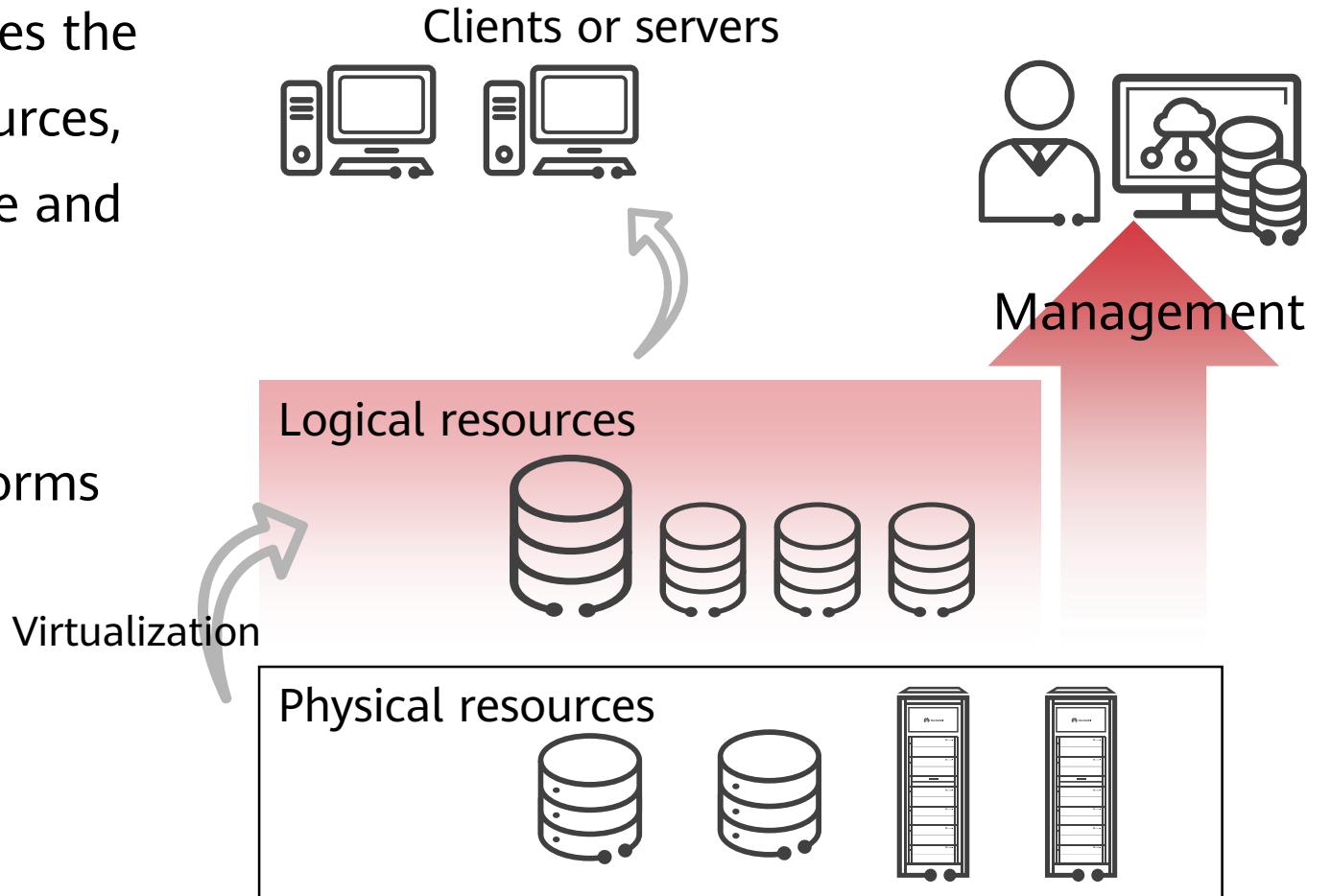
Scale-out Storage

- Physical resources are organized using software to form a high-performance logical storage pool, ensuring reliability and providing multiple storage services.
- Generally, scale-out storage scatters data to multiple independent storage servers in a scalable system structure. It uses those storage servers to share storage loads and uses location servers to locate storage information.



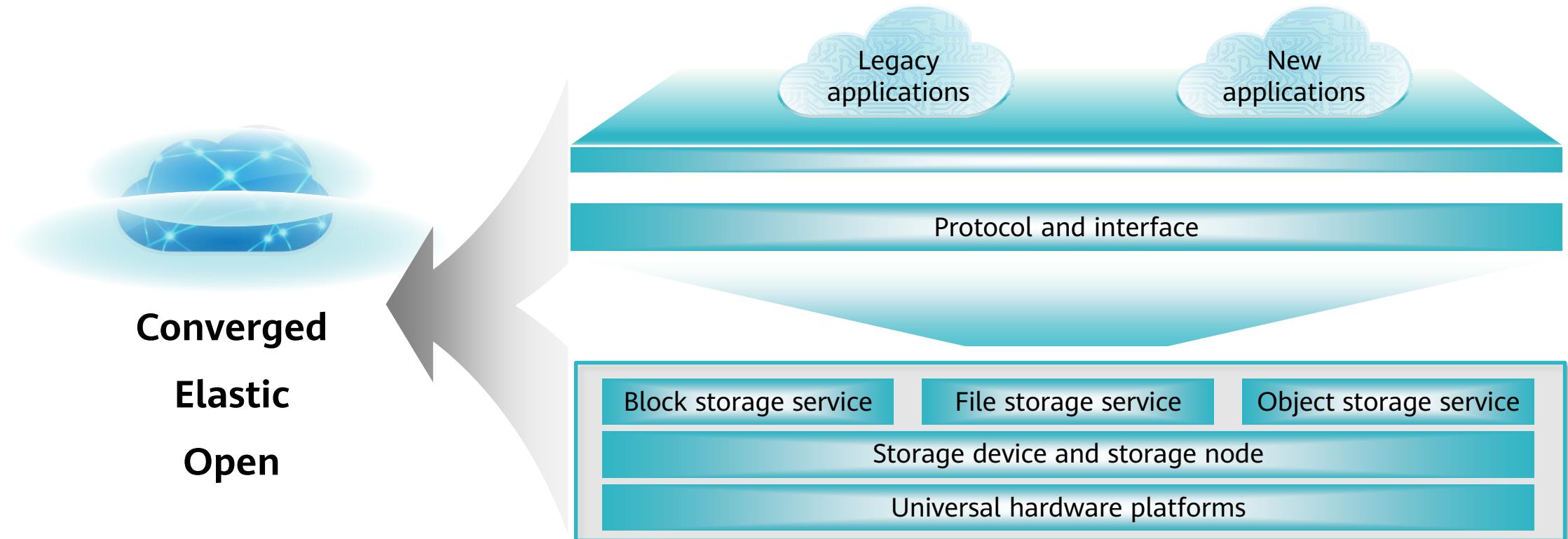
Storage Virtualization

- Storage virtualization consolidates the storage devices into logical resources, thereby providing comprehensive and unified storage services.
- Unified functions are provided regardless of different storage forms and device types.



Cloud Storage

- The cloud storage system combines multiple storage devices, applications, and services. It uses highly virtualized multi-tenant infrastructure to provide scalable storage resources for enterprises. Those storage resources can be dynamically configured based on organization requirements.



Contents

1. Data and Information

2. Data Storage

3. Development of Storage Technologies

- Storage Architecture

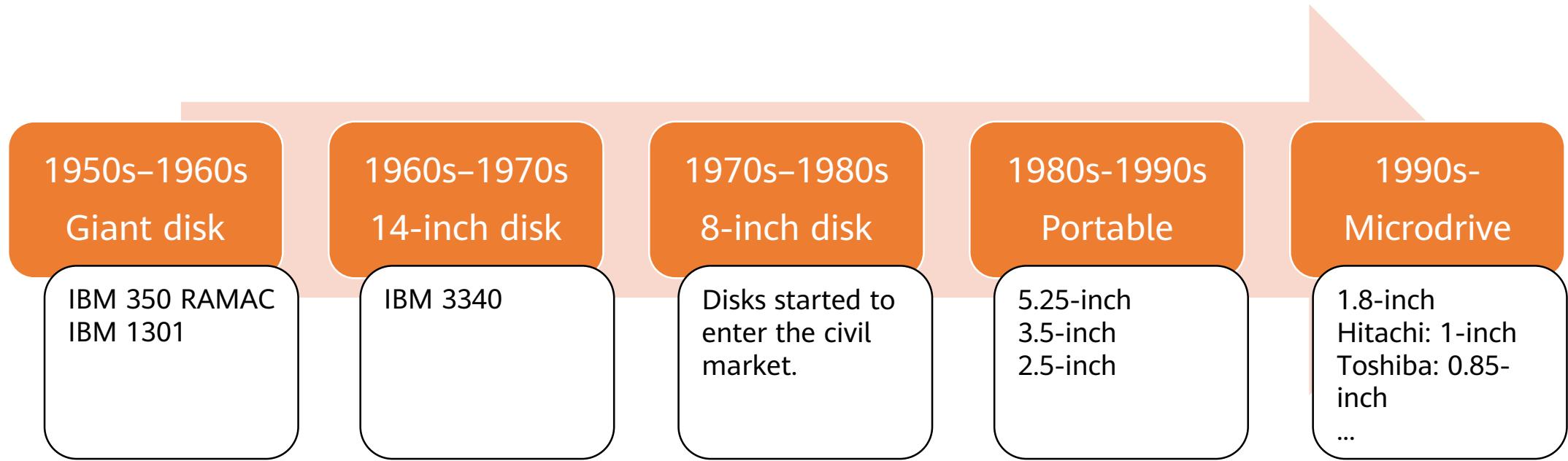
- Storage Media

- Interface Protocols

4. Development Trend of Storage Products

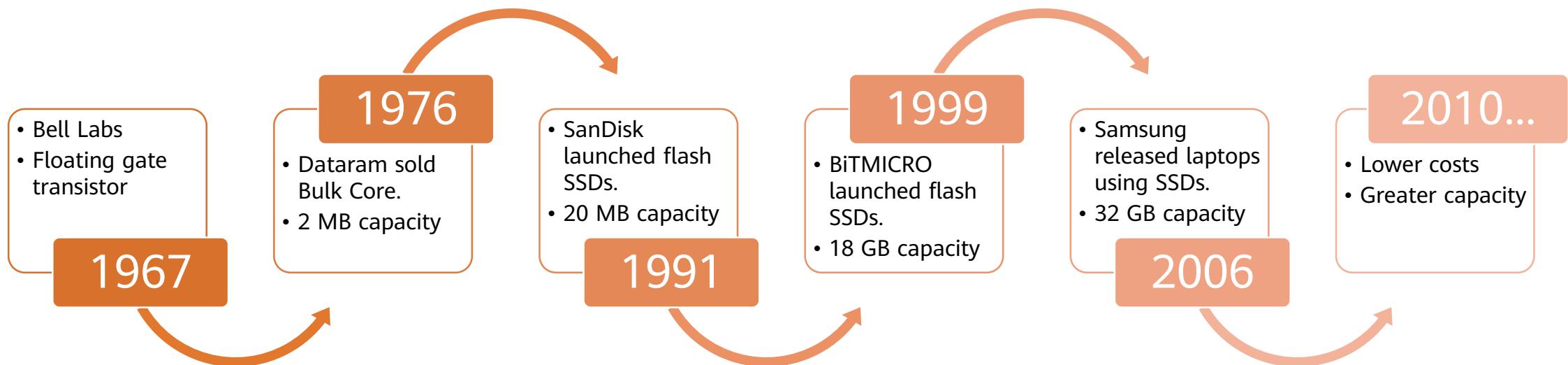
History of HDDs

- Larger capacity with the smaller size.



History of SSDs

- Solid-state drives (SSDs) were invented almost as early as HDDs, but were not popular at that time due to its high price and the rapid development of HDDs at the end of the 20th century.
- With the requirement for high access speed, SSDs are booming.

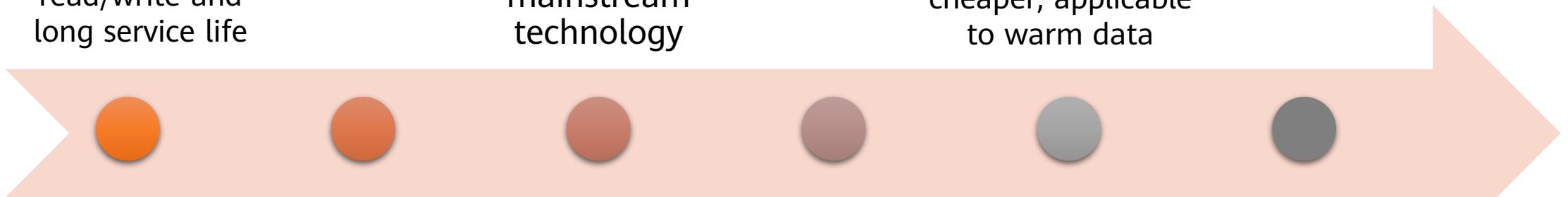


Development of Flash Memory

Single-level cell
(SLC): fast
read/write and
long service life

Triple-level cell
(TLC): mature
mainstream
technology

Quad-level cell
(QLC): larger and
cheaper, applicable
to warm data



Multi-level cell
(MLC): moderate
read/write speed
and service life

3D TLC: improves
the storage density
through multi-
layer overlaying

SCM: with the access
speed slightly slower than
memory, but much faster
than NAND media

Storage Class Memory (SCM)

- Storage class memory (SCM) is non-volatile memory, which is slightly slower than memory but much faster than NAND in terms of the access speed.
- There are various types of SCM media under development, but the mainstream SCM media are PCRAM, ReRAM, MRAM, and NRAM.

	Current Storage Technology		SCM			
	DRAM	NAND Flash	PCRAM	ReRAM	MRAM	NRAM
Non-volatility	No	Yes	Yes	Yes	Yes	Yes
Read latency	10–60 ns	25 μ s	48 ns	< 10 ns	< 10 ns	< 30 ns
Write latency	10–60 ns	200 μ s	40–150 ns	~ 10 ns	12.5 ns	50 ns
Erasable times	$> 10^{15}$	10^4	10^8	10^5	$> 10^{15}$	$> 10^{14}$
Addressing unit	Byte	Page	Byte	Byte	Byte	Byte

Contents

1. Data and Information

2. Data Storage

3. Development of Storage Technologies

- Storage Architecture

- Storage Media

- Interface Protocols

4. Development Trend of Storage Products

Interface Protocols

- Disk interfaces connect disks to hosts.
- Interface protocols refer to the communication modes and requirements that interfaces for exchanging information must comply with.

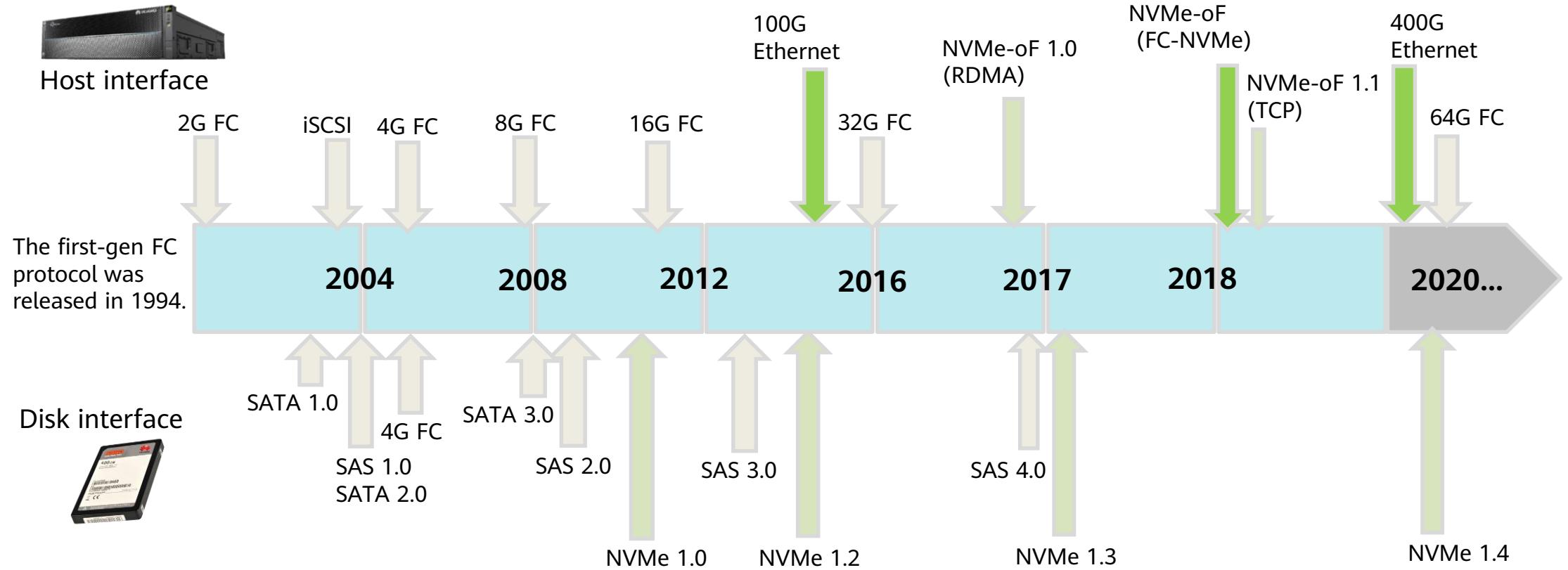


Parallel interface



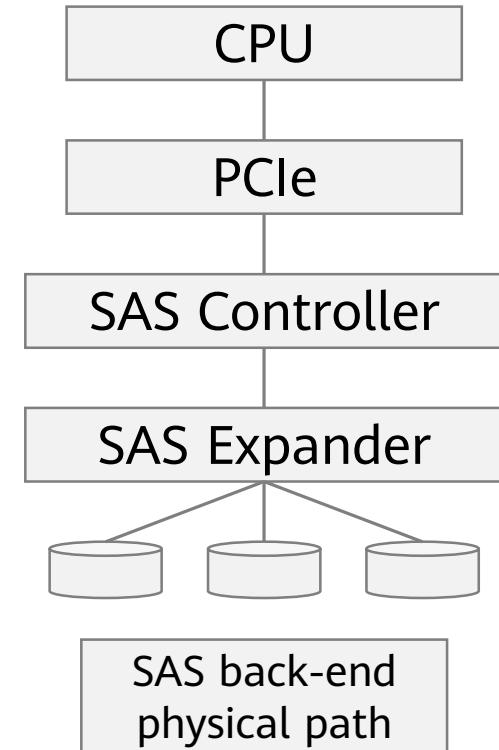
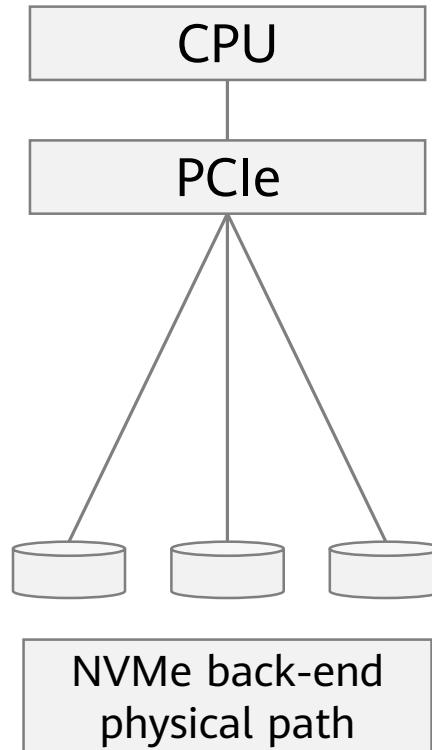
Serial interface

History of Interface Protocols



NVMe and NVMe-oF

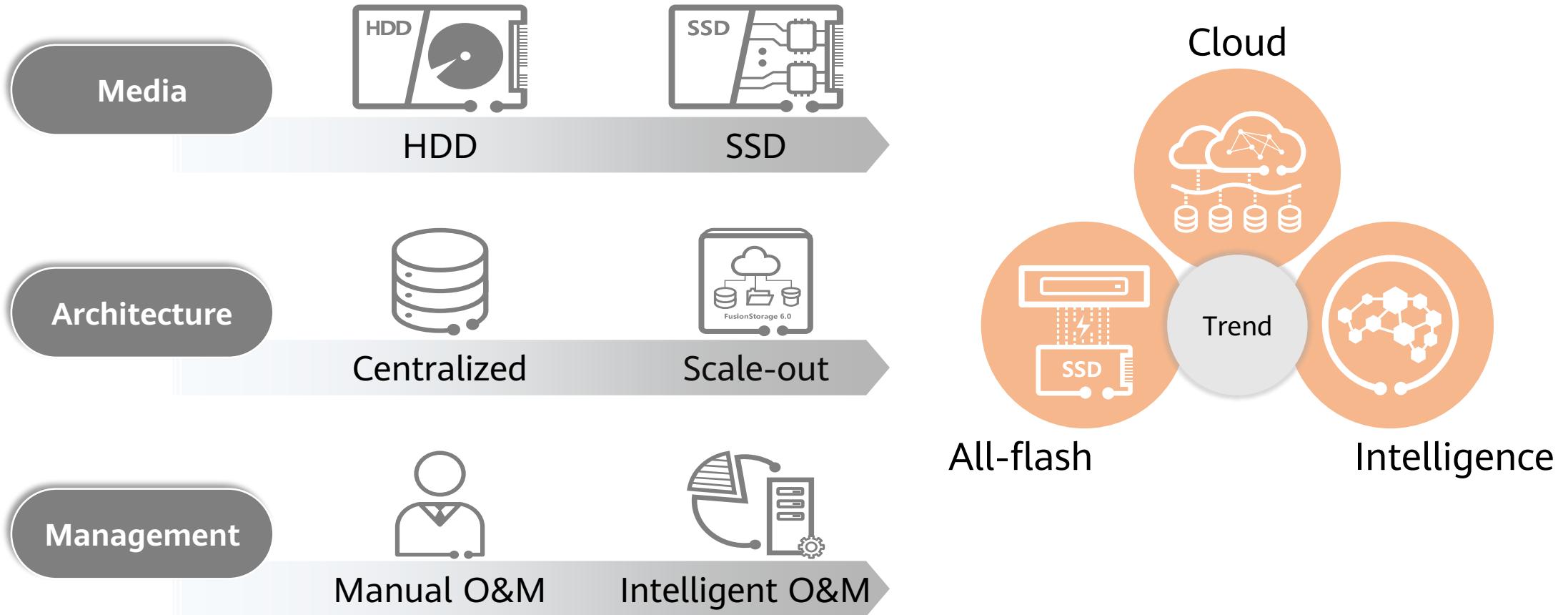
- NVMe, Non-Volatile Memory Express
 - Improves the performance
 - Reduces the latency
- NVMe-oF, NVMe over Fabrics
 - Potential: low latency and high bandwidth
 - Purpose: accelerates the data transmission among the storage network



Contents

1. Data and Information
2. Data Storage
3. Development of Storage Technologies
- 4. Development Trend of Storage Products**

History of Storage Products



The Intelligence Era is Coming



Steam Age



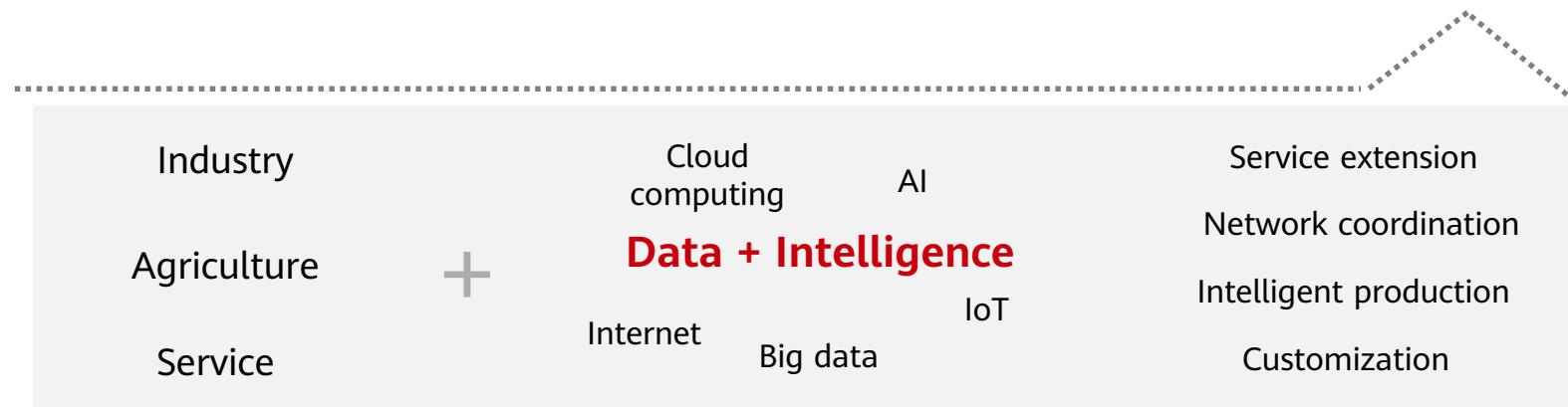
Electricity Age



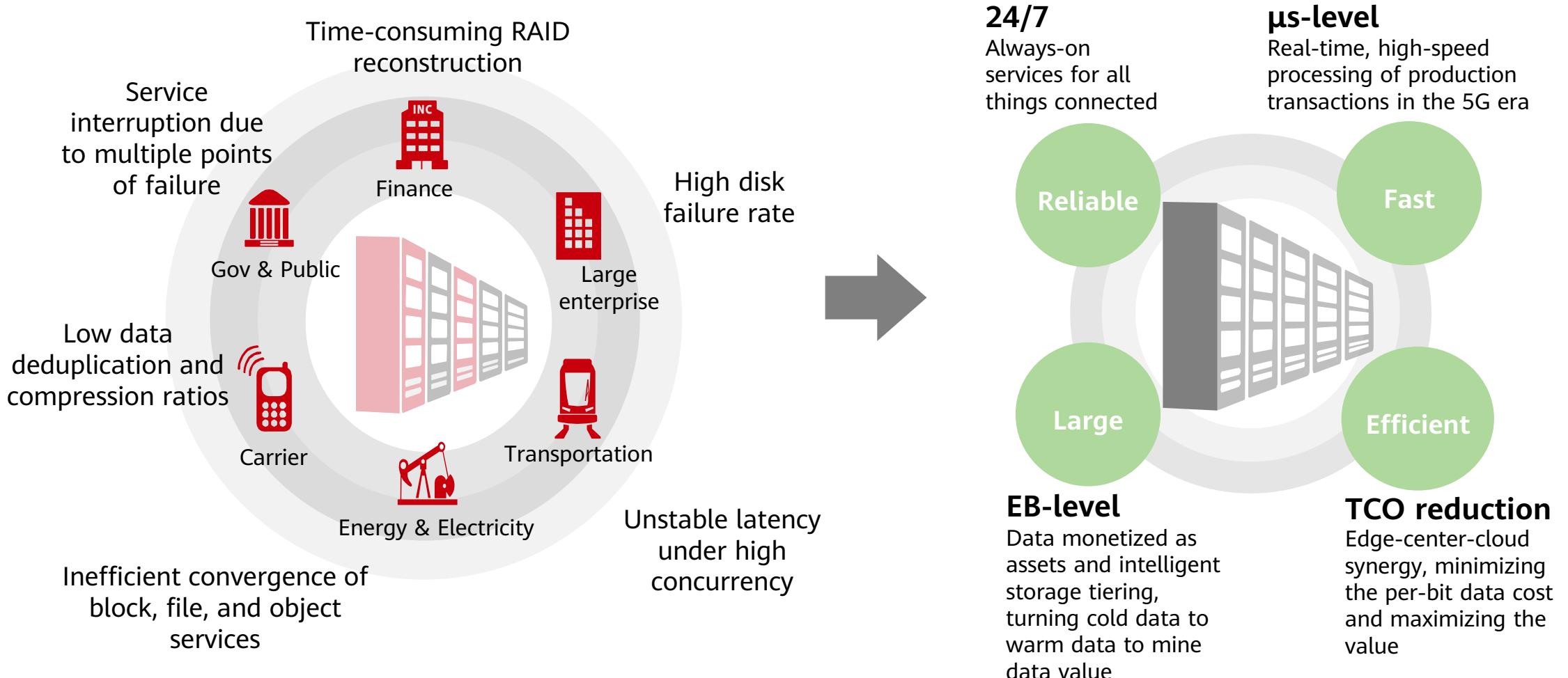
Information Age



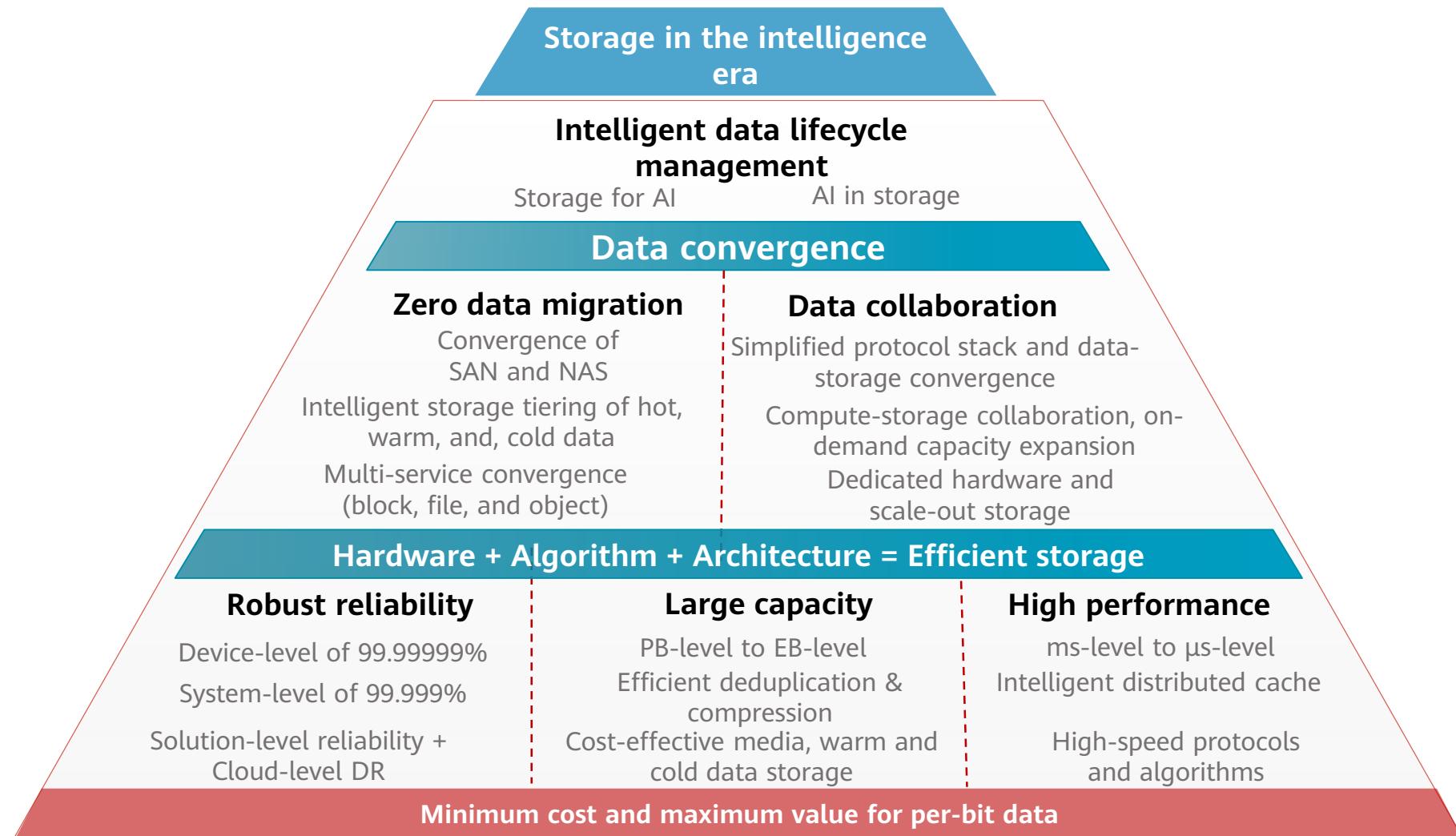
Intelligence Age



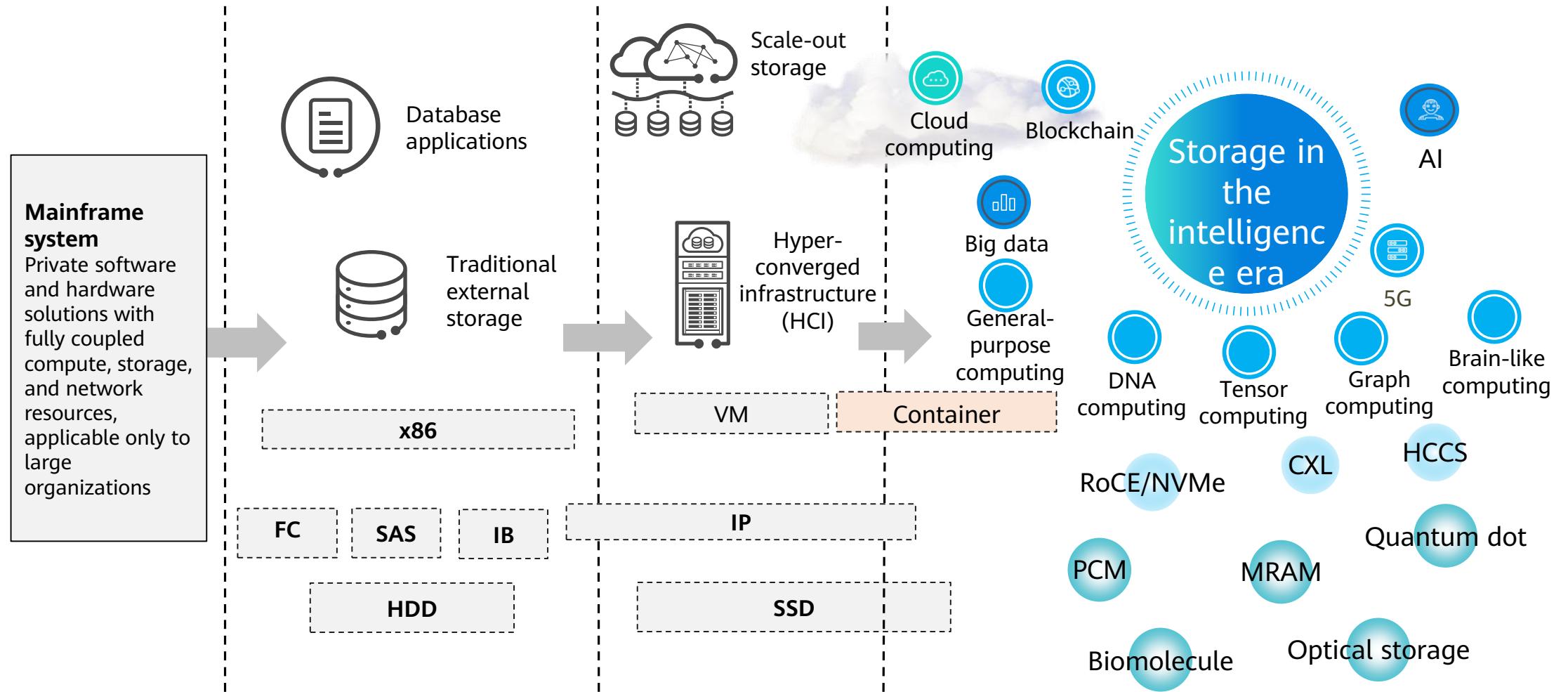
Challenges to Data Storage



Characteristics of Storage in the Intelligence Era



Data Storage Trend



Optical Storage Technology

Blu-ray storage



Long service
life

High
reliability

100 GB+ per
disk

Gold nanostructured glass



Low power
consumption

Long-term
stable storage

10 TB per
disk

50 years >> 600 years

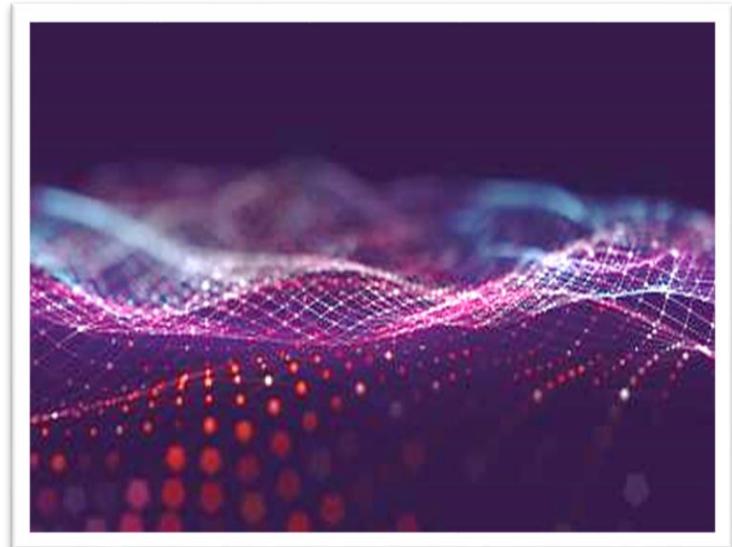
DNA Data Storage

- A small number of synthetic DNA molecules can store a large amount of data, and can freeze, dry, transport, and store data for thousands of years.
- Advantages of using DNA as storage media:
 - Small size
 - High density
 - Strong stability
- Bottlenecks and limitations:
 - High costs of DNA molecular synthesis
 - Slow data read and search



Atomic Storage

- In 1959, physicist Richard Feynman suggested that it was possible to use atoms to store information if they could be arranged the way we wanted.
- Because an atom is so small, the capacity of atomic storage will be much larger than that of existing storage media in the same size.
- With the development of science and technology, arranging the atoms the way we want has become a reality.
- Bottlenecks and limitations:
 - Strict requirements on the operating environment



Quantum Storage

- Now, information in electronic devices is stored and moved through the flow of electrons.
- If electrons are replaced by photons, the movement of information within a computer may occur at the speed of light.
- Although the storage efficiency and service life are improved, the quantum storage is still difficult to be widely applied at present.
- Quantum storage cannot meet the following requirements:
 - High storage efficiency
 - Low noise
 - Long service life
 - Operating at room temperature



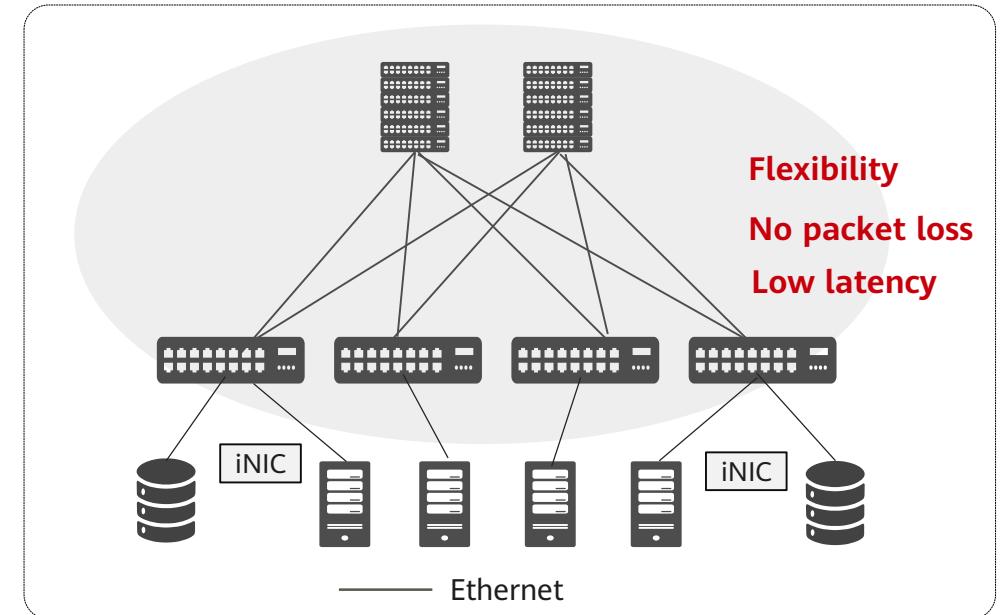
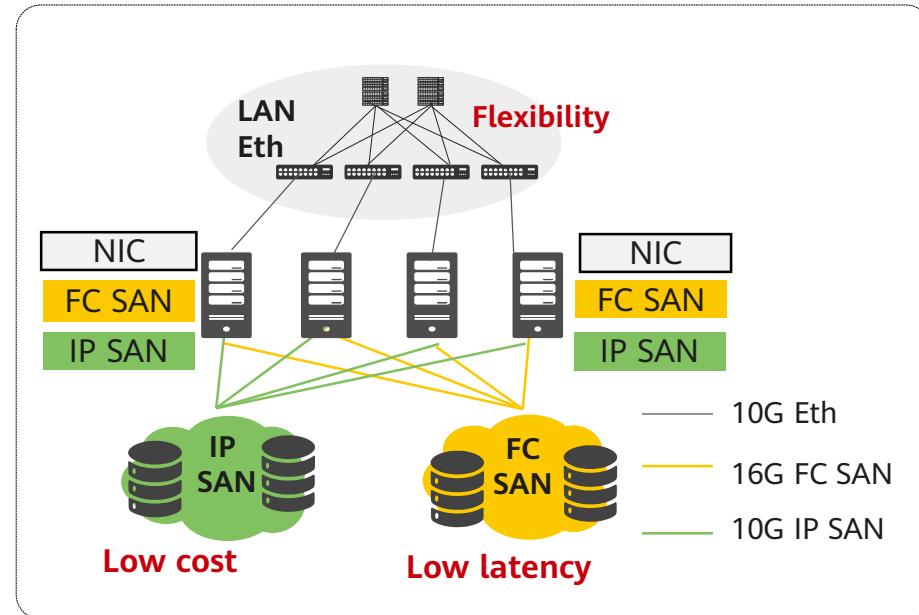
Storage Network Trend

AS-IS FC SAN and IP SAN

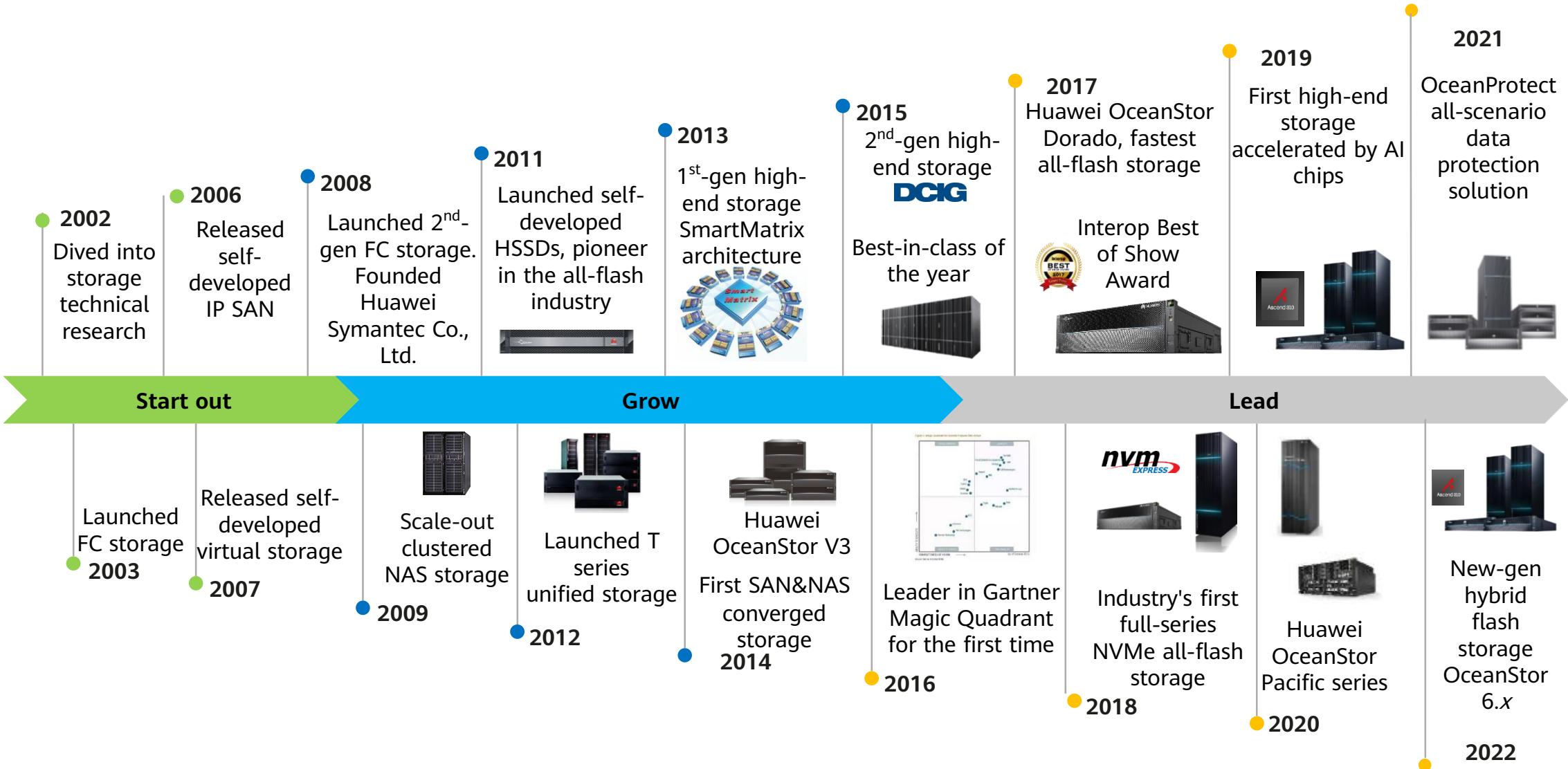
High network costs: The FC private network has low latency but high costs. The IP SAN has low costs but high latency and poor performance.
High O&M costs: IP SAN and FC SAN require dedicated O&M personnel separately, and do not support cloud-and-network synergy.

TO-BE Converged AI Fabric network

Reduced network costs: Open Ethernet carries high-performance, low-latency, and low-cost storage networks.
Reduced O&M costs: No dedicated O&M skills are required, and unified network management is supported for data centers.



History of Huawei Storage Products



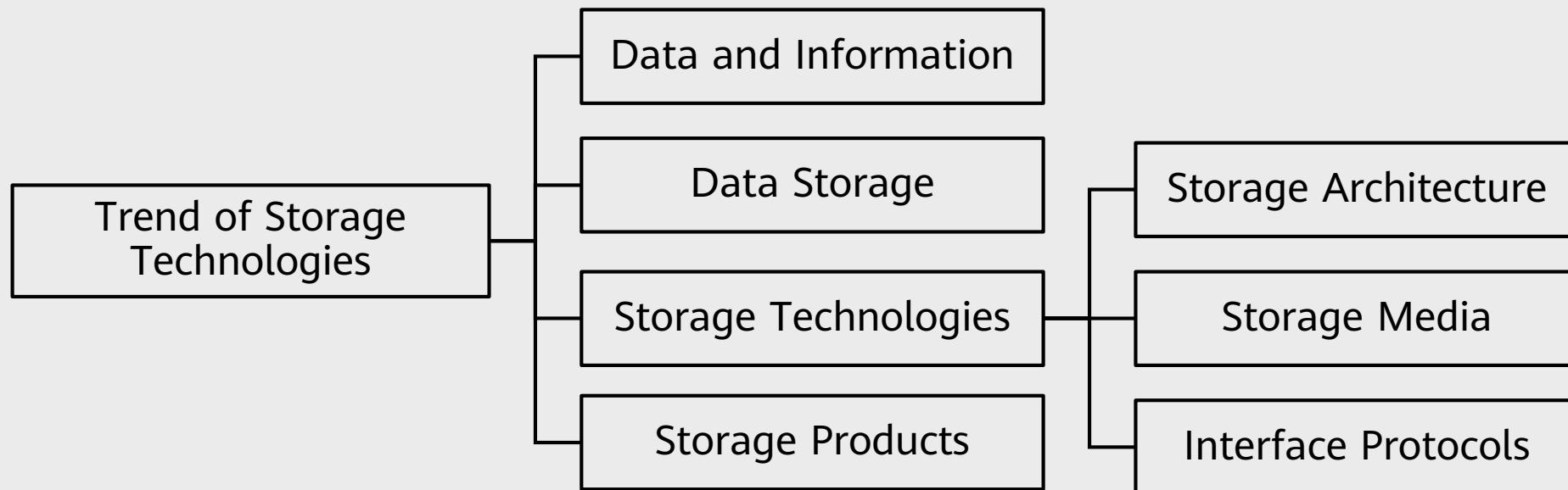
Quiz

1. (Choose multiple options) Which of the following are data types?
 - A. Structured data
 - B. Semi-structured data
 - C. Unstructured data
 - D. Massive amounts of data
2. (Choose multiple options) Which of the following statements about storage are correct?
 - A. Storage refers to disks.
 - B. A storage system consists of storage hardware, software, and solutions.
 - C. Storage types include block storage, file storage, and object storage.
 - D. File storage is used to store data of data applications.

Quiz

3. (Choose multiple options) Which of the following are characteristics of cloud storage?
 - A. Convergence
 - B. Open
 - C. Elasticity
 - D. Scale-up
4. (Choose multiple options) Which of the following are the objectives of integrating AI into storage?
 - A. Simple
 - B. Efficient
 - C. High power consumption
 - D. Easy to use

Summary



Recommendations

- Huawei official websites
 - Enterprise business: <https://e.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/en/index.html>
 - Online learning: <https://www.huawei.com/en/learning>
- Popular tools
 - HedEx Lite
 - Network Document Tool Center
 - Information Query Assistant

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.



Intelligent Data Storage System



Foreword

- This course describes the components of the storage system, including the controller enclosure, disk enclosure, disks, and interface modules, the expansion methods of the storage system, and the working principles of the storage media and the components.

Objectives

Upon completion of this course, you will understand:

- Storage product forms
- Functions and components of controller enclosures and disk enclosures
- Working principles of HDDs and SSDs
- Concepts of scale-up and scale-out, and related cables and interface modules

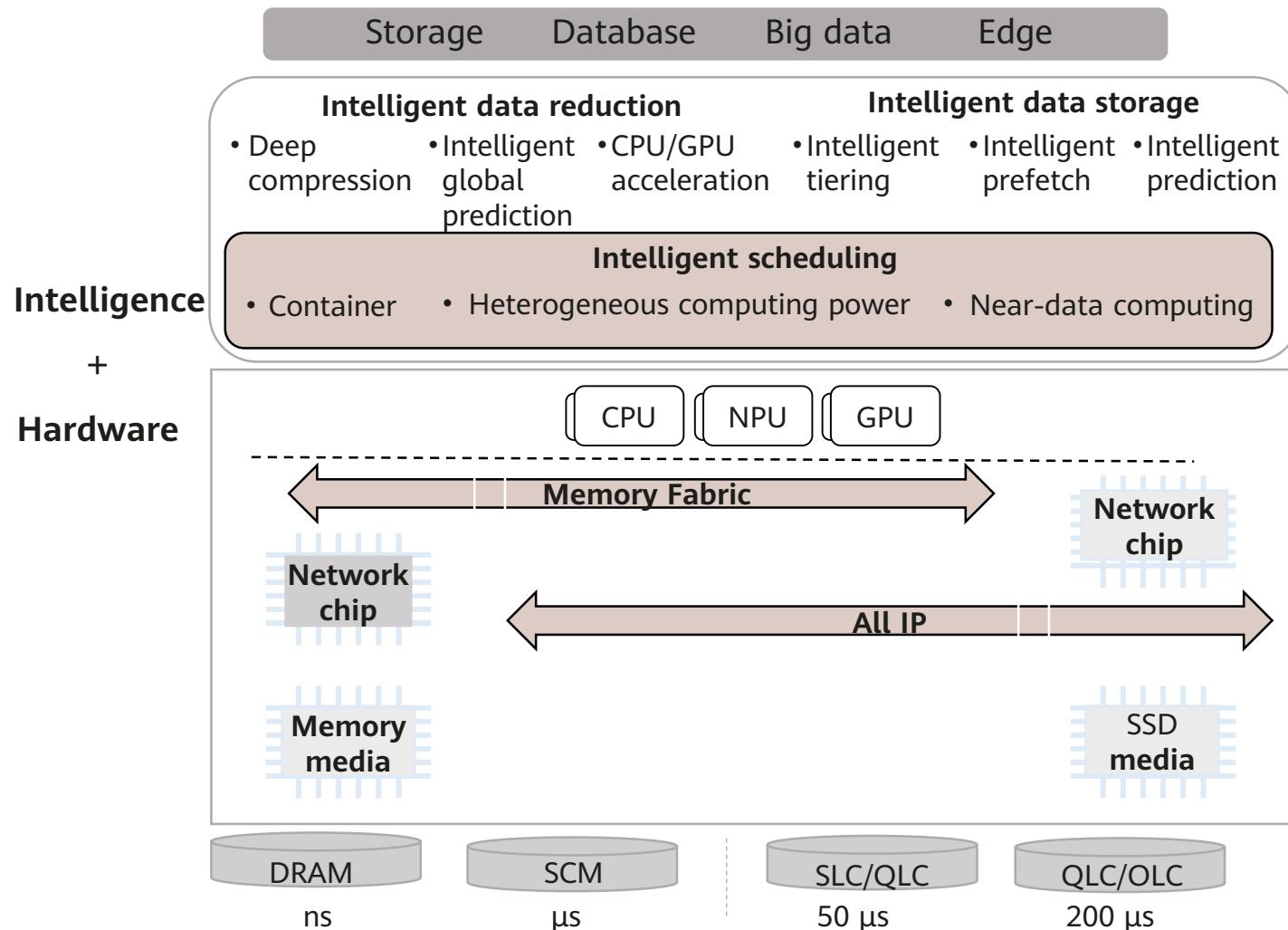
Contents

- 1. Intelligent Data Storage System**
- Intelligent Data Storage Components
- Storage System Expansion Methods

Storage in the Intelligent Era

- As we move rapidly toward an intelligent world, data is being generated at an unprecedented rate.
- More companies have realized that the key to achieving smartization is data infrastructure, with storage at its core.
- An intelligent world calls for intelligent storage.
 - Storage for AI
 - Support companies to adopt intelligent technologies to accomplish AI training and application.
 - AI in Storage
 - Integrate AI technologies into the full-lifecycle management to improve storage management, performance, efficiency, and stability.

Intelligent Data Storage Architecture



Intelligent data reduction

- AI-based prediction
- CPU and GPU intelligent reduction algorithm

Intelligent data storage

- Intelligent prefetch, data tiering, hotspot identification, data caching and other technologies for optimal media combination

Intelligent scheduling

- Dynamic management of heterogeneous computing resources, near-data computing scheduling, and quick start of containers

Memory Fabric: building a memory-centric and high-performance network

- Supporting high-performance networks with nanosecond-level latency
- Memory pooling and tiering

All IP: ultimate cost reduction based on SSD media

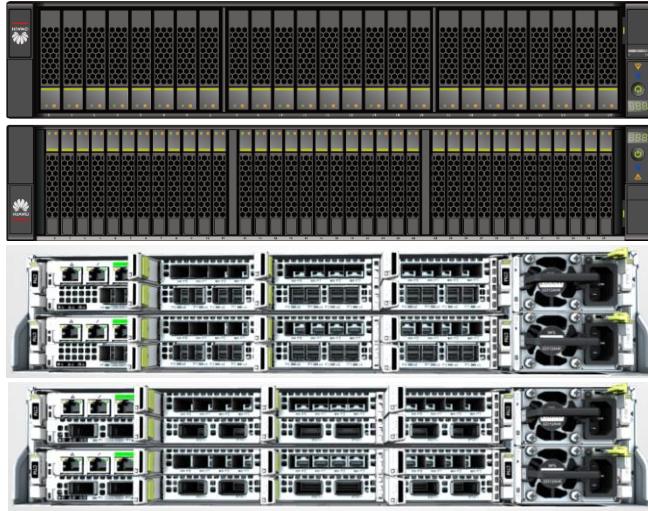
- In-depth disk-controller collaboration, and SLC, QLC, and OLC evolution
- Building a simplified network with all IP

Contents

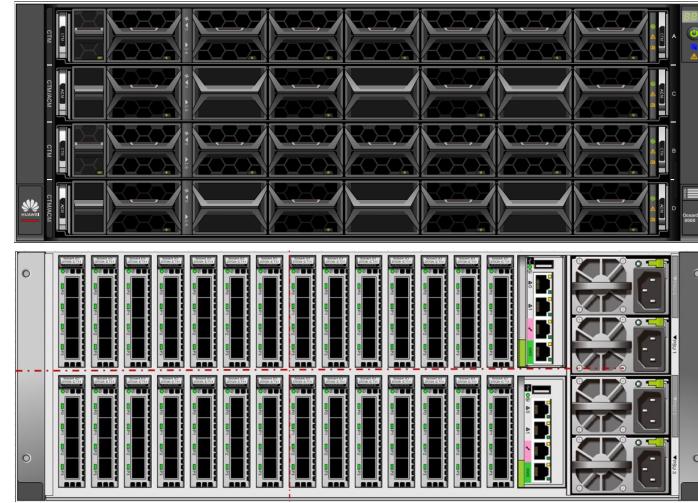
2. Intelligent Data Storage Components

- Controller enclosure
 - Disk enclosure
 - Expansion module
 - Disk
 - Interface module

Storage Product Form



2 U, disk and controller integration



4 U, disk and controller separation

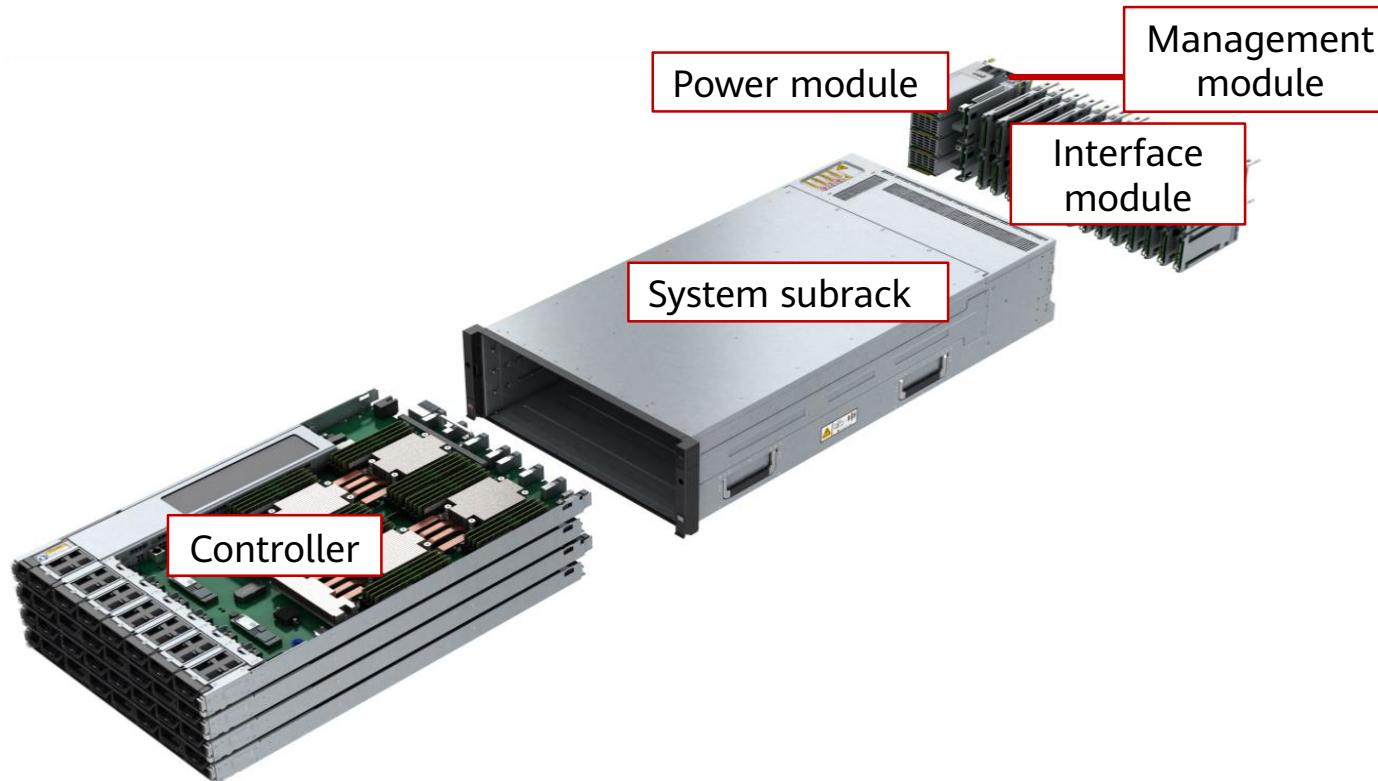


Integrated bay

Note: Huawei OceanStor Dorado V6 is used as the example.

Controller Enclosure

- The controller enclosure uses a modular design and consists of a system subrack, controllers (with built-in fan modules), BBUs, power modules, management modules, and interface modules.



Front View of a Controller Enclosure



2 U controller enclosure (disk and controller integration)

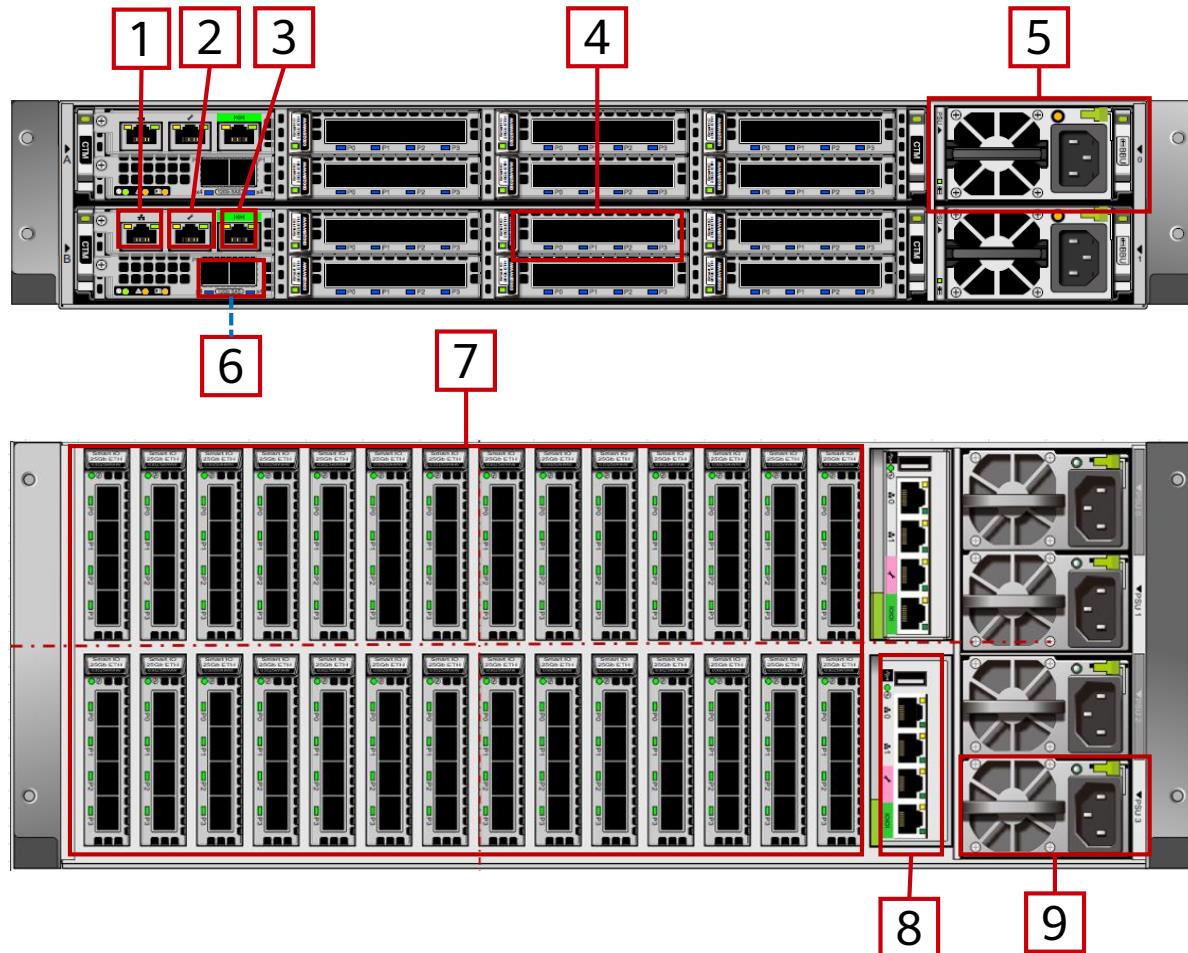


4 U controller enclosure (disk and controller separation)

Icon	Description
	Enclosure ID indicator
	Enclosure location indicator 1. Blinking blue: The controller enclosure is being located. 2. Off: The controller enclosure is not located.
	Enclosure alarm indicator 1. Steady amber: An alarm is reported by the controller enclosure. 2. Off: The controller enclosure is working properly.
	Power indicator/Power button

Note: Huawei OceanStor Dorado V6 is used as the example.

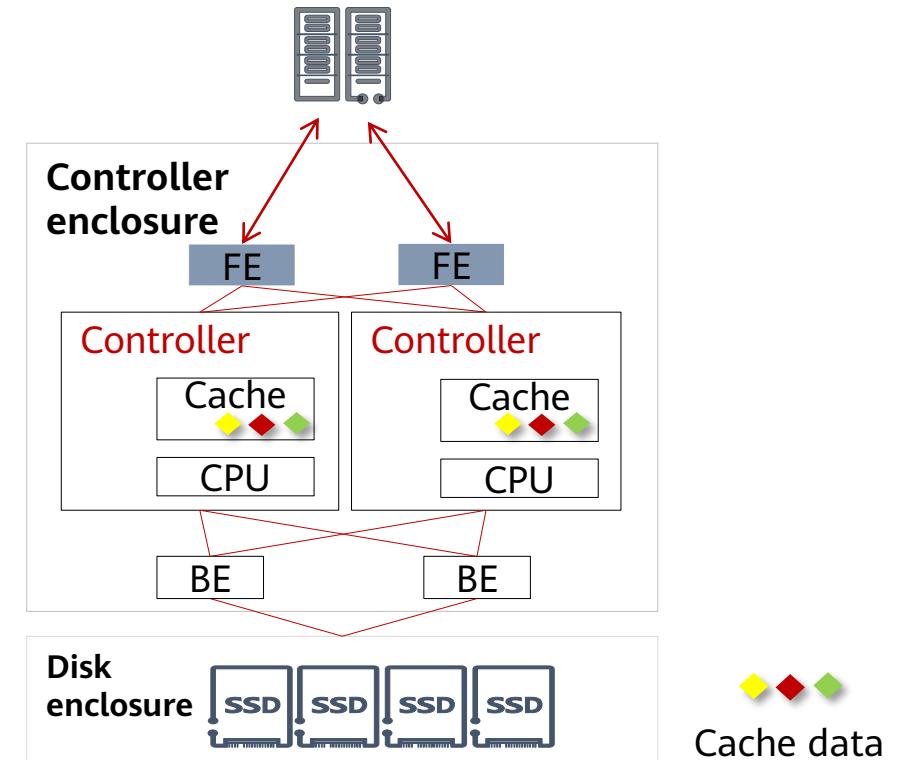
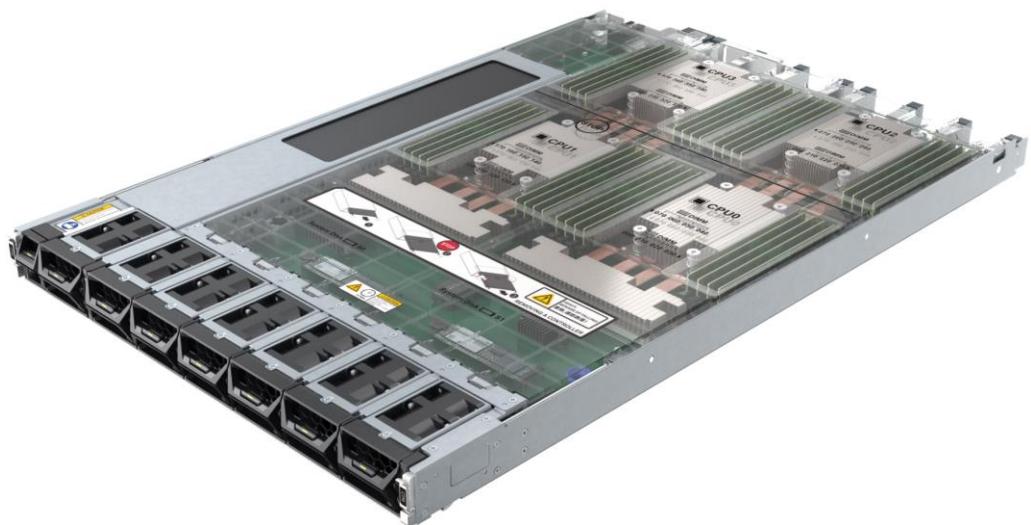
Rear View of a Controller Enclosure



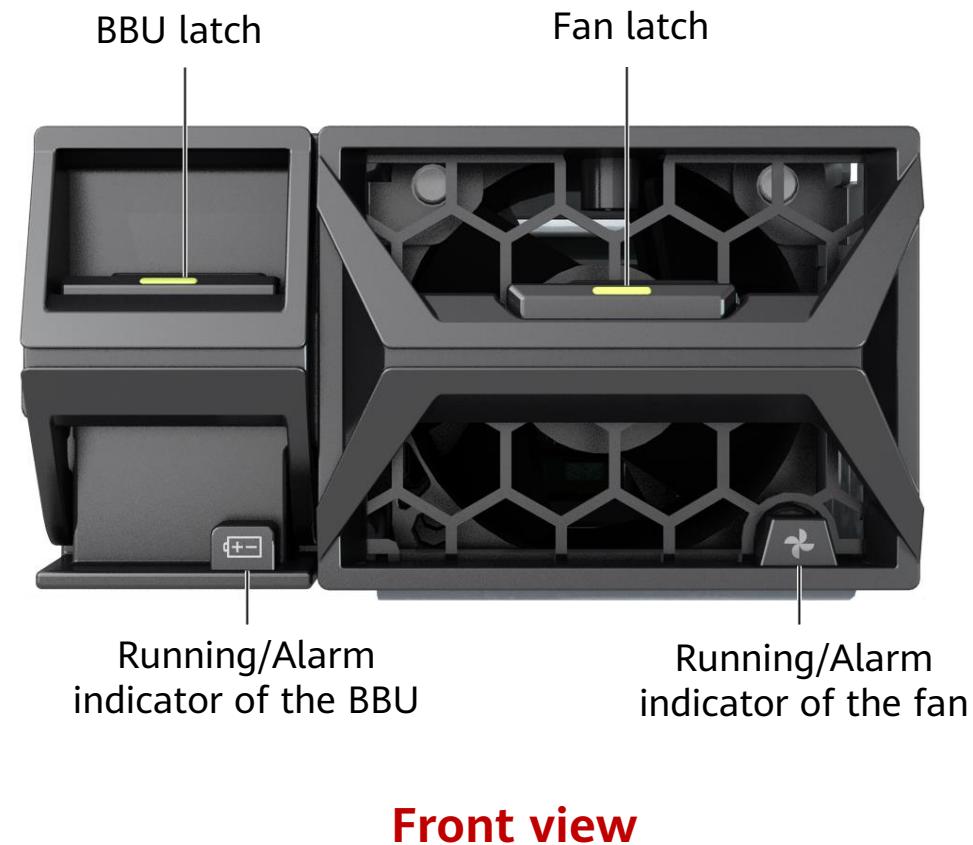
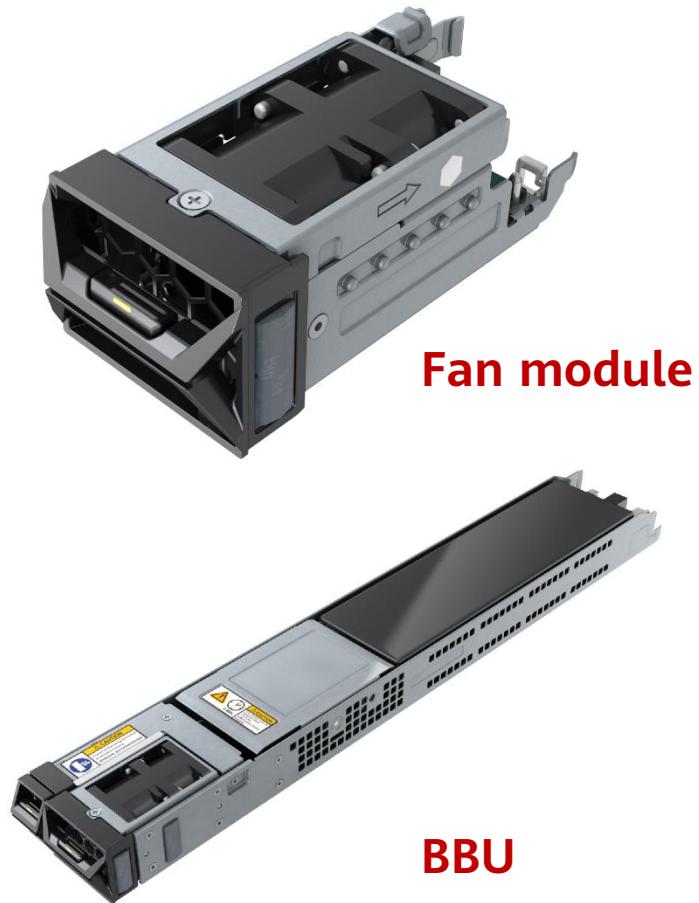
No.	Description
1	Management port
2	Maintenance port
3	Serial port
4	Interface module
5	Power-BBU module
6	SAS expansion port
7	Interface module
8	Management module
9	Power module

Controller

- A controller is the core component of a storage system. It processes storage services, receives configuration management commands, saves configuration data, connects to disks, and saves critical data to cache disks.

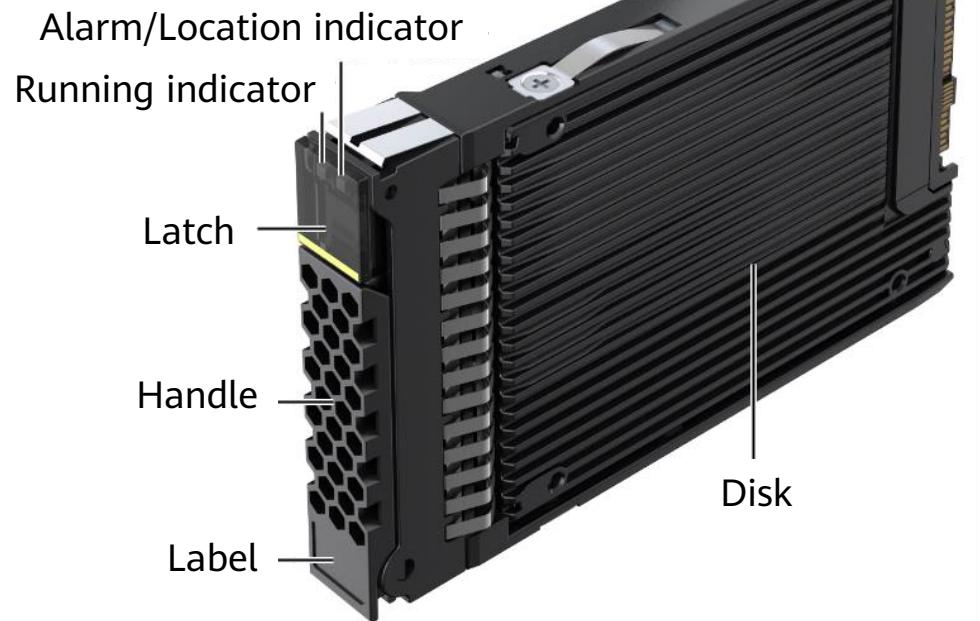


BBU and Fan Module

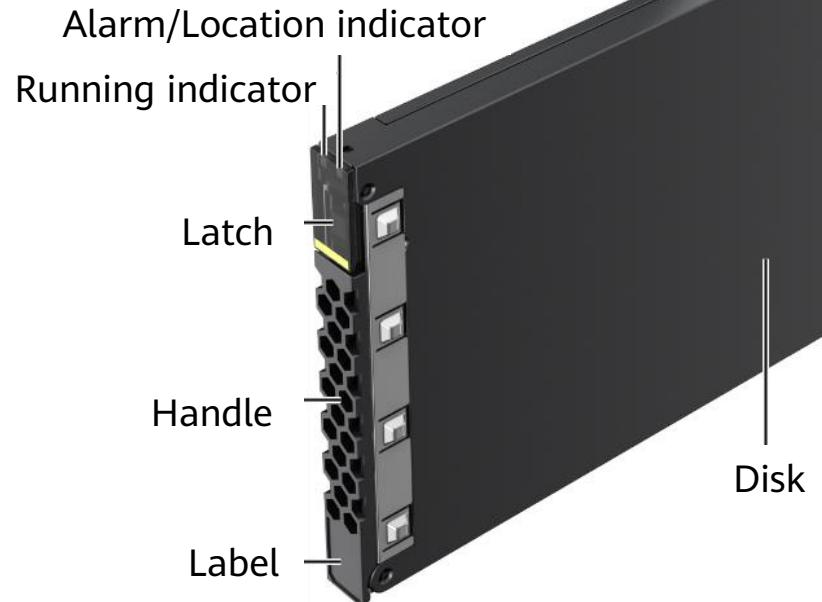


Coffer Disk

2.5-inch coffer disk



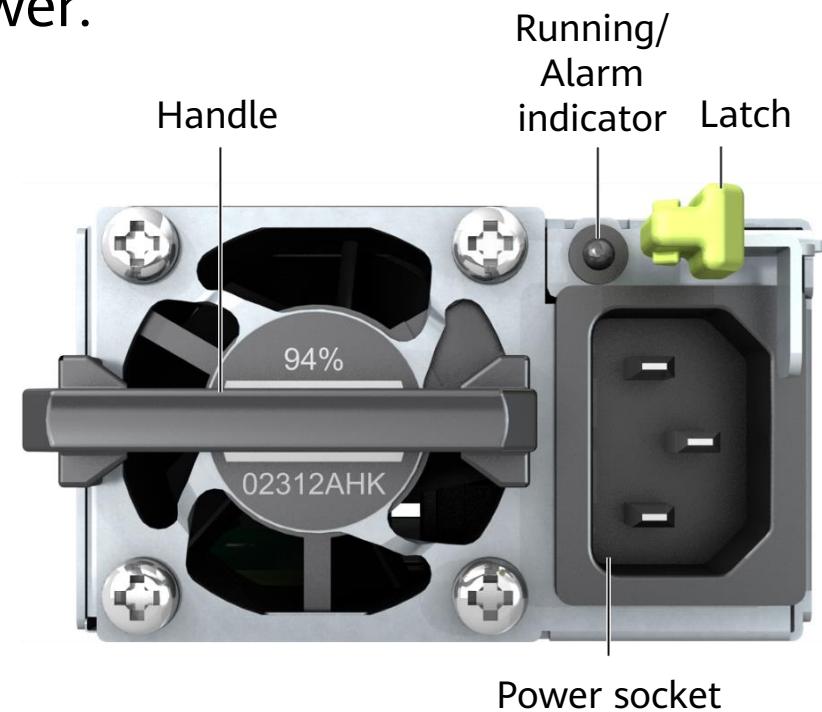
Palm-sized NVMe coffer SSD



Note: Huawei OceanStor Dorado V6 is used as the example.

Power Module

- The AC power module supplies power to the controller enclosure, allowing the enclosure to operate normally at maximum power.



Note: Huawei OceanStor Dorado V6 is used as the example.

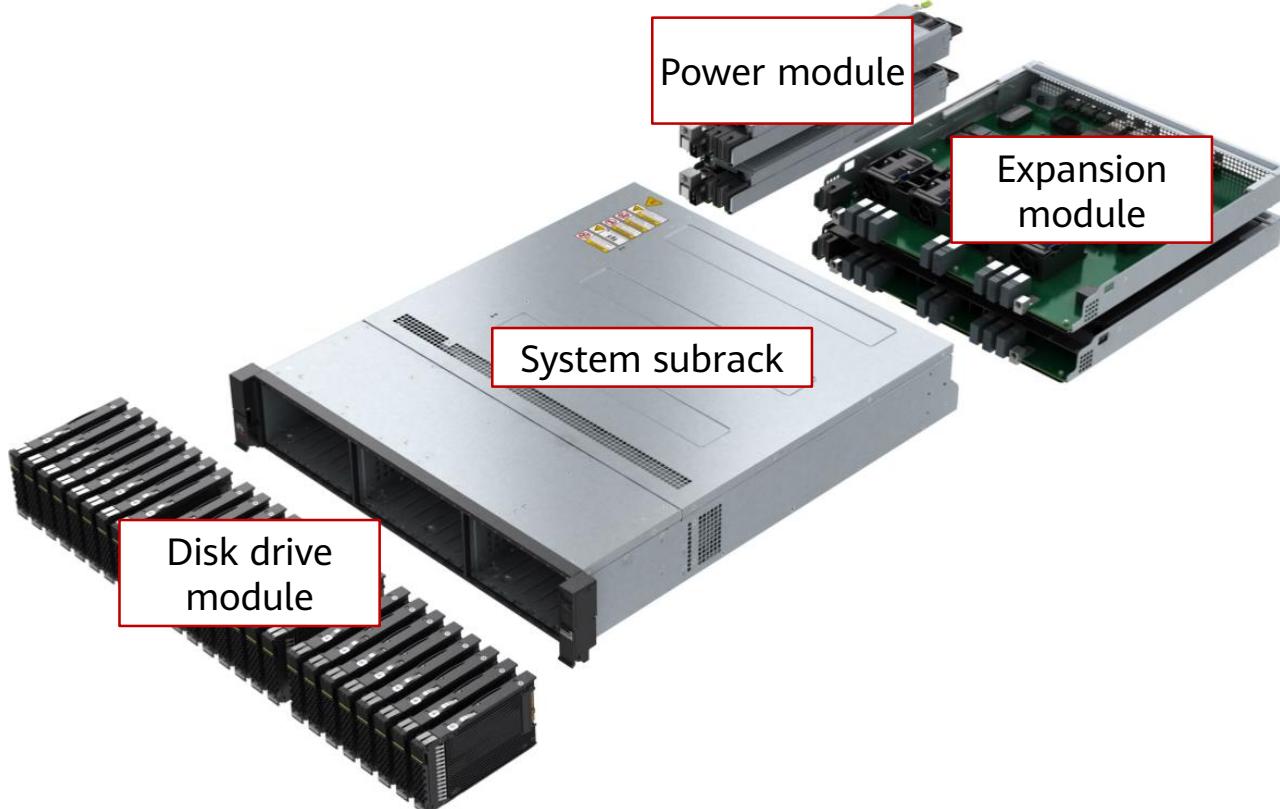
Contents

2. Intelligent Data Storage Components

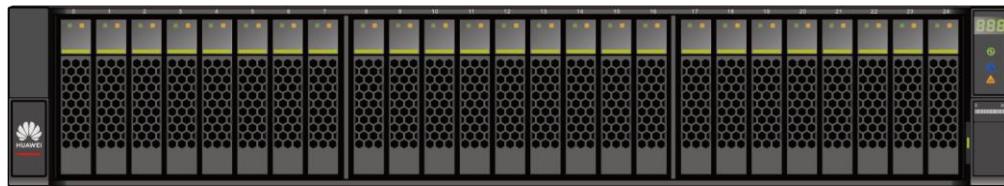
- Controller enclosure
- Disk enclosure
 - Expansion module
 - Disk
 - Interface module

Disk Enclosure

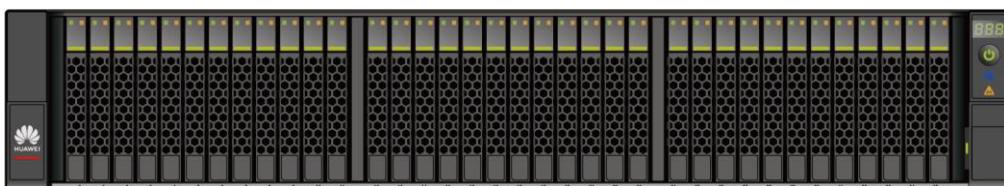
- The disk enclosure uses a modular design and consists of a system subrack, expansion modules, power modules, and disks.



Front View of a Disk Enclosure



2 U 25-slot smart SAS disk enclosure

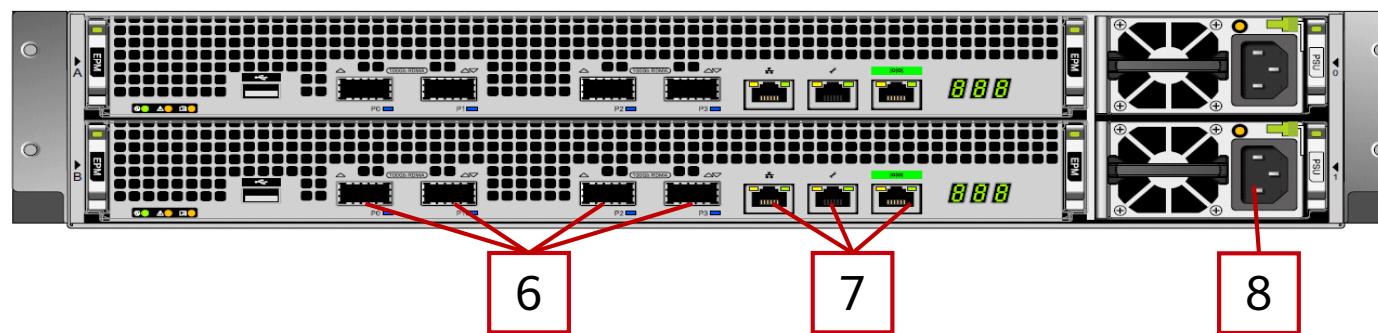
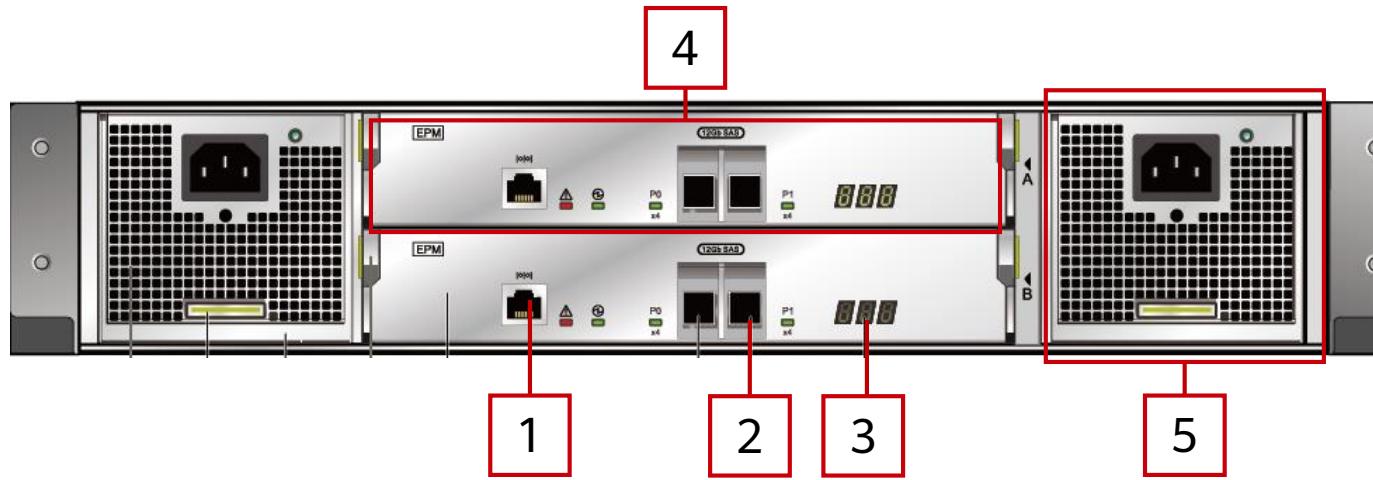


2 U 36-slot smart NVMe disk enclosure

Note: This slide shows the front views of a 2 U SAS disk enclosure and a 2 U smart NVMe disk enclosure of Huawei OceanStor Dorado V6.

Icon	Description
	ID indicator of the disk enclosure
	Location indicator of the disk enclosure 1. Blinking blue: The disk enclosure is being located. 2. Off: The disk enclosure is not located.
	Alarm indicator of the disk enclosure 1. Steady yellow: An alarm is reported by the disk enclosure. 2. Off: The disk enclosure is working properly.
	Power indicator of the disk enclosure 1. Steady green: The disk enclosure is powered on. 2. Off: The disk enclosure is powered off.
	Power indicator/Power button 1. The disk enclosure is powered on and off with the controller enclosure. The power button on the disk enclosure is invalid and cannot be used to power on or off the disk enclosure separately.

Rear View of a Disk Enclosure



No.	Description
1	Serial port
2	Mini SAS HD expansion port
3	ID display
4	Expansion module
5	Power module
6	Onboard expansion port
7	Onboard management port
8	Power module

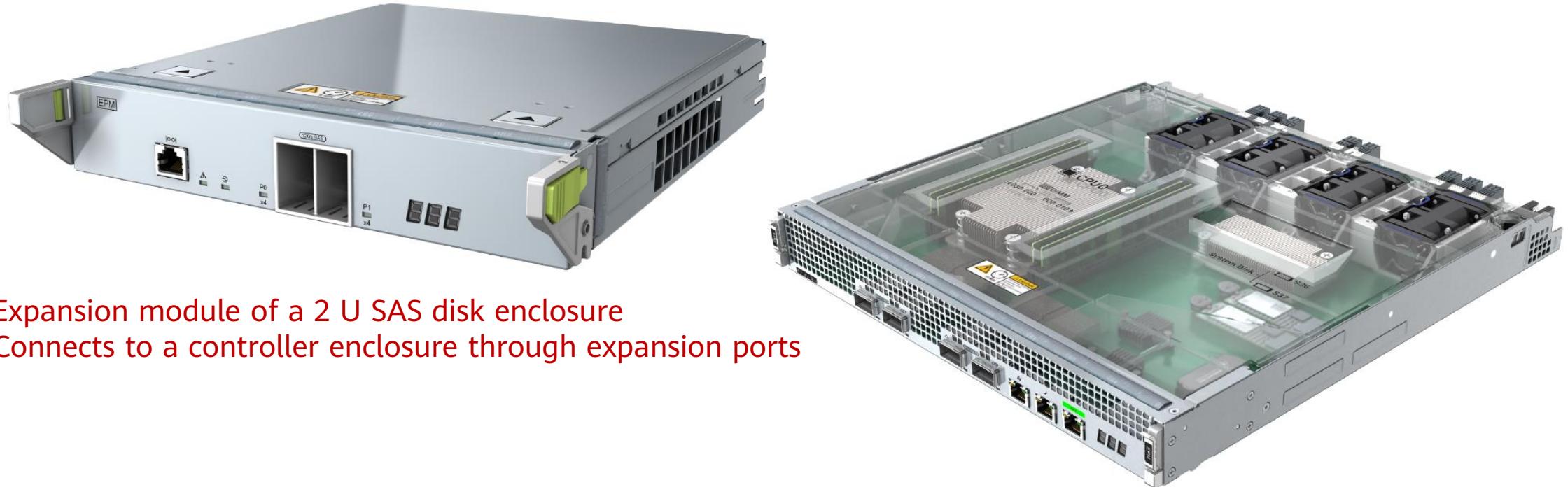
Note: This slide shows the rear views of the 2 U smart SAS and smart NVMe disk enclosures of Huawei OceanStor Dorado V6.

Contents

2. Intelligent Data Storage Components

- Controller enclosure
- Disk enclosure
- Expansion module
- Disk
- Interface module

Expansion Module



Expansion module of a 2 U SAS disk enclosure
Connects to a controller enclosure through expansion ports

Expansion module of a 2 U smart NVMe disk enclosure
Connects to a controller enclosure through expansion ports

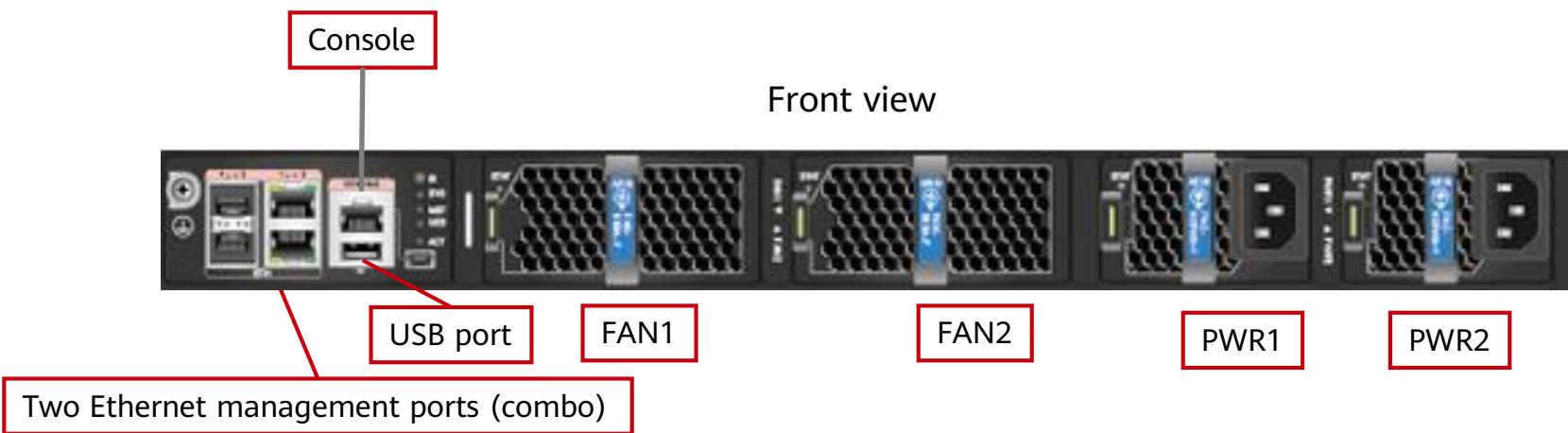
CE Switch

Rear view



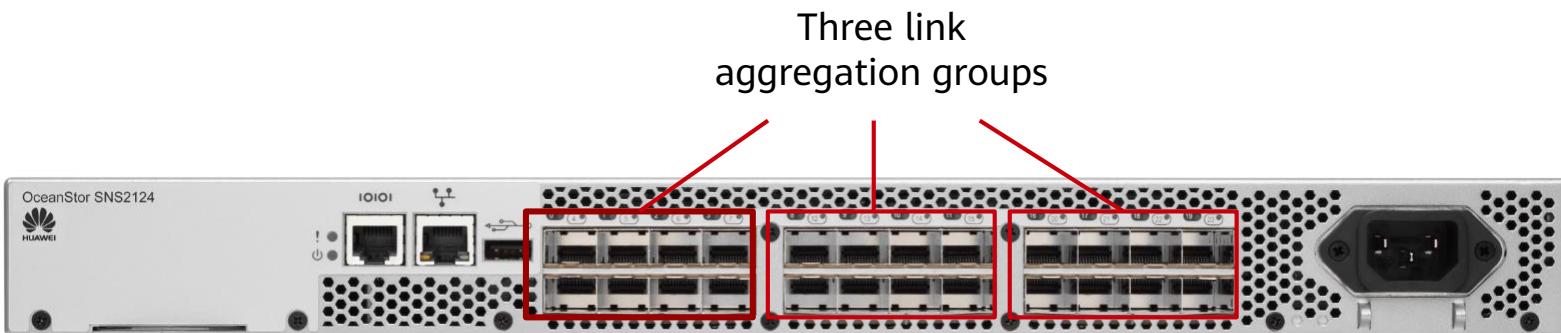
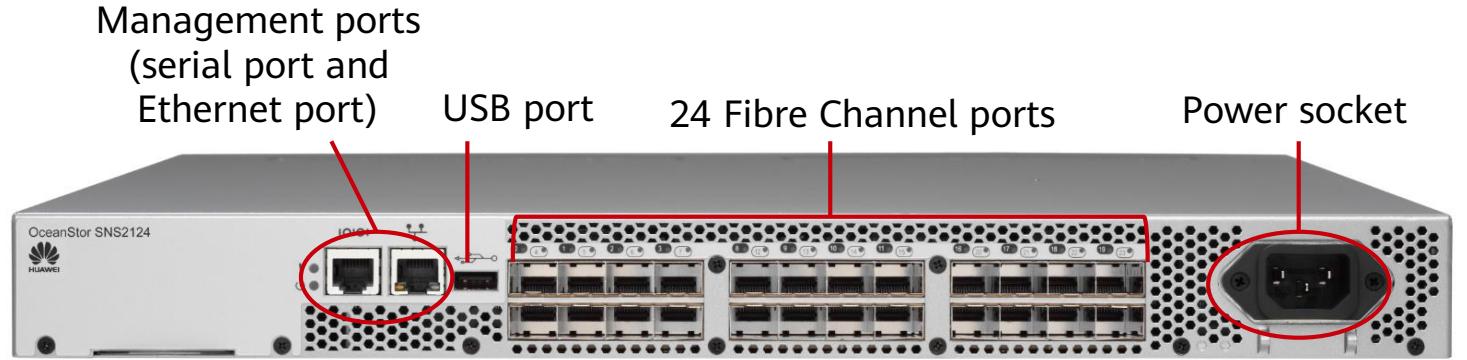
Console

Front view



Note: Huawei CE6800 series switches are used as an example.

Fibre Channel Switch



Note: Huawei SNS2124 is used as an example.

Device Cables



1. Serial cable



2. Mini SAS HD
electrical cable



3. Mini SAS HD
optical cable



4. 100G
QSFP28 cable



5. 25G SFP28 cable



6. MPO-4*DLC
optical fiber



7. Optical fiber

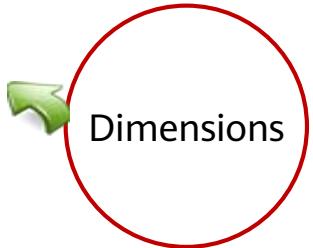
Contents

2. Intelligent Data Storage Components

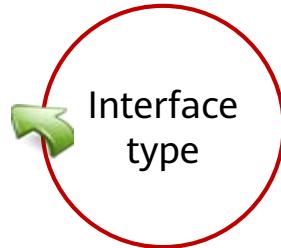
- Controller enclosure
- Disk enclosure
- Expansion module
- Disk
 - HDD
 - SSD
- Interface module

Disk Type

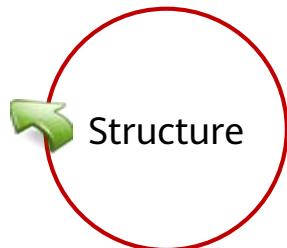
1.8-inch
2.5-inch
3.5-inch
5.25-inch
...



IDE
SCSI
SATA
SAS
FC
NVMe
...



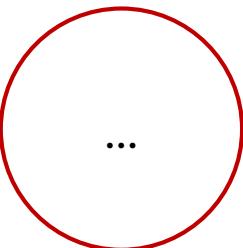
HDD
SSD



What are the types of disks?

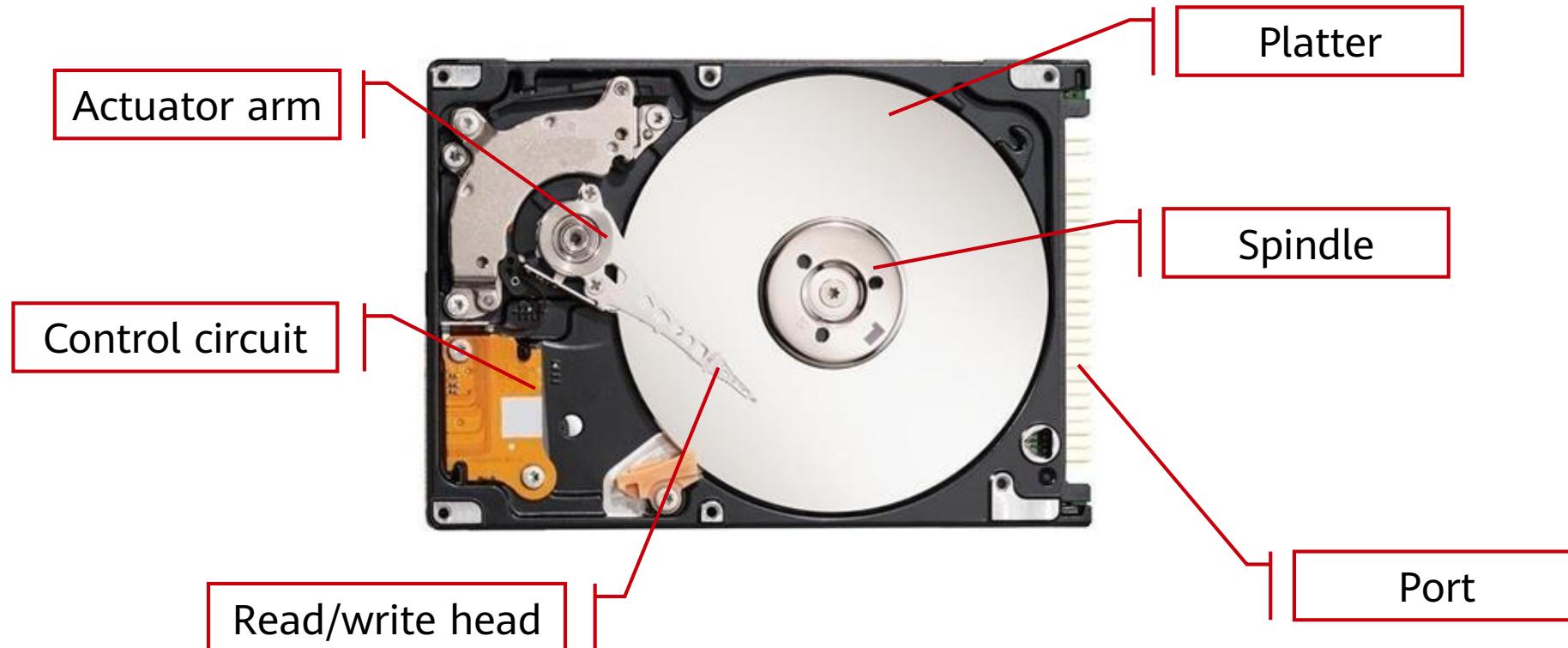


Enterprise-class
Desktop-class
...



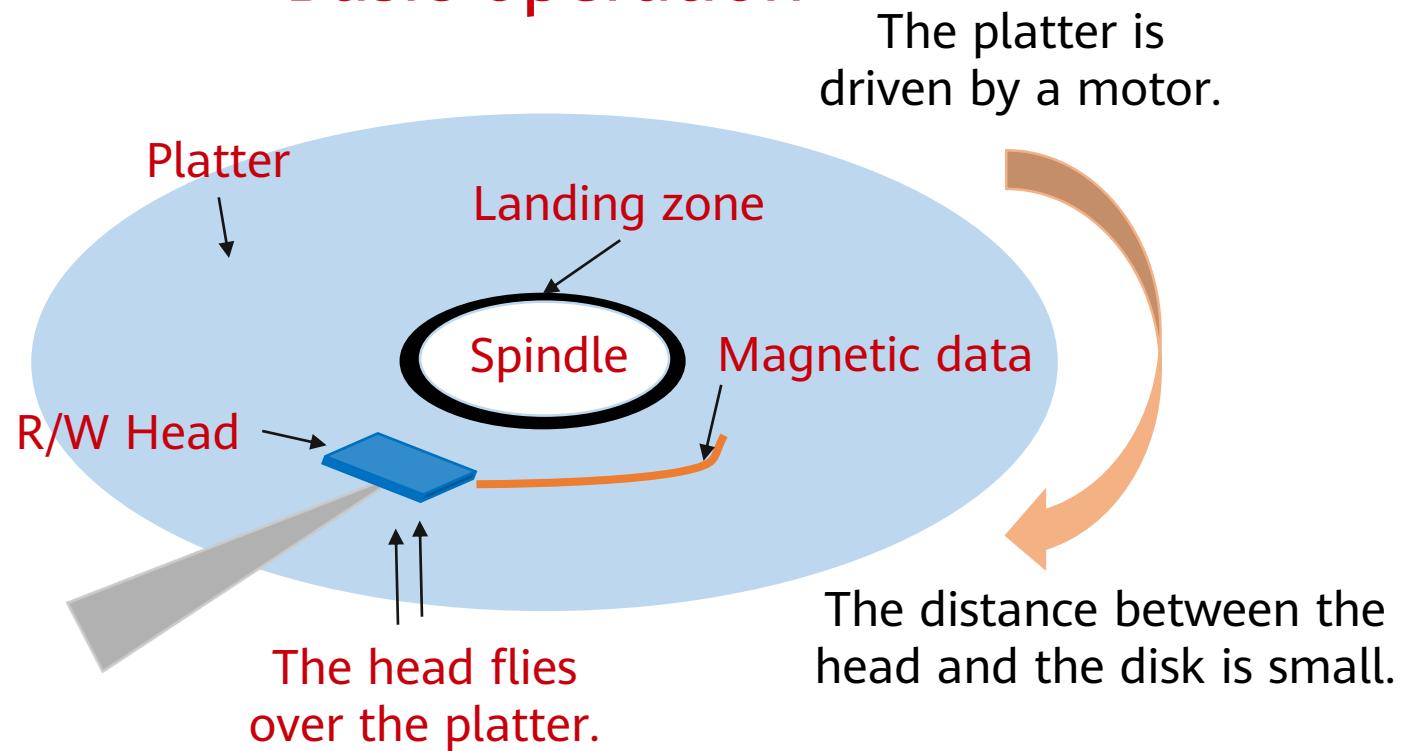
HDD Structure

- An HDD consists of platters, an actuator arm, read/write heads, a spindle, a port, and control circuits.

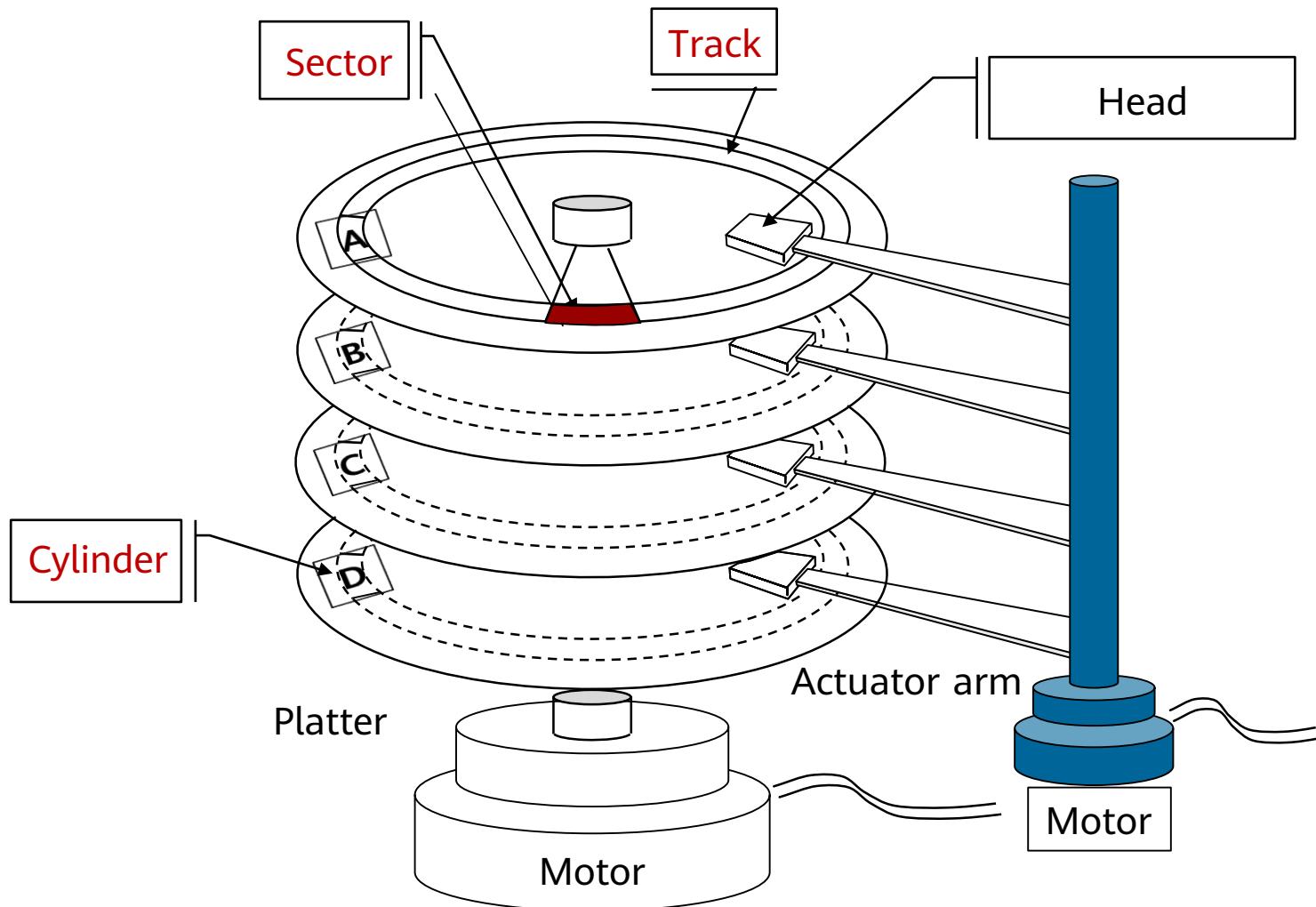


HDD Working Principles

Basic operation



Data Organization on a Disk



Disk Capacity and Cache

- Disk capacity
 - Disk capacity = Number of cylinders x Number of heads x Number of sectors x 512 bytes.
The unit is MB or GB. The disk capacity is determined by the **capacity of a single platter** and the **number of platters**.
- Cache
 - Because the processing speed of a CPU is much faster than that of a disk, the CPU must wait until the disk completes a read/write operation before issuing a new command. To solve this problem, a cache is added to the disk to improve the read/write speed.

Factors Relevant to Disk Performance



Rotation speed

Primary factor that determines the throughput in the case of sequential I/Os

Seek speed

Primary factor that affects the random I/O performance

Single platter capacity

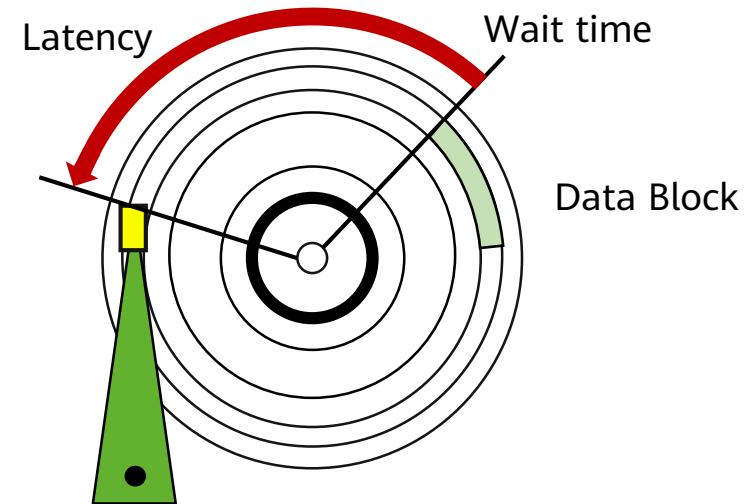
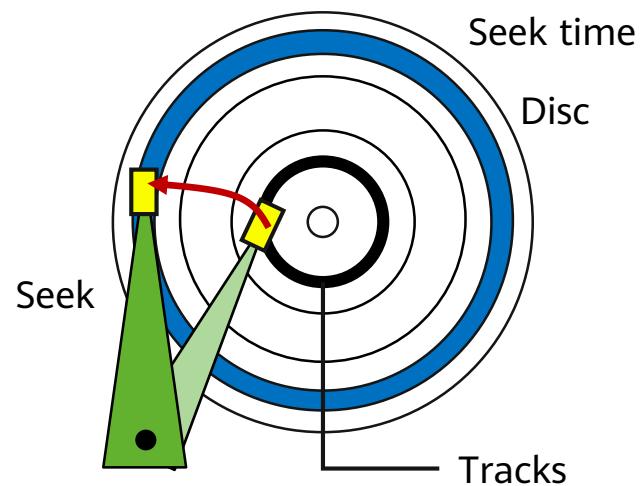
Indirect factor for disk performance

Port speed

The least important factor for disk performance

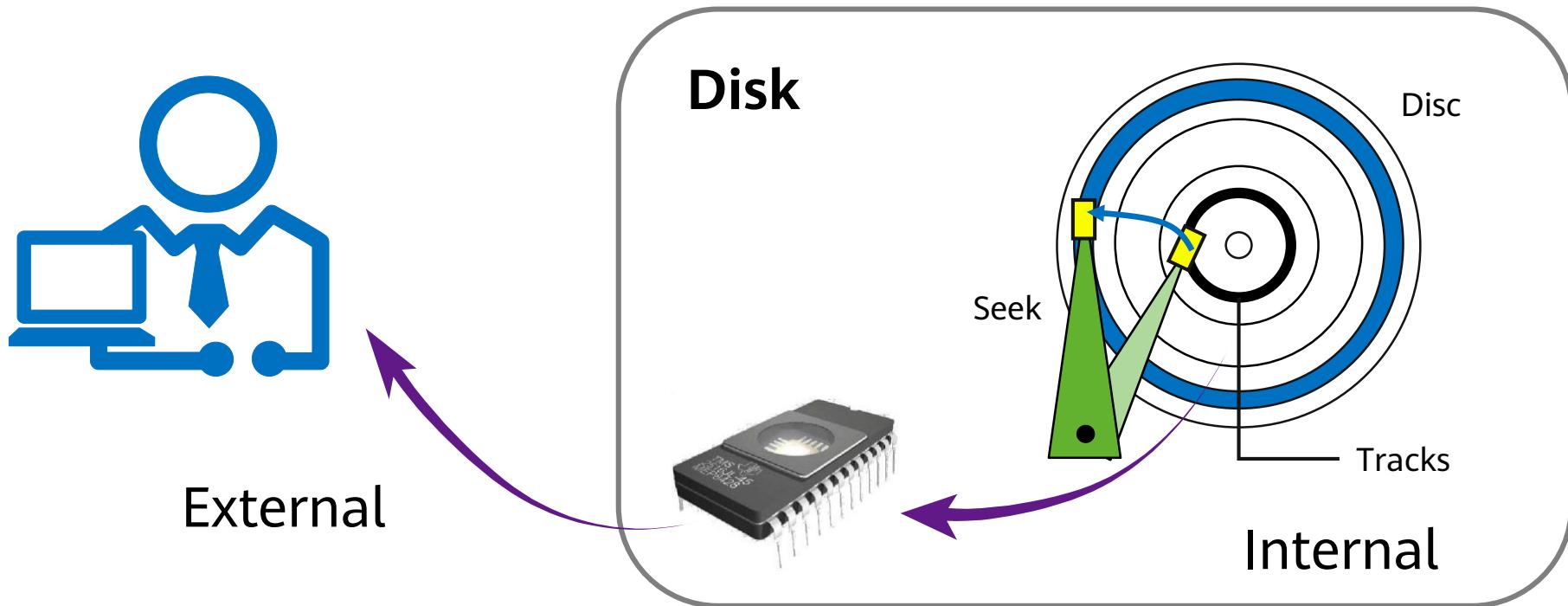
Average Access Time

- The average access time is determined by:
 - Average seek time
 - Average latency time



Data Transfer Rate

- The data transfer rate is determined by:
 - Internal transfer rate
 - External transfer rate/Interface transfer rate



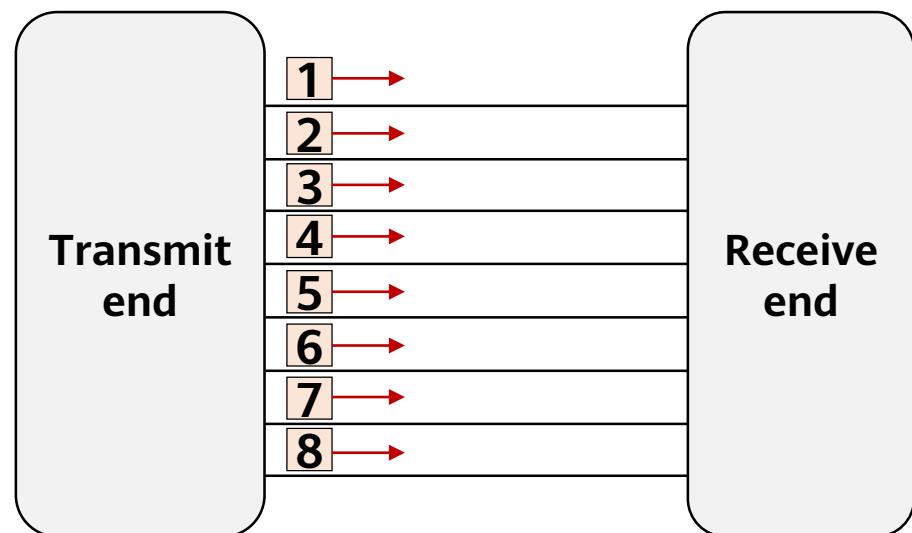
Disk IOPS and Transmission Bandwidth

- IOPS
 - Input/Output operations per second (IOPS) is a key indicator to measure disk performance.
 - IOPS is calculated by the seek time, rotation latency, and data transmission time.
- Transmission bandwidth (throughput)
 - Indicates the amount of data that is successfully transmitted in a unit time, that is, the speed at which data streams are transmitted. For example, if it takes 10s to write 10,000 files of 1 KB size, the transmission bandwidth is only 1 MB/s; if it takes 0.1s to write a 10 MB file, the transmission bandwidth is 100 MB/s.

Parallel and Serial Transmission

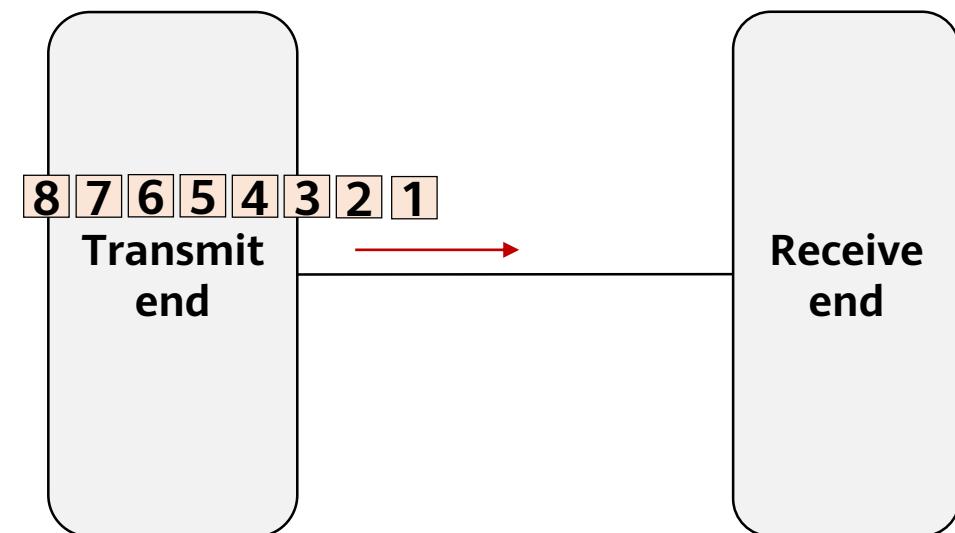
- For example, the methods for transmitting numbers 1 to 8 are as follows:

Parallel transmission



Multiple lines are connected between two ends, and one number is transmitted on each line.

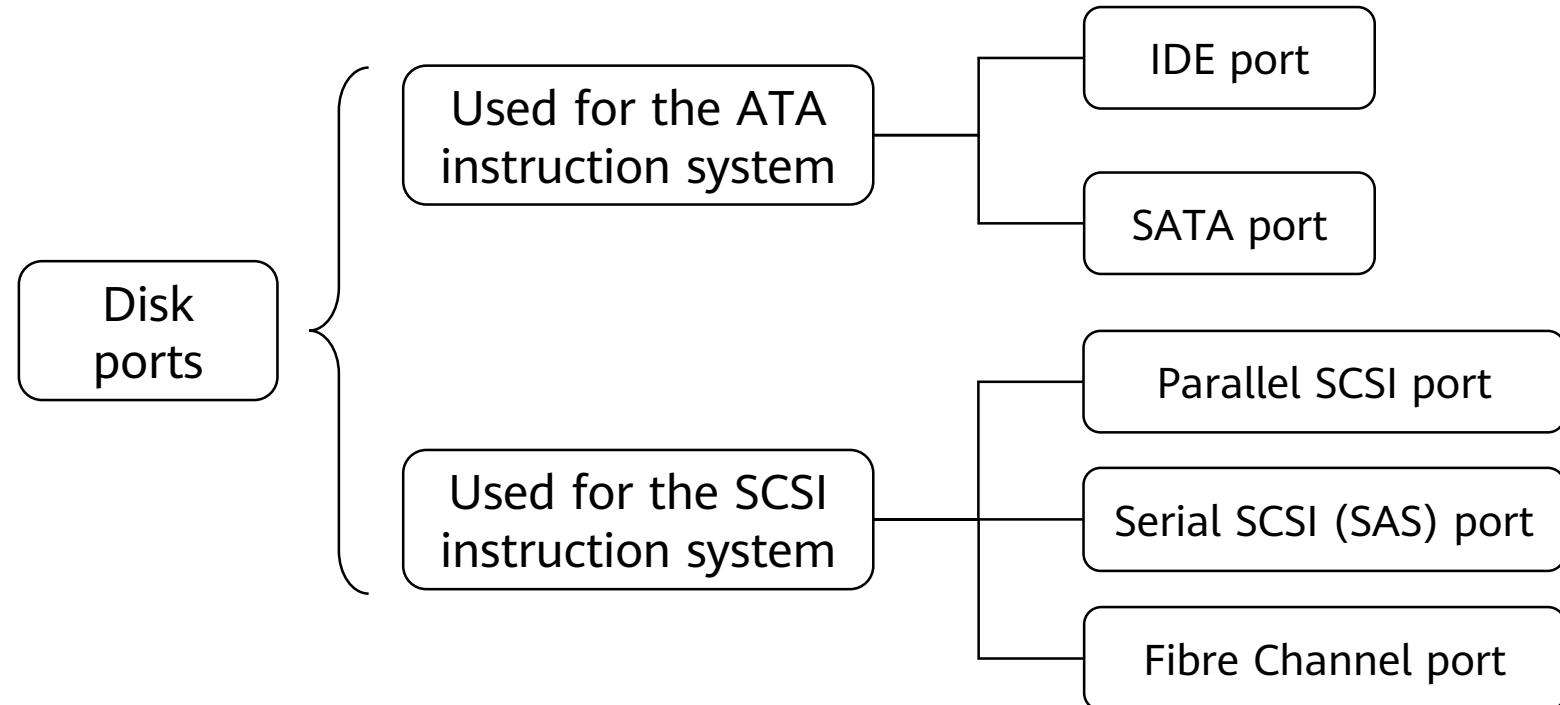
Serial transmission



Only one line is connected between two ends. Eight numbers are sent on this line in sequence. The receive end has all numbers after eight transmissions.

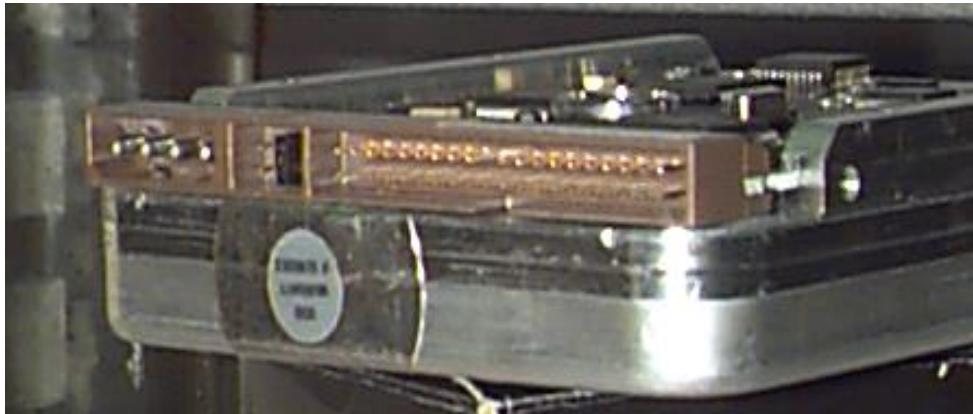
HDD Port Technology

- A disk must provide a simple port for users to access its data. Generally, disks provide the following physical ports:



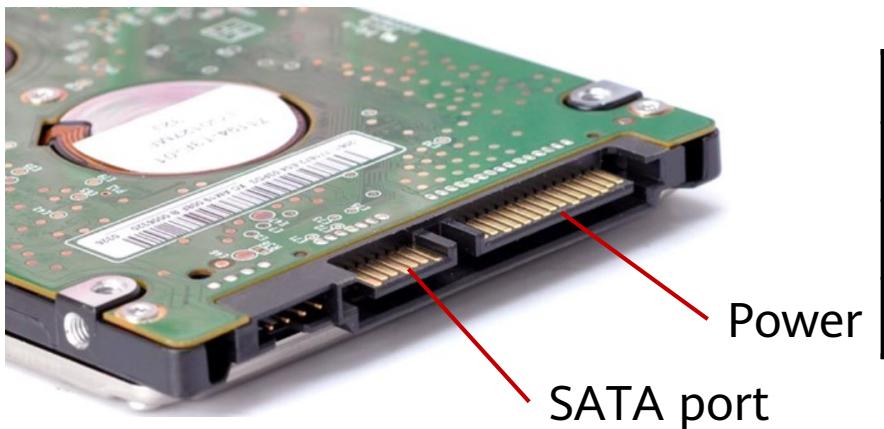
IDE Disk Port

- The integrated drive electronics (IDE) port is also called the parallel ATA port.
 - ATA stands for Advanced Technology Attachment.
 - The ATA disk is also called the IDE disk.
 - The ATA port uses the parallel ATA technology.



SATA Port

- SATA is short for serial ATA.
 - SATA ports use serial transmission and provide a higher rate than IDE ports.
 - SATA uses a point-to-point architecture and supports hot swap.



SATA Version	Line Code	Transfer Rate	Throughput
1.0	8b/10b	1.5 Gbit/s	150 MB/s
2.0	8b/10b	3 Gbit/s	300 MB/s
3.0	8b/10b	6 Gbit/s	600 MB/s

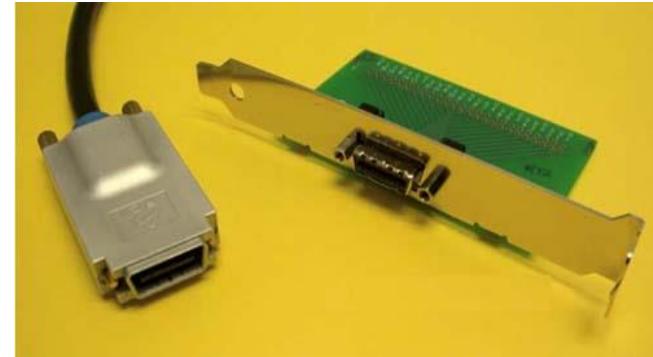
SCSI Port

- SCSI is short for Small Computer System Interface.



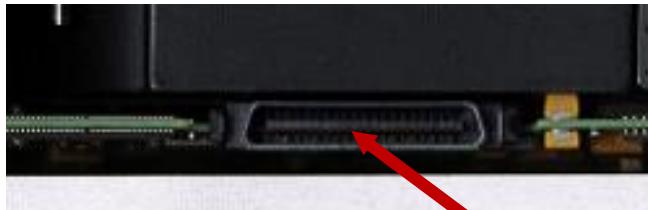
SAS Port

- SAS is short for Serial Attached SCSI.
 - SAS is a point-to-point, full-duplex, and dual-port interface.
 - SAS is backward compatible with SATA.
 - Rate: 600 Mbit/s per channel
 - SAS features high performance, high reliability, and powerful scalability.



Fibre Channel Port

- Fibre Channel disks use the Fibre Channel arbitrated loop (FC-AL).
 - FC-AL is a dual-port serial storage interface based on the SCSI protocol.
 - FC-AL supports full-duplex mode.
 - Fibre Channel provides a universal hardware transmission platform for upper-layer protocols (SCSI and IP). It is a serial data transmission interface that features high speed, high reliability, low latency, and high throughput.



40-pin Male FC-SCA II Connector



Contents

2. Intelligent Data Storage Components

- Controller enclosure
- Disk enclosure
- Expansion module
- **Disk**
 - HDD
 - **SSD**
- Interface module

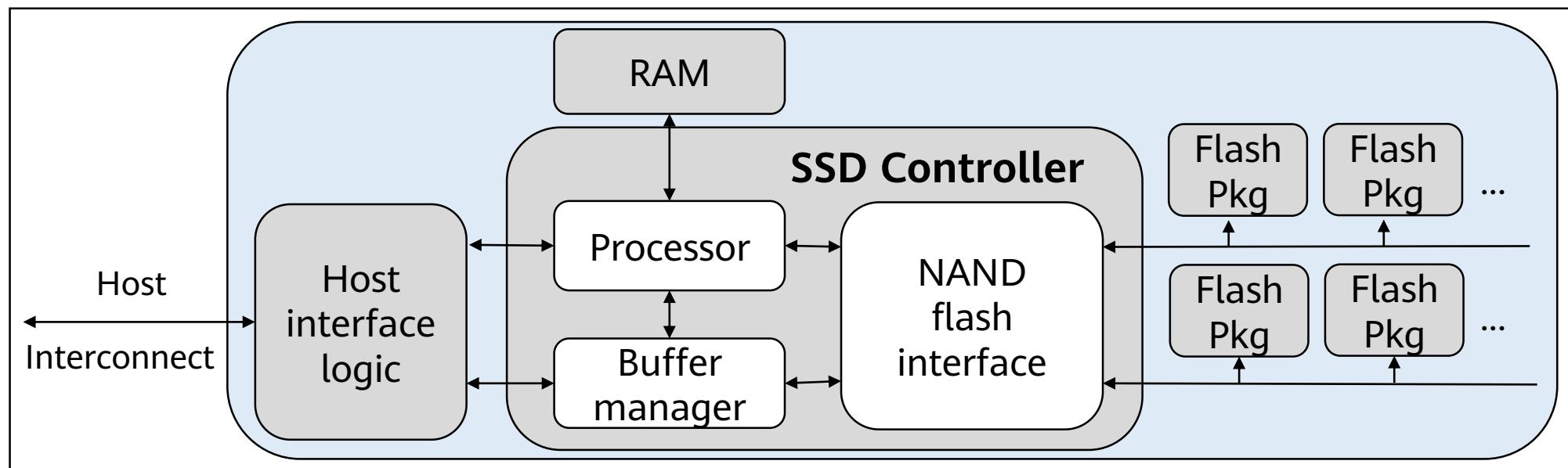
SSD Overview

- Compared to HDDs, SSDs have absolute advantages in terms of performance, reliability, power consumption, and portability. SSDs have been widely used in various industries.
- SSD characteristics:
 - Uses NAND flash to save data, providing a faster speed than HDDs.
 - Has no mechanical structure inside, so it consumes less power, dissipates less heat, and generates less noise.
 - Its service life is determined by the number of program/erase (P/E) cycles.



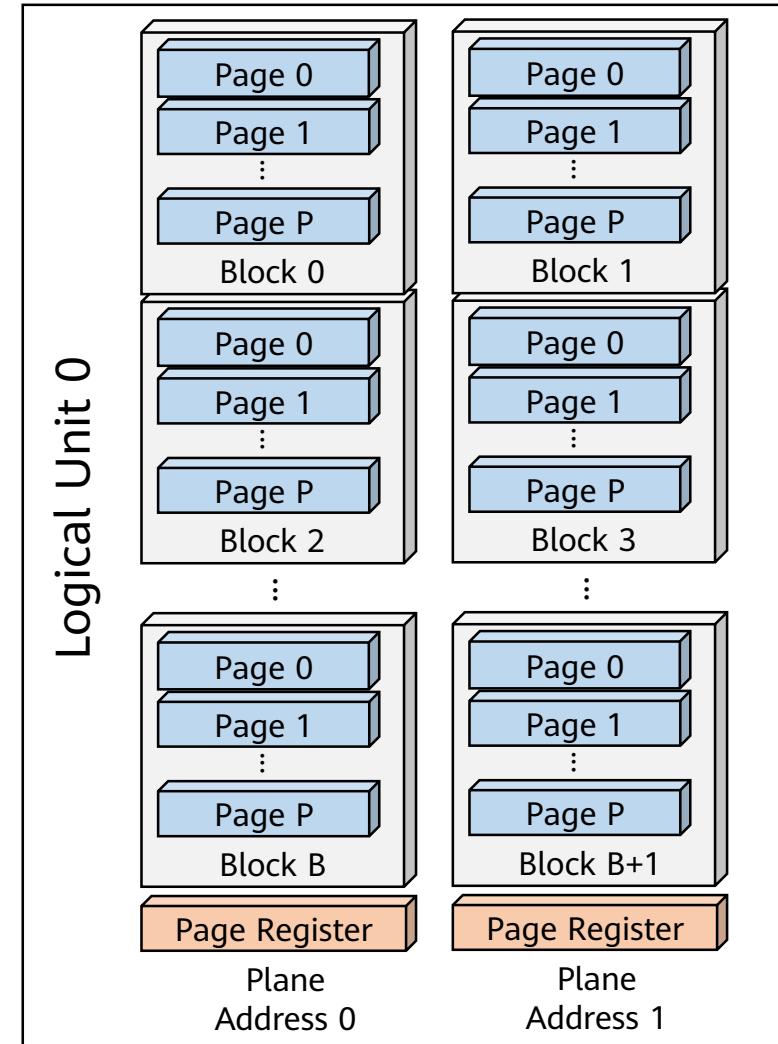
SSD Architecture

- An SSD consists of a control unit and a storage unit (mainly flash memory chips).
 - Control unit: SSD controller, host interface, and DRAM
 - Storage unit: NAND flash

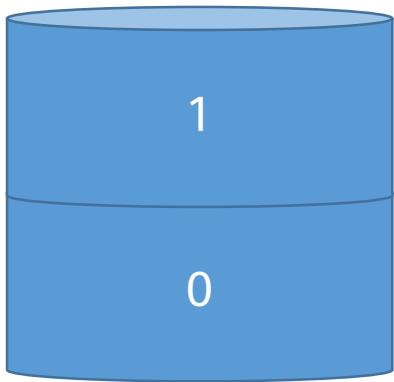


NAND Flash

- Internal storage units in NAND flash include:
 - LUNs, planes, blocks, pages, and cells
- Operations on the NAND flash include erase, program, and read.
- NAND flash is a non-volatile medium. A block must be erased before new data is written to it. A program/erase (P/E) cycle is the process of erasing a block and then writing it again.

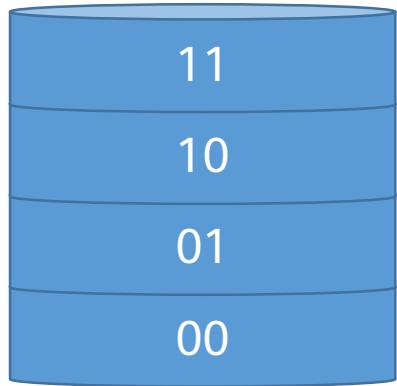


SLC, MLC, TLC, and QLC



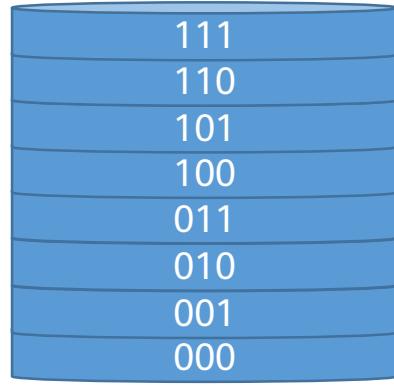
SLC-1bit

- SLC
- 1. Supports 50,000 to 100,000 P/E cycles, providing the best reliability.
 - 2. The storage capacity is small.
 - 3. The cost is the highest.



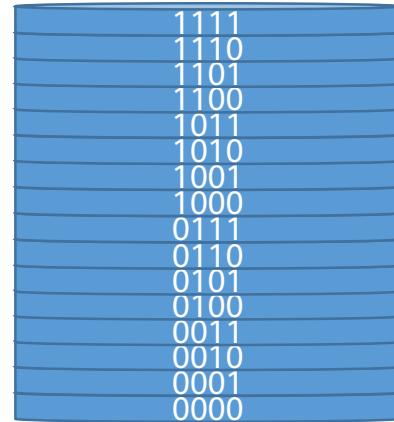
MLC-2bit

- MLC
- 1. Supports about 3,000 P/E cycles.
 - 2. The speed is slower than that of SLC.
 - 3. The storage capacity is relatively large.
 - 4. The price is relatively low.



TLC-3bit

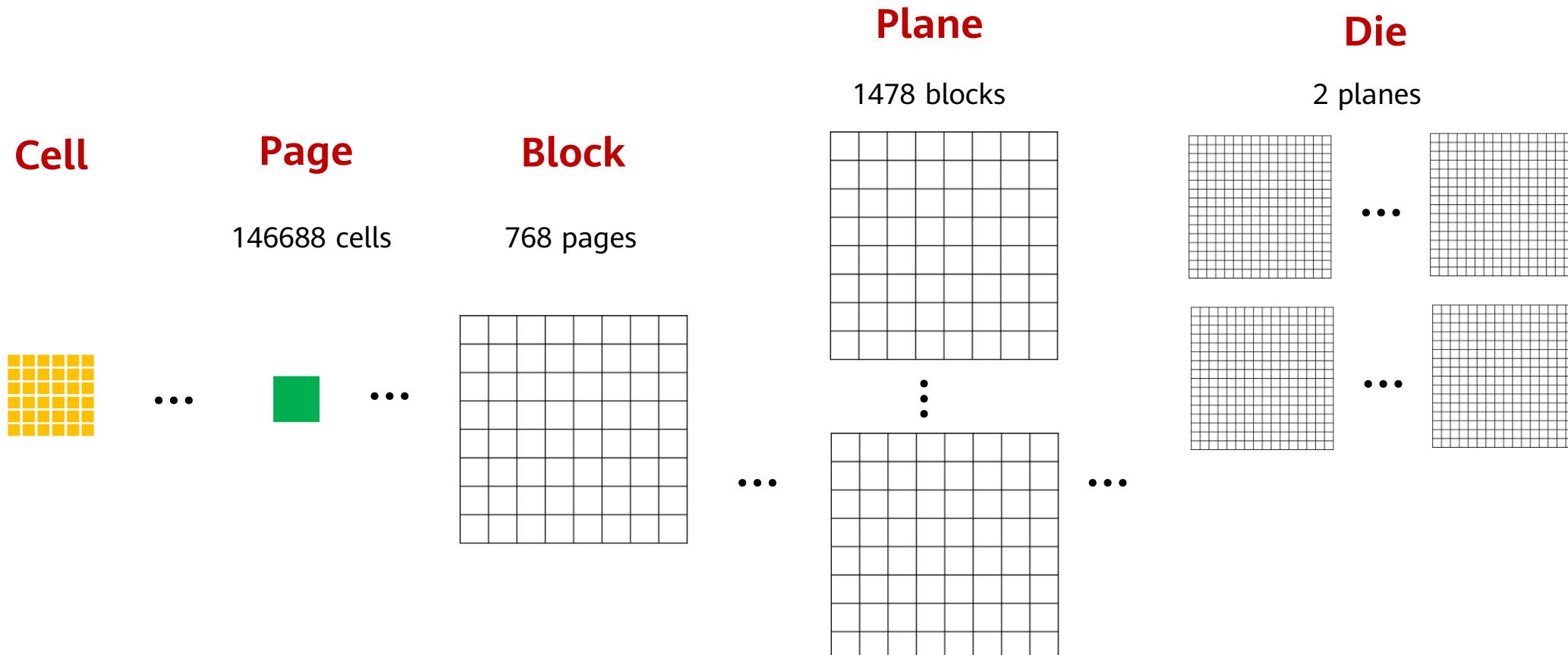
- TLC
- 1. Provides higher data density and supports only several hundred to 1,000 P/E cycles.
 - 2. The reliability and performance are low.
 - 3. Generally used in personal devices due to the cost advantage, but cannot meet the requirements of enterprise products.



QLC-4bit

- QLC
- 1. The capacity is further improved by 33%.
 - 2. The performance and life cycle are further reduced.

Flash Chip Data Relationship



Address Mapping Management

Logical block address (LBA) → No. 26, XX Road, Binjiang District, Hangzhou City, Zhejiang Province, People's Republic of China

Physical block address (PBA) → 120° 12' east longitude, 30° 16' north latitude

HDD: The relationship between LBA and PBA is fixed.

- Overwrite

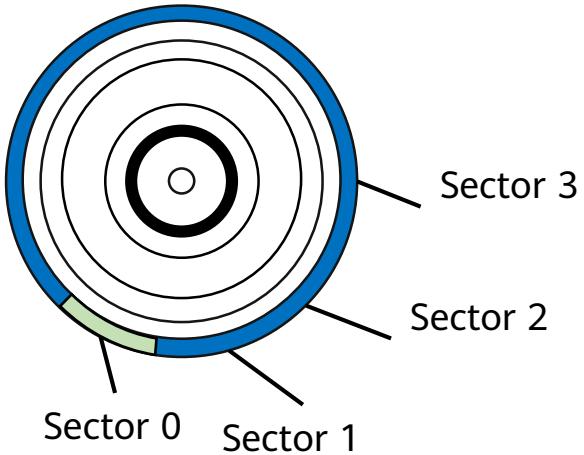


The Flash Translation Layer (FTL) is responsible for the conversion between the LBA and PBA.

SSD: The relationship between LBA and PBA is not fixed.

- Non-overwrite: A block must be erased before new data is written to it. New data and old data are at different locations.

FTL



OS sector (512 bytes). File systems read/write data in the unit of 512 bytes.

Main controller
FTL mapping table: saved in the internal SRAM/DRAM, external DRAM, or NAND flash.

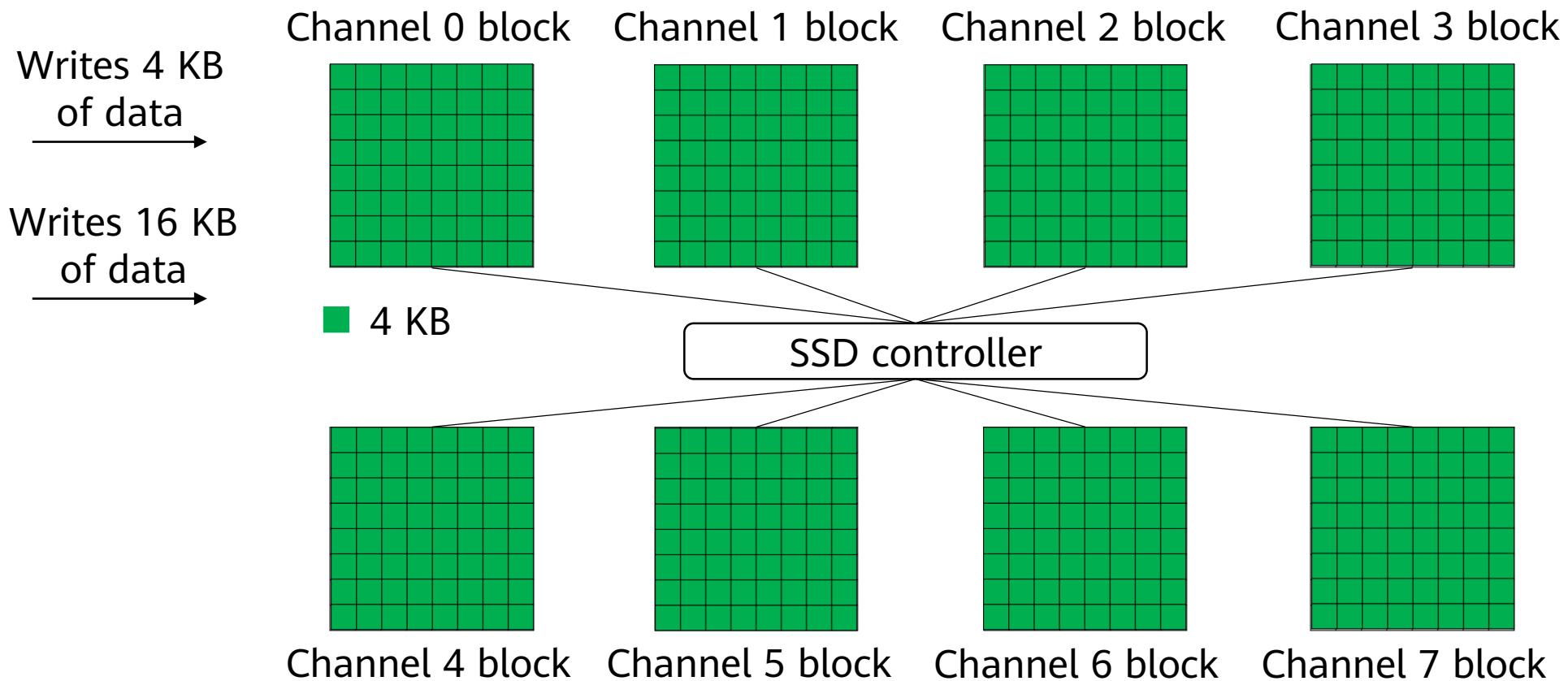
Sector 0			
			Sector 2
	Sector 1		
		Sector 4	
Sector 3			Sector 5

FTL mapping operation. The main controller maps the addresses based on the mapping table.

Data is stored in the NAND physical addresses based on the mapping table.

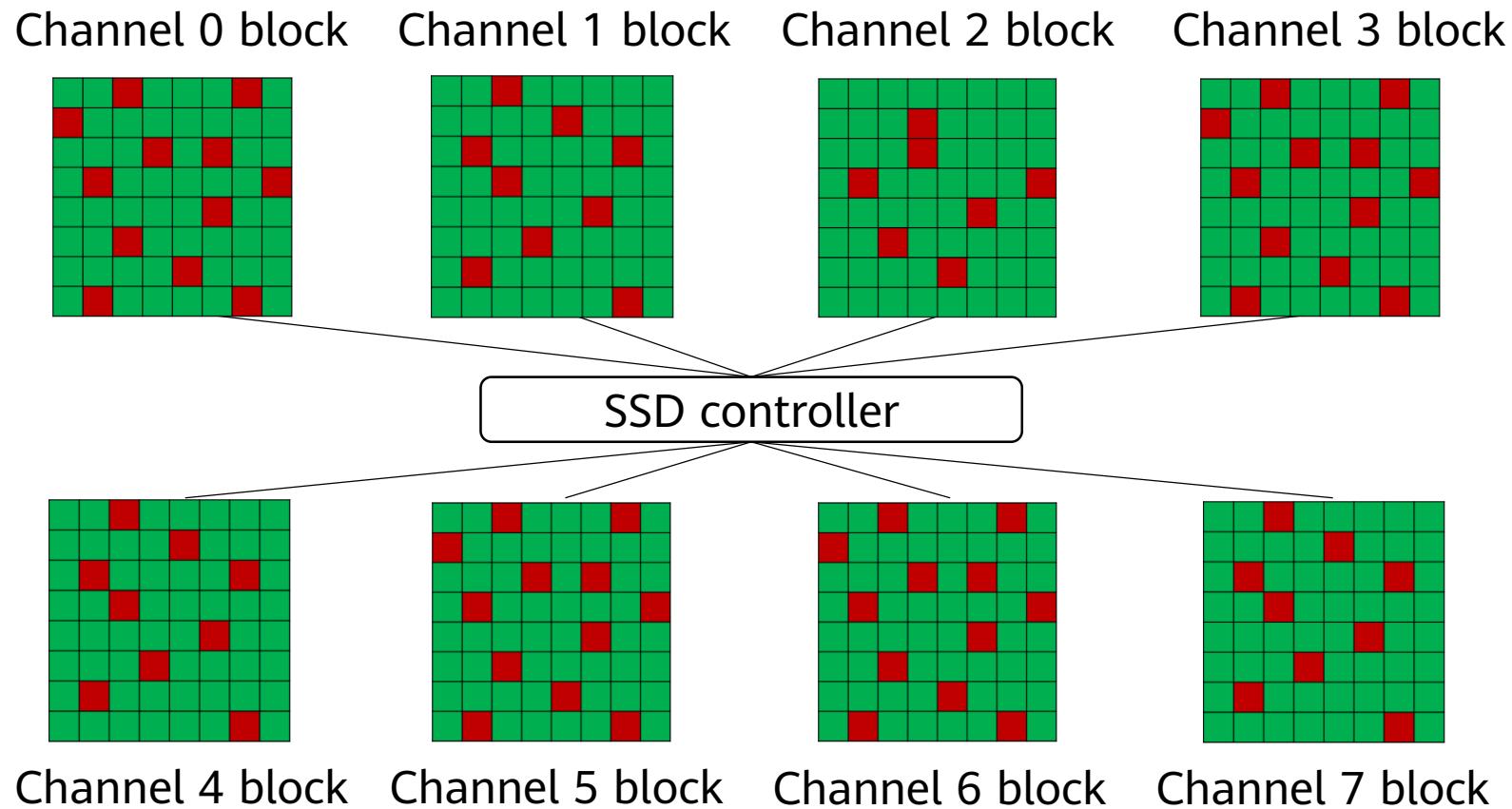
Data Write Process on an SSD (1)

- The following uses eight channels as an example to demonstrate how the host writes data to the SSD.

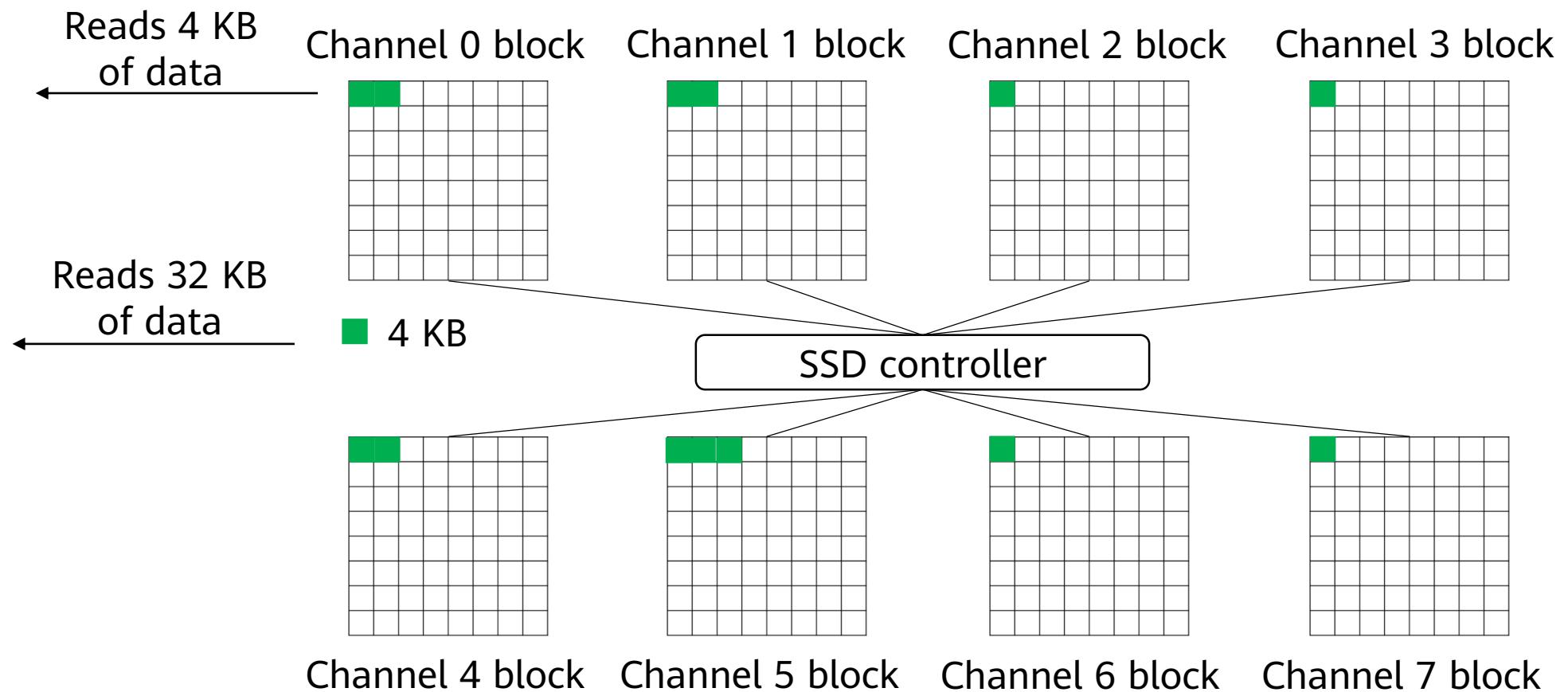


Data Write Process on an SSD (2)

- When the SSD is full, old data must be deleted to release space for new data. When a user deletes and writes data, data in some blocks becomes invalid or aged.

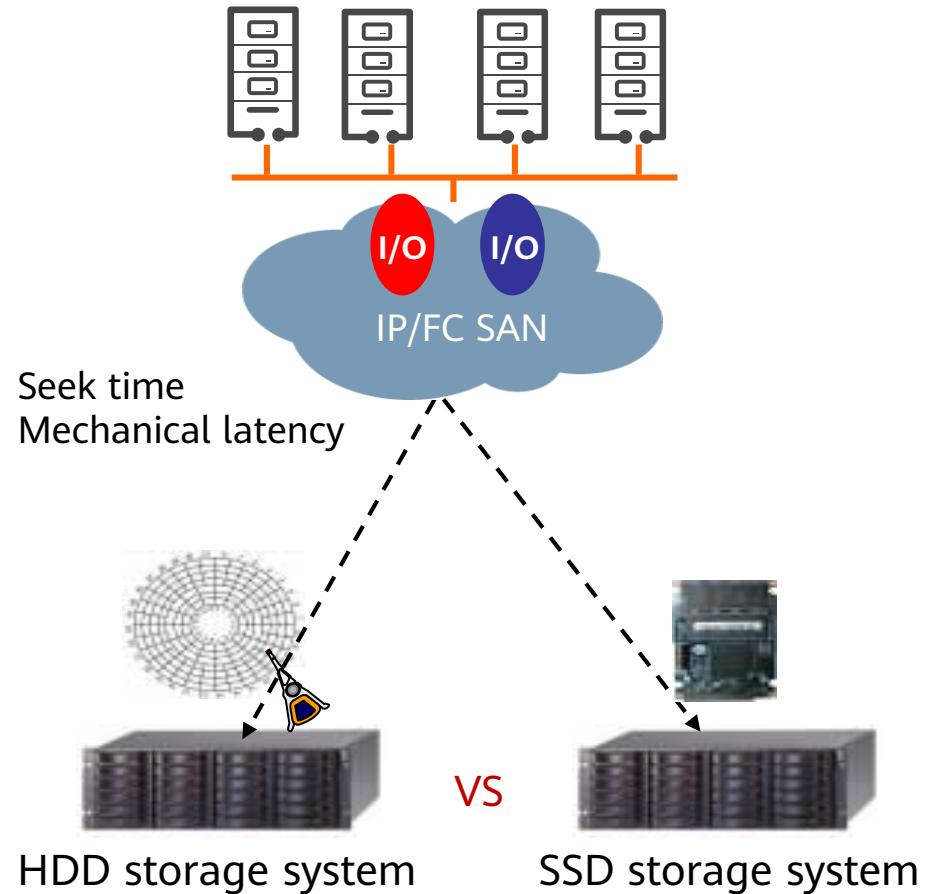


Data Read Process on an SSD

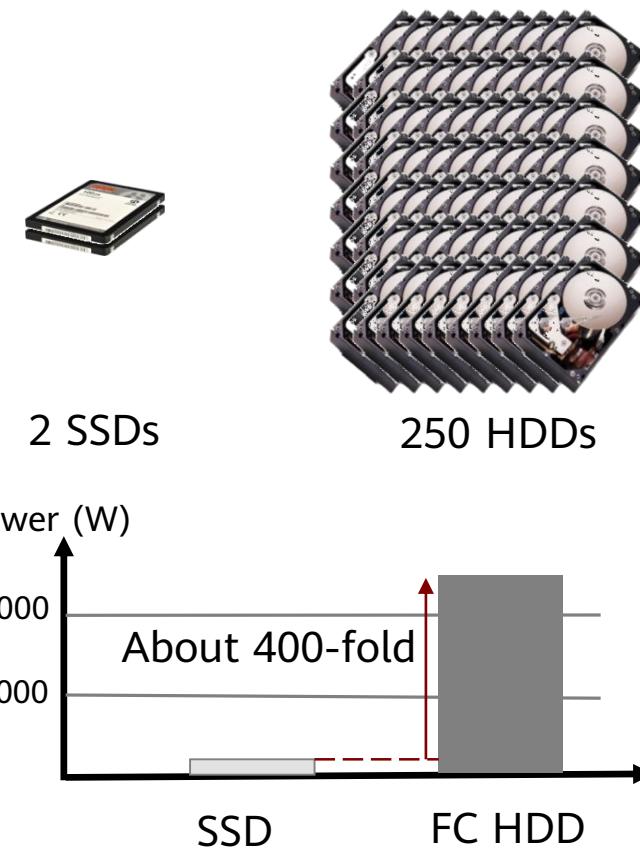


SSD Performance Advantages

SSD Performance Advantages

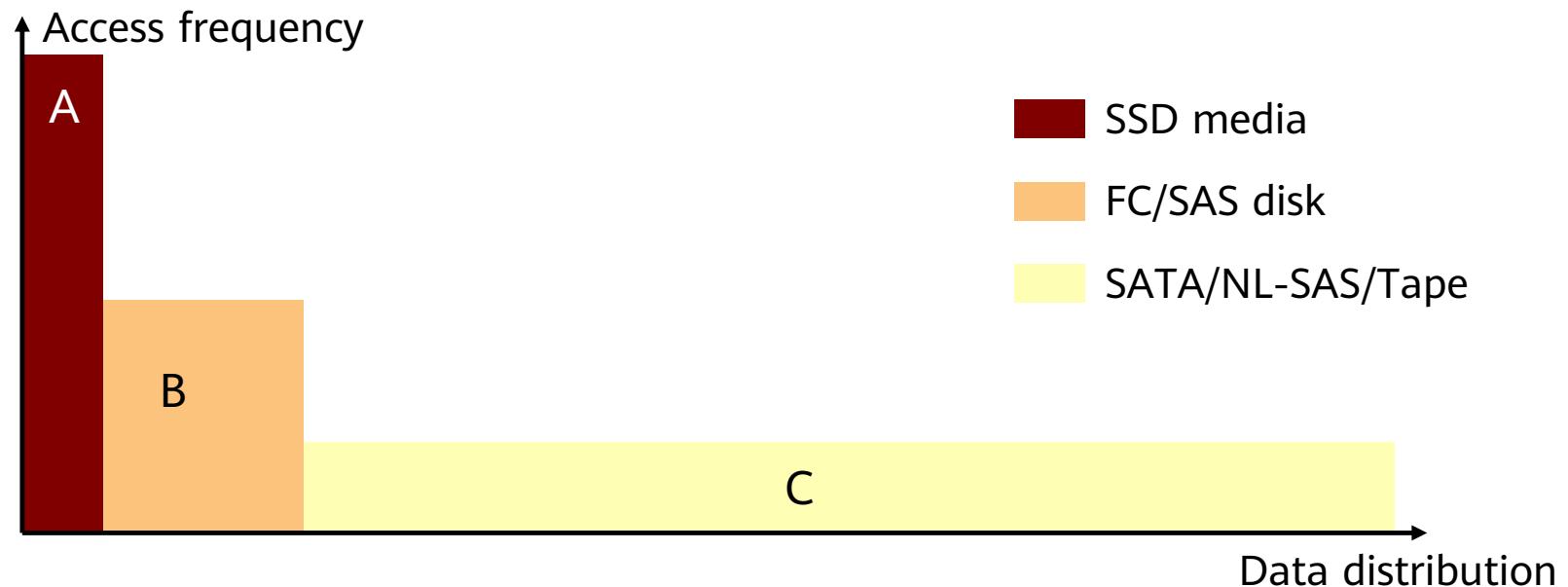


Power consumption under 100,000 read IOPS



Use of SSDs in Storage Systems

- Class A applications: high-concurrency applications featuring random reads and writes, such as databases
- Class B applications: large files, images, and streaming media featuring sequential reads and writes
- Class C applications: data backup or rarely used applications



SCM Card

- Storage class memory (SCM) is a new class of non-volatile memory. It is slightly slower than memory but much faster than NAND in terms of the access speed.
- An SCM card is a cache acceleration card of the SCM media type. To use SmartCache for OceanStor Dorado V6 all-flash storage (6.1.0 and later versions), install an SCM card on the controller enclosure.

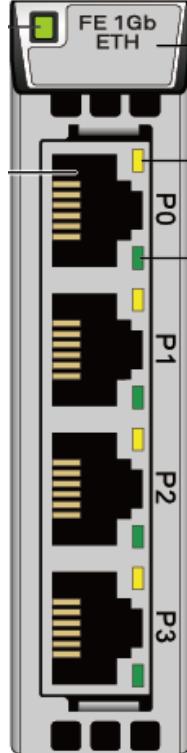


Contents

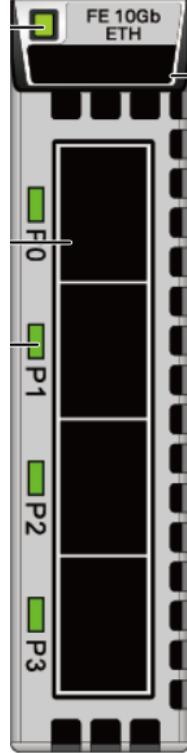
2. Intelligent Data Storage Components

- Controller enclosure
- Disk enclosure
- Expansion module
- Disk
- Interface module

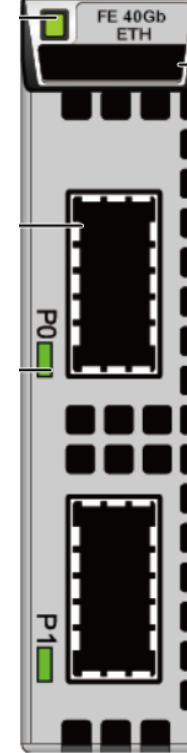
Front-End: GE Interface Modules



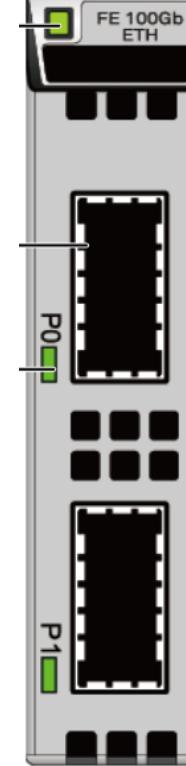
GE electrical
interface module



10GE electrical
interface module

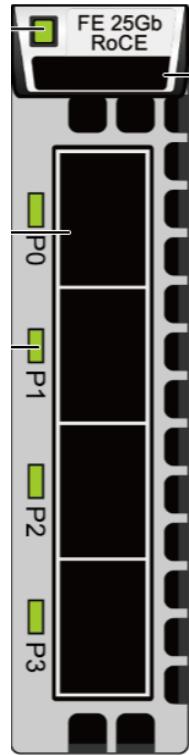


40GE interface
module

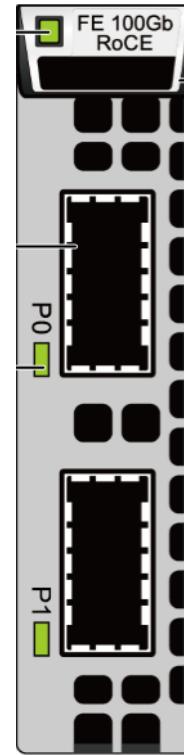


100GE interface
module

Front-End: RoCE Interface Modules

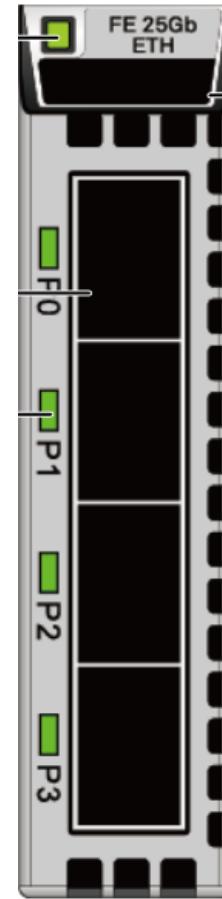
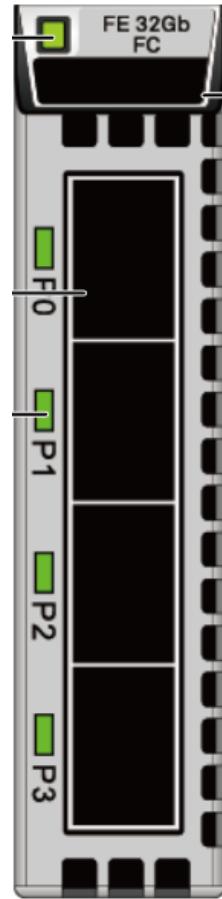


25 Gbit/s RoCE interface module

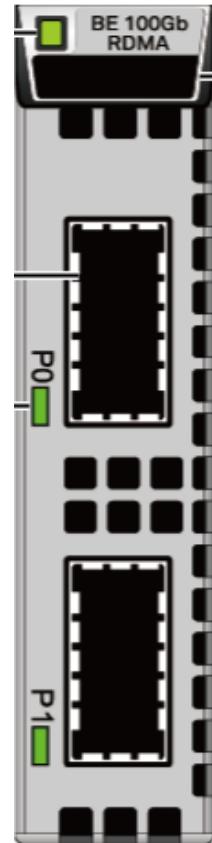


100 Gbit/s RoCE interface module

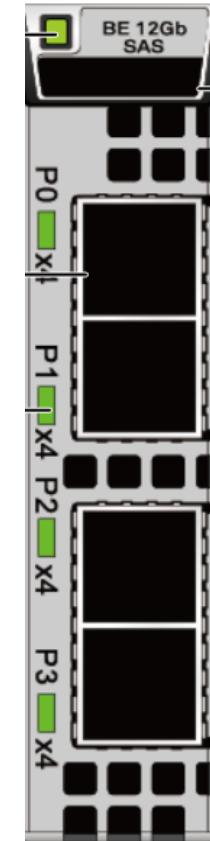
Front-End: SmartIO Interface Modules



Back-End: 100 Gbit/s RDMA Interface Module and 12 Gbit/s SAS Expansion Module

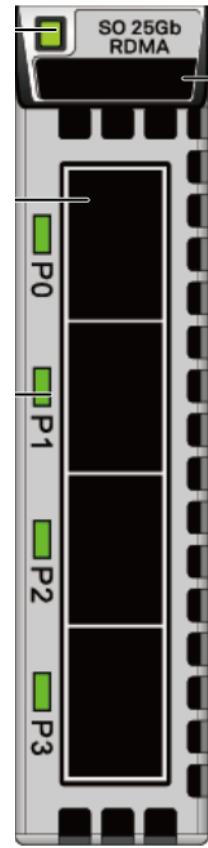


100 Gbit/s RDMA interface module

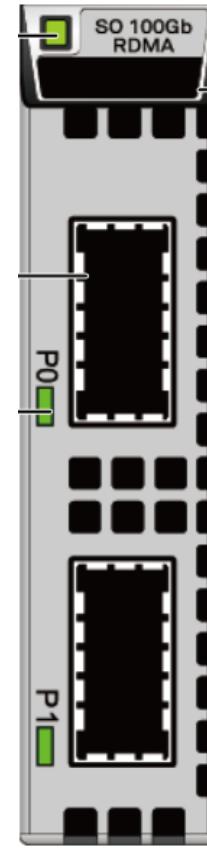


12 Gbit/s SAS expansion module

Scale-out: 100 Gbit/s RDMA Interface Module and 25 Gbit/s RDMA Interface Module



25 Gbit/s RDMA interface module



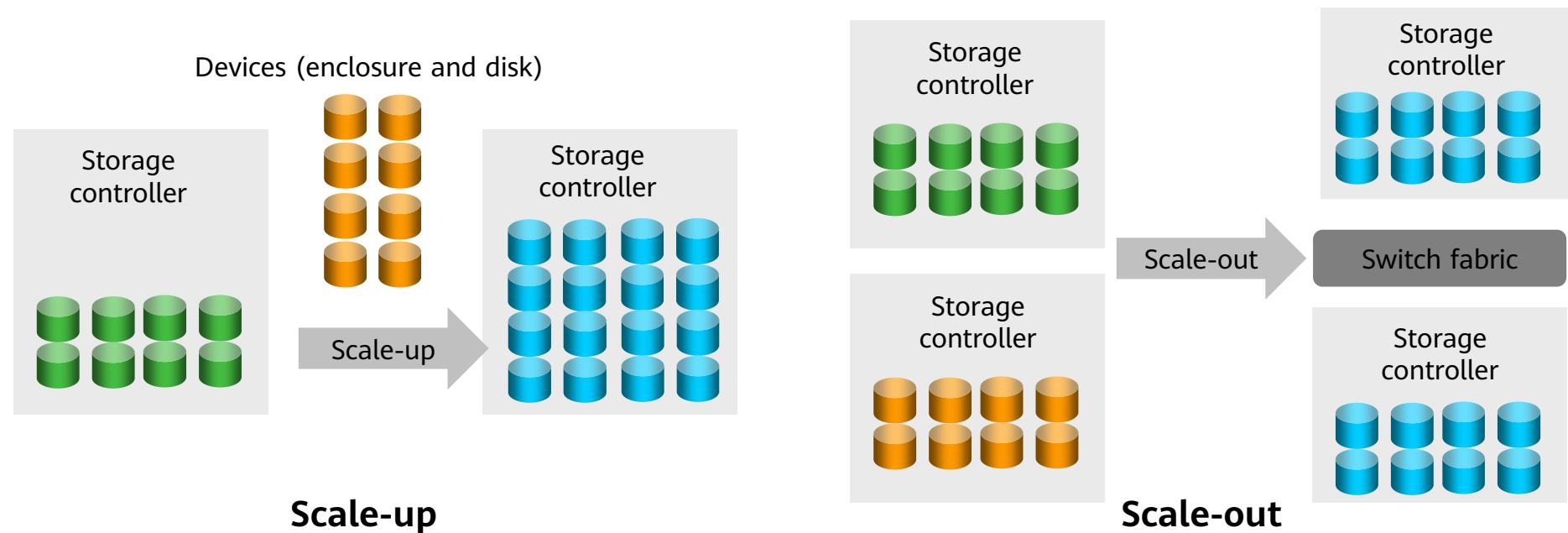
100 Gbit/s RDMA interface module

Contents

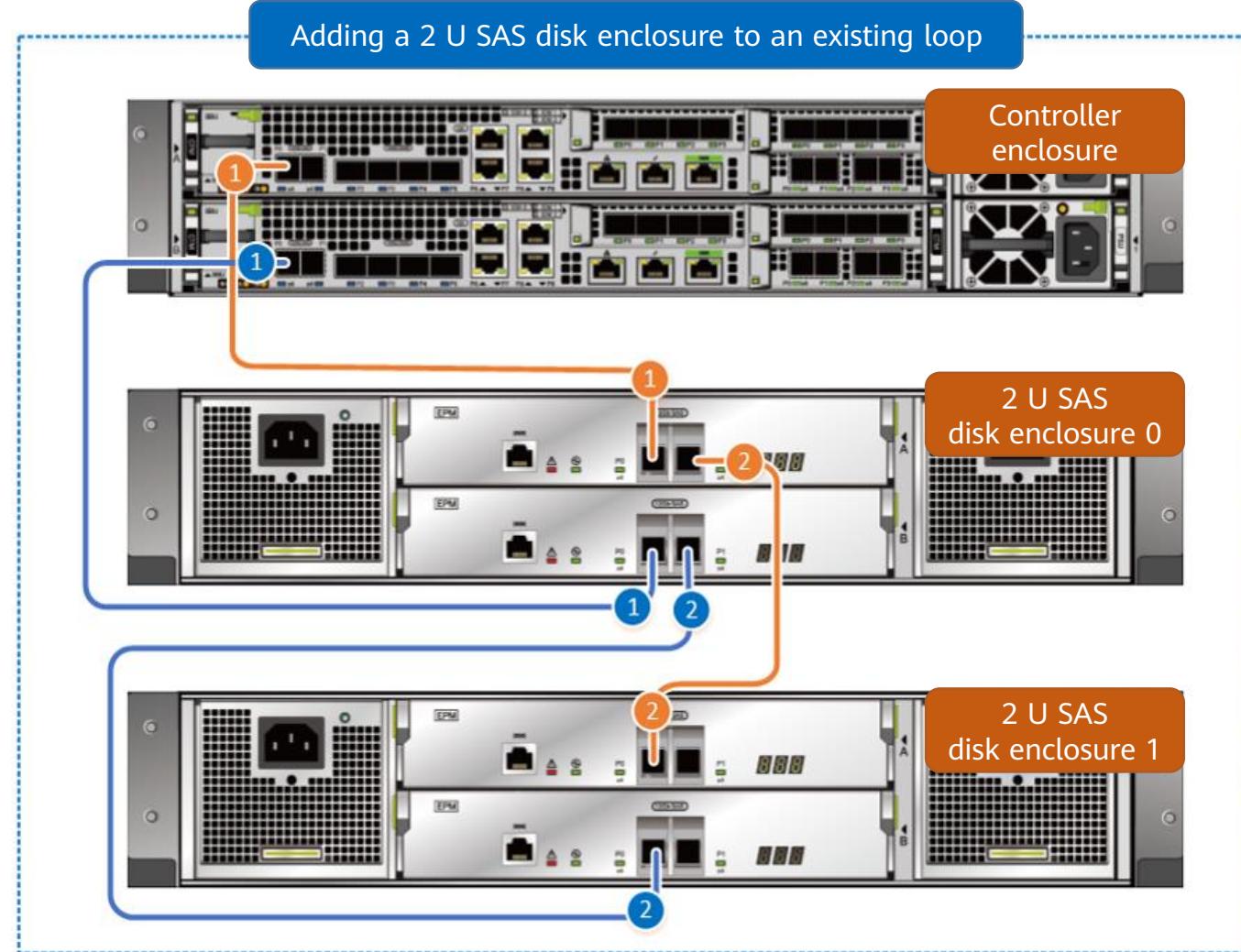
1. Intelligent Data Storage System
2. Intelligent Data Storage Components
- 3. Storage System Expansion Methods**

Scale-up and Scale-out

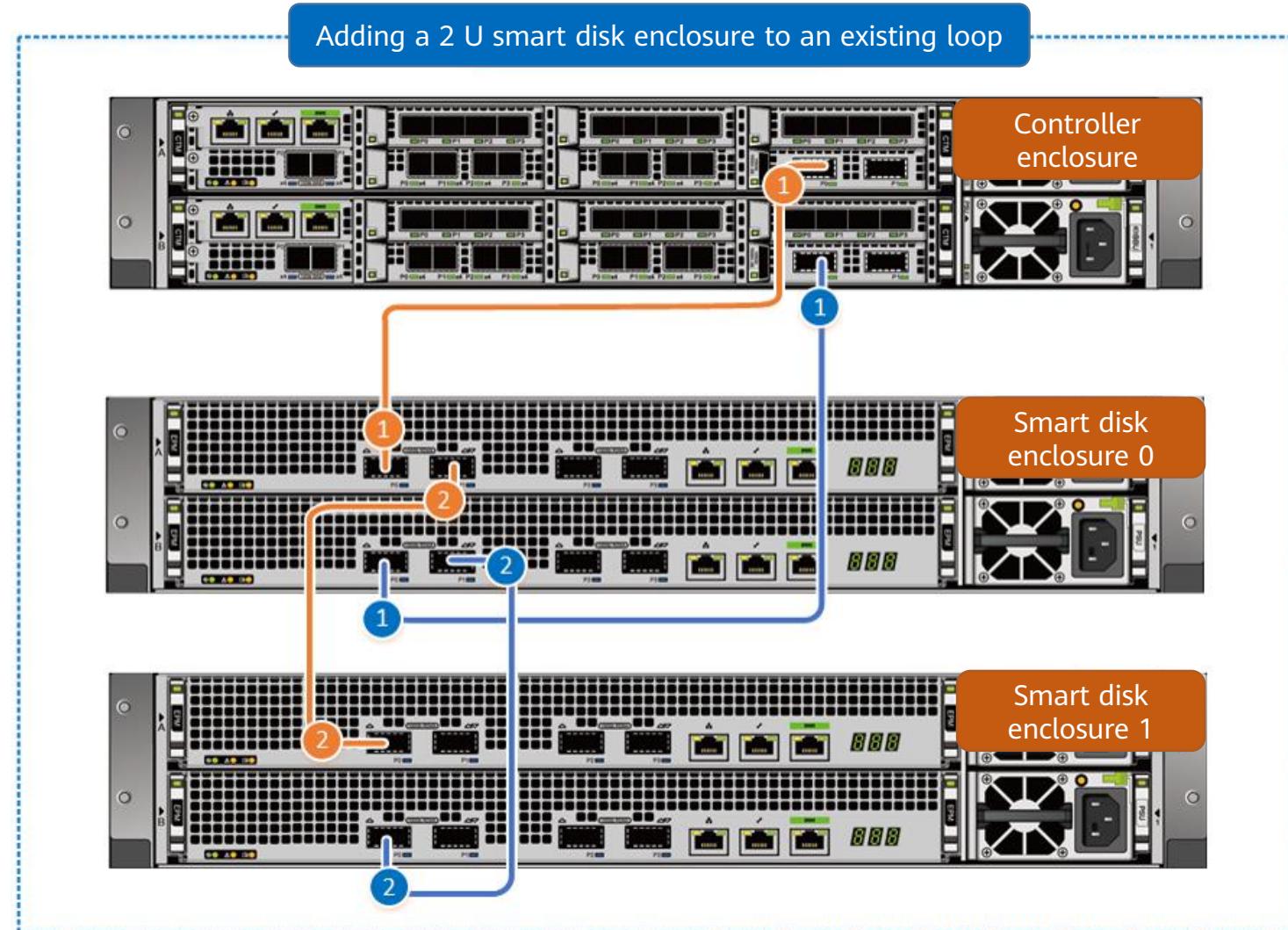
- Developments in enterprise IT systems and service expansion have caused data to skyrocket. The initial configuration of storage systems cannot meet the demands, and now capacity expansion is a top concern for system administrators. There are two capacity expansion methods: scale-up and scale-out. The following uses Huawei storage products as an example to describe the two methods.



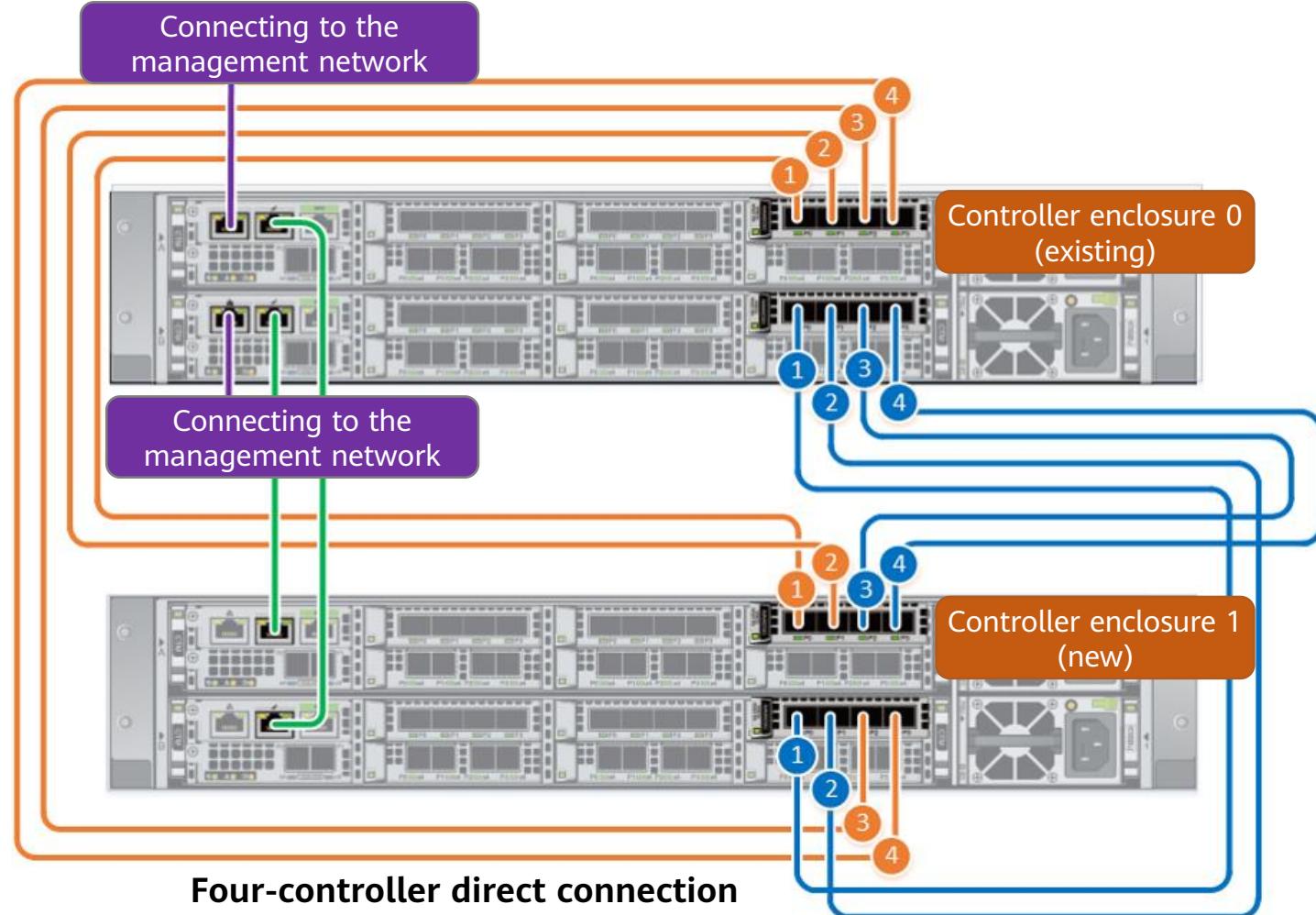
SAS Disk Enclosure Scale-up Networking Principles



Smart Disk Enclosure Scale-up Networking Principles



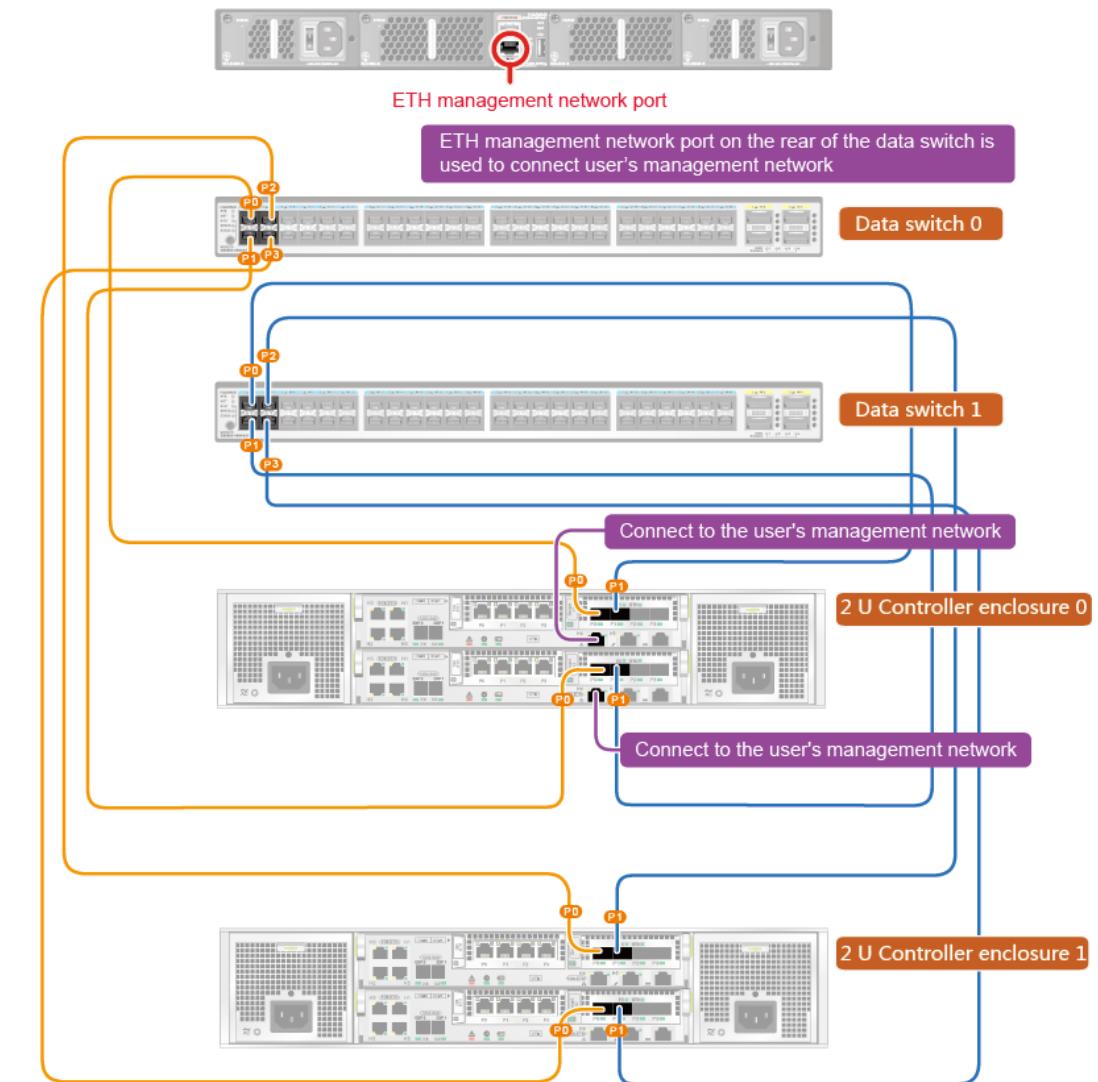
Scale-out Direct-Connection Networking



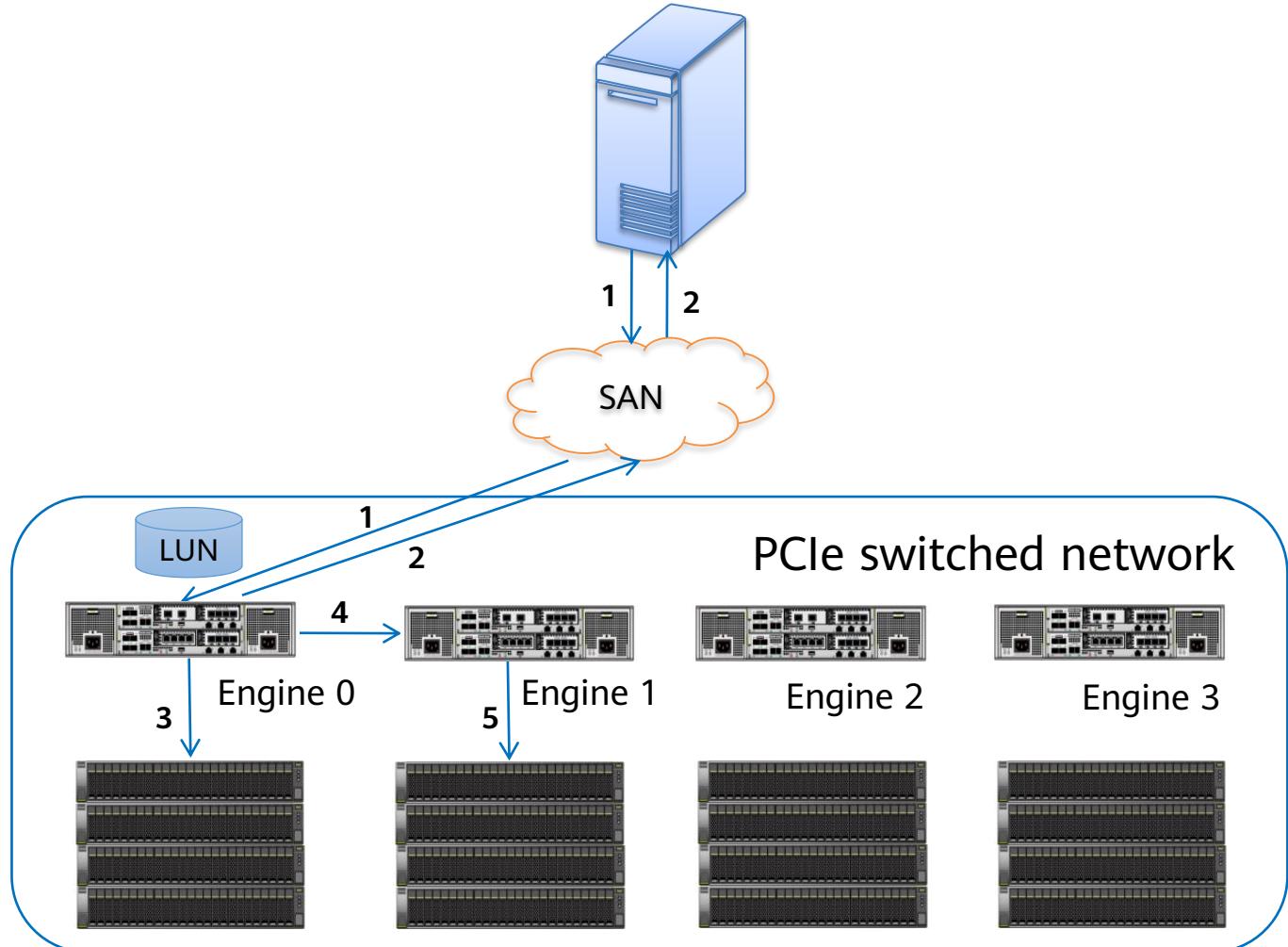
The above figure shows the scale-out networking of Huawei OceanStor Dorado 5000 V6 and 6000 V6.

Scale-out Switched Networking

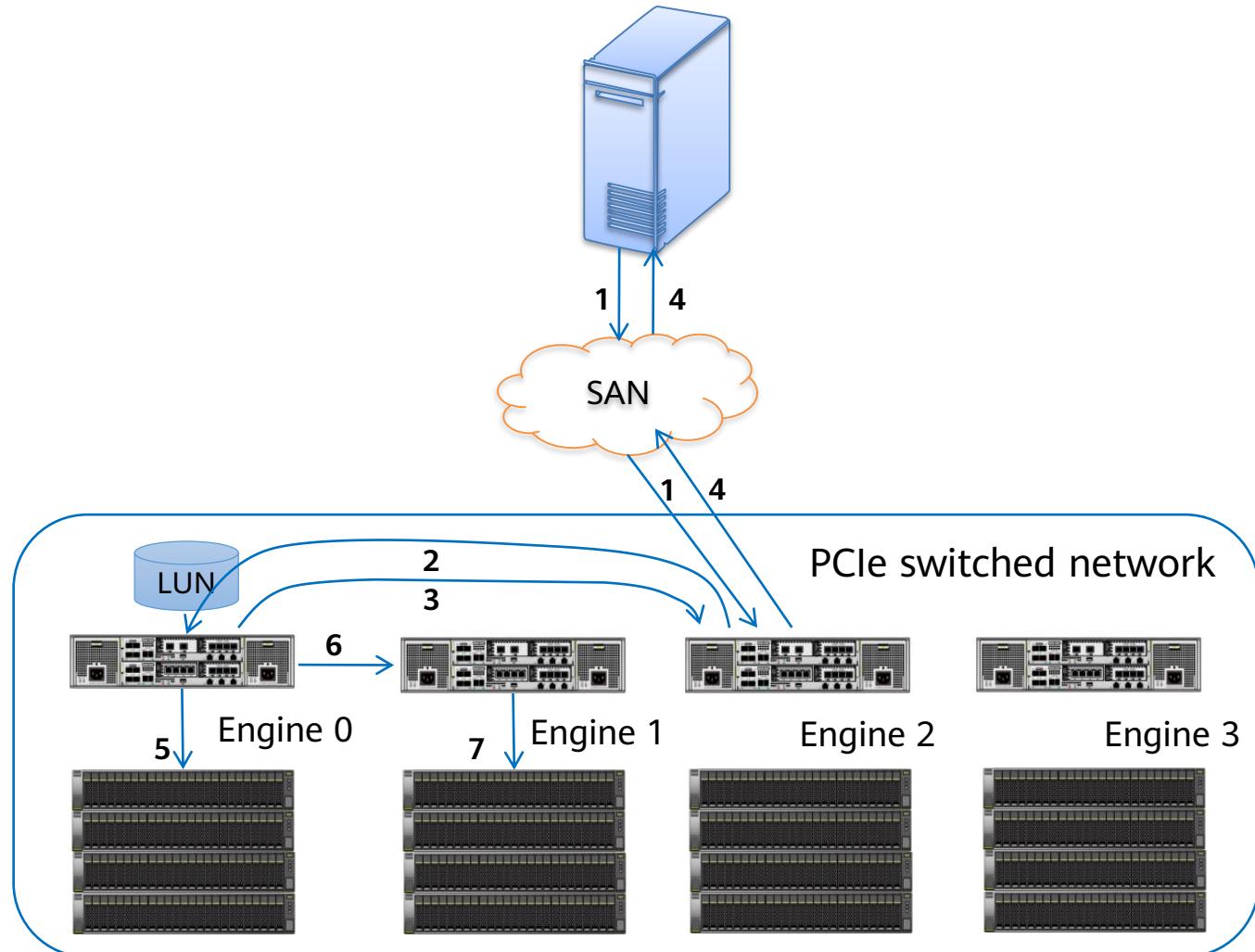
- The networking can be switching networking or switch-free networking.
- The management networks involve external and internal management networks.
- The internal management network and internal data network share the same physical network. For details, see the orange and blue cables in the figure. QoS is used, minimizing the impact on the management channel and data channel.
- The external management network only needs to enable the management network of two controllers in the first controller enclosure to connect to the user's management network. There is no need to connect controllers in other controller enclosures to the user's management network.
- The ETH management network ports in the rear of a CE6850 switch are used to connect to the user's network.



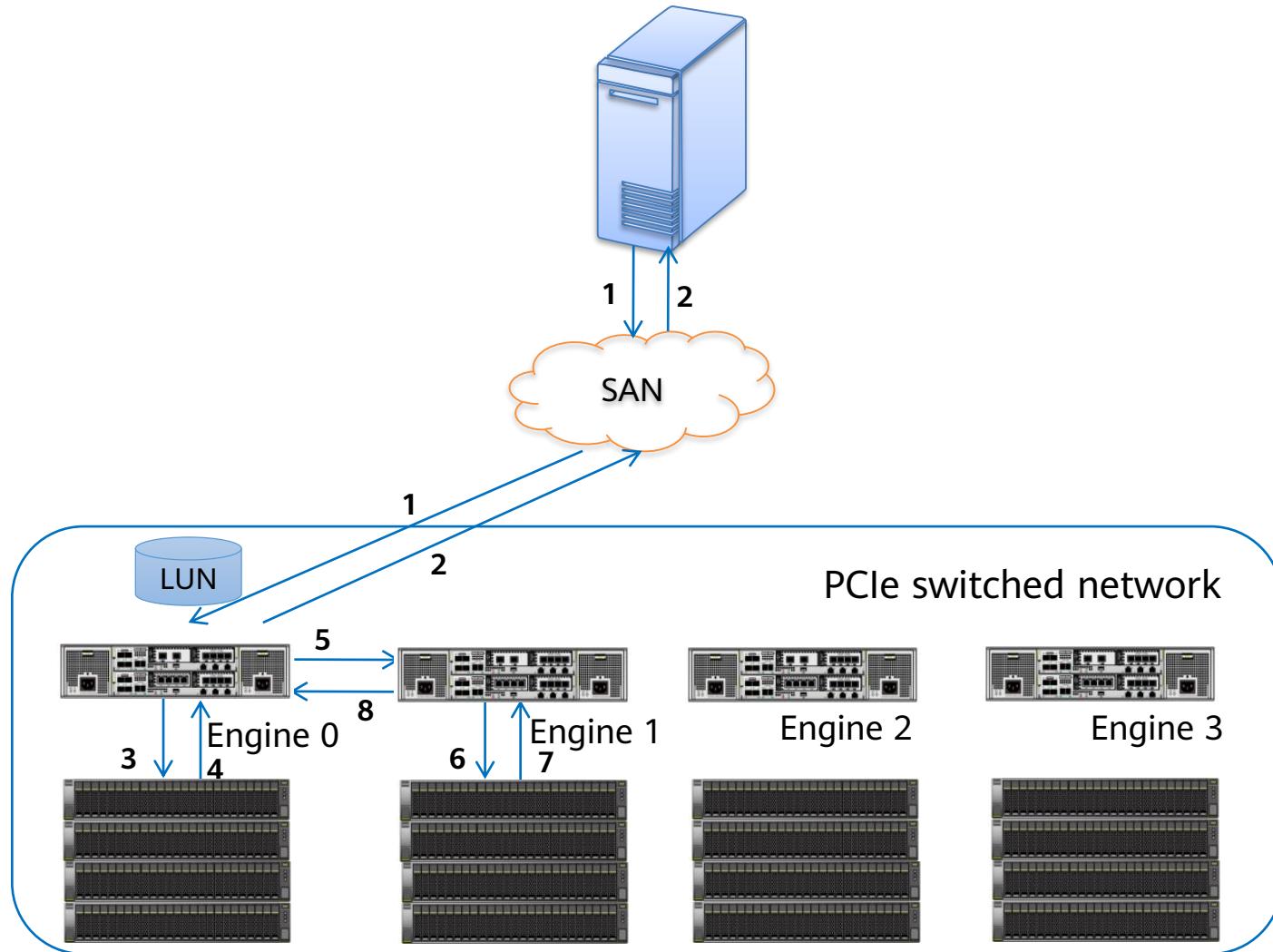
Local Write Process



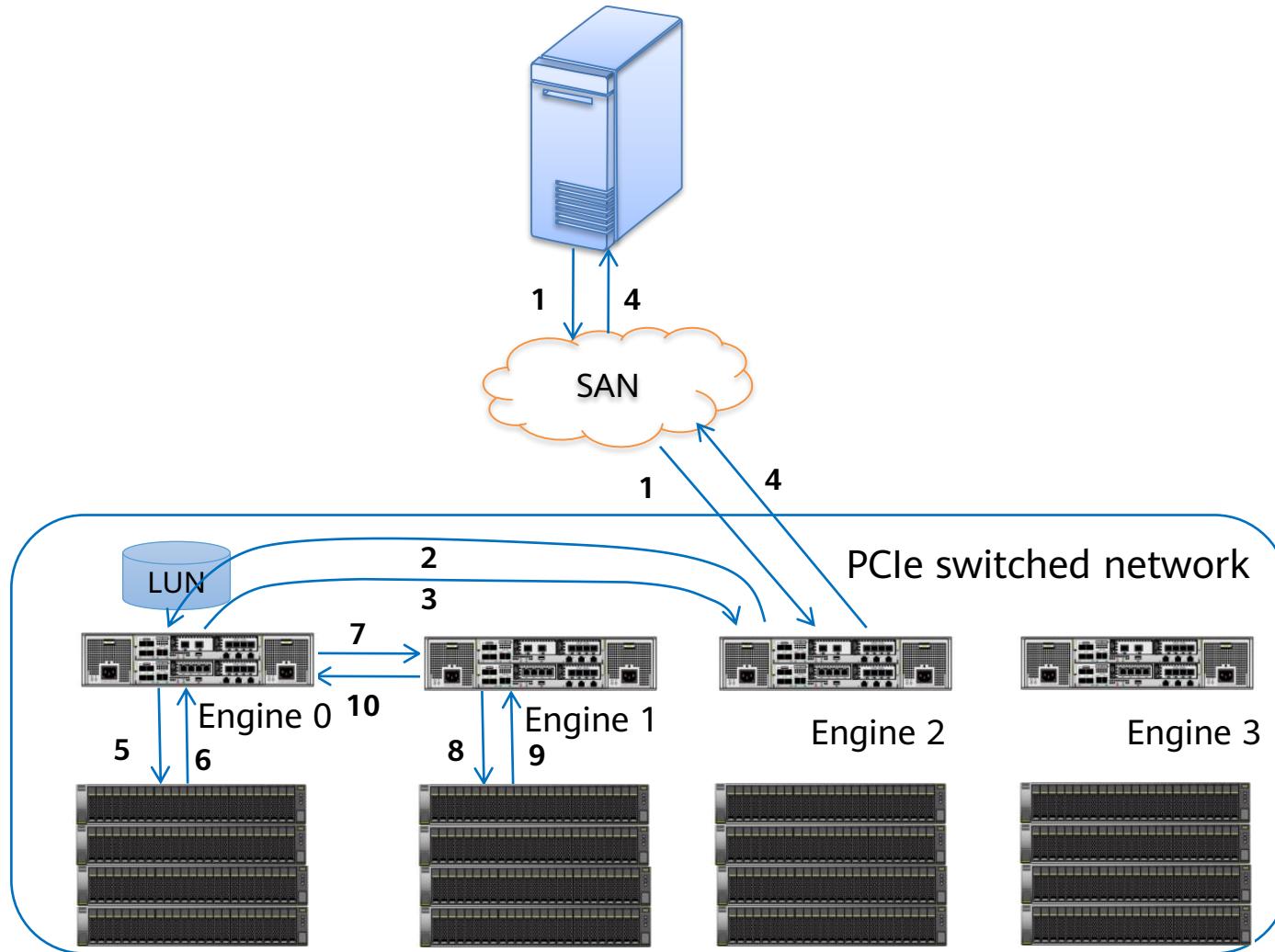
Non-local Write Process



Local Read Process



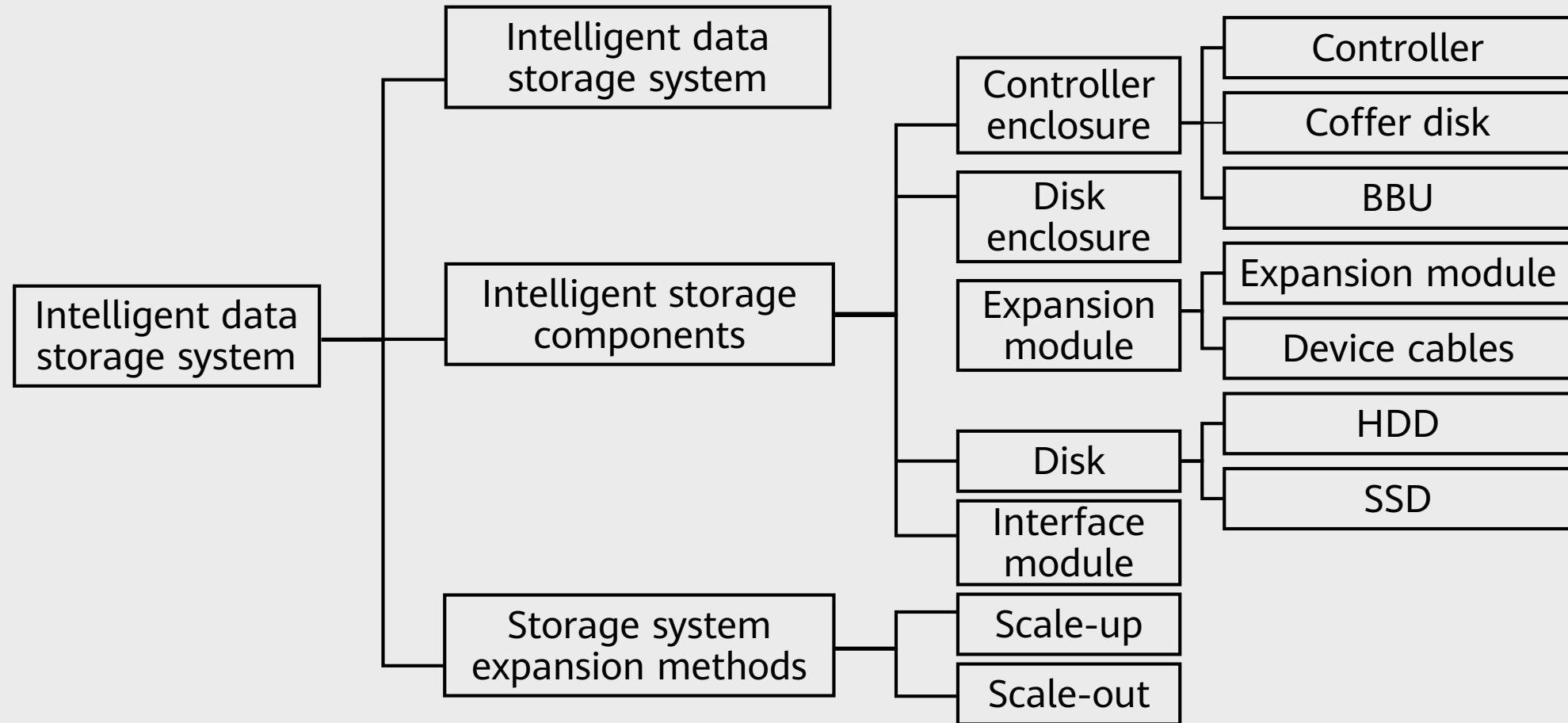
Non-local Read Process



Quiz

1. (Multiple-answer question) What are the types of SSDs?
 - A. SLC
 - B. MLC
 - C. TLC
 - D. QLC
2. (Multiple-answer question) Which of the following can be used to measure the performance of an HDD?
 - A. Disk capacity
 - B. Rotation speed
 - C. Data transfer rate
 - D. Average access time

Summary



Recommendations

- Huawei official websites
 - Enterprise business: <https://e.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/>
 - Online learning: <https://www.huawei.com/en/learning>
- Popular tools
 - HedEx Lite
 - Network Document Tool Center
 - Information Query Assistant

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.



RAID Technologies



Foreword

- This course introduces technologies of traditional RAID and RAID 2.0+. The evolution of RAID technologies aims at data protection and performance improvement.

Objectives

After completing this course, you will be able to understand:

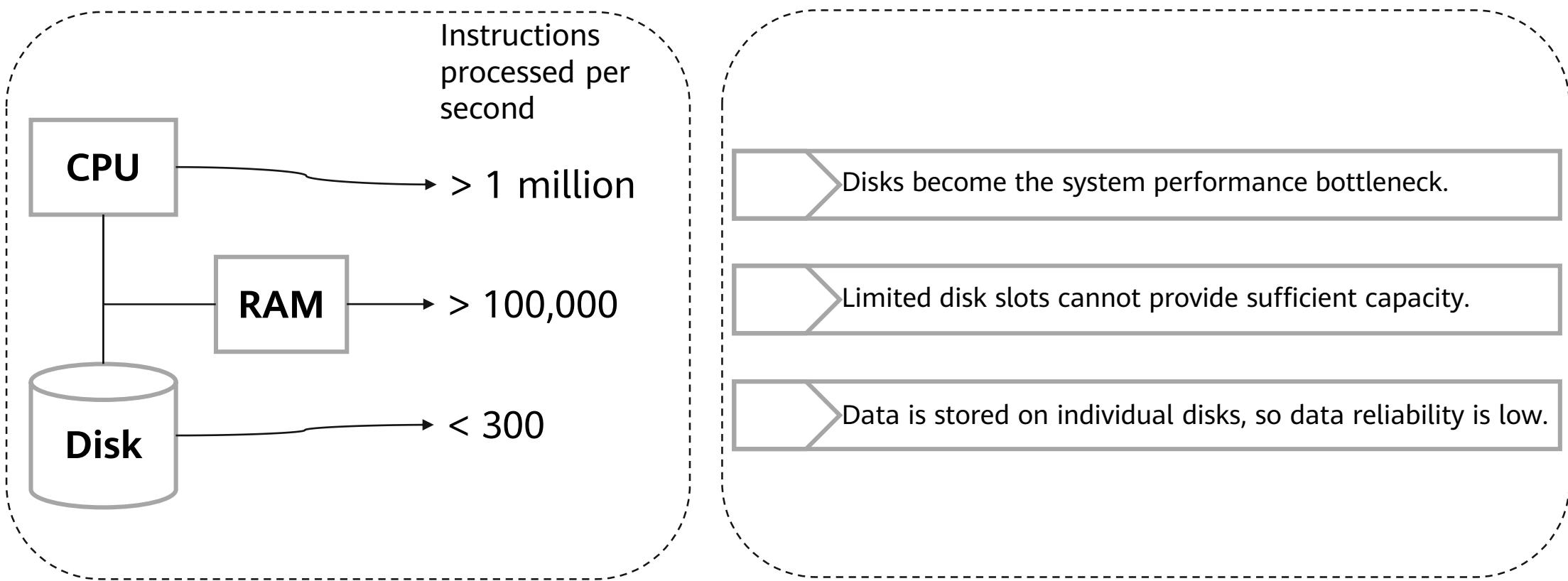
- Common RAID levels
- Different levels of data protection provided by different RAID levels
- Working principles of RAID 2.0+
- Dynamic RAID and RAID-TP

Contents

- 1. Traditional RAID**
2. RAID 2.0+
3. Other RAID Technologies

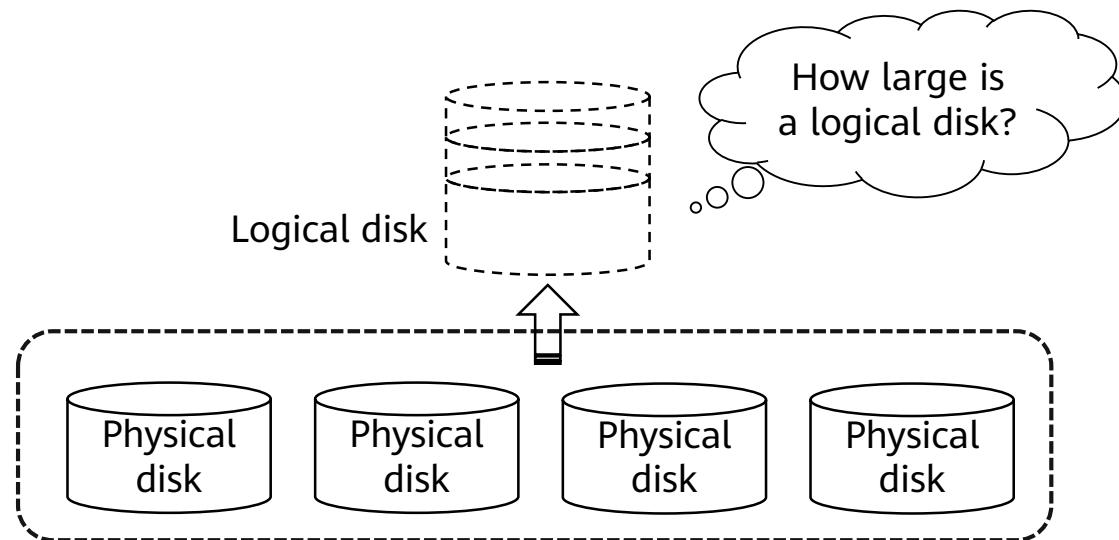
Background

- Problems in traditional computer systems must be addressed.



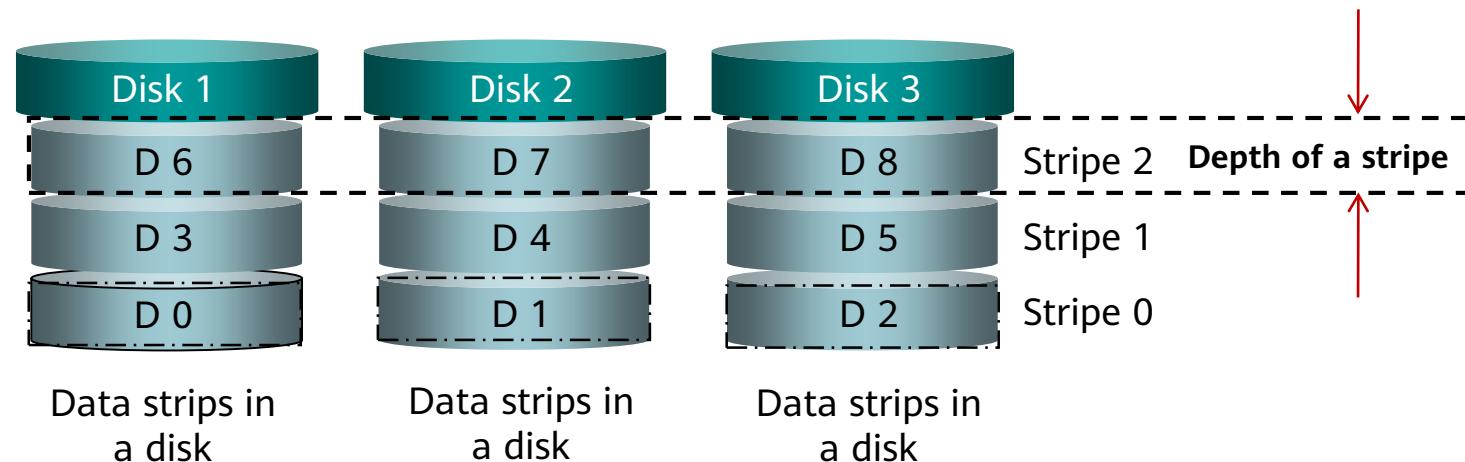
What Is RAID?

- Redundant Array of Independent Disks (RAID) combines multiple physical disks into one logical disk in different ways, improving read/write performance and data security.
- Implementations: hardware RAID and software RAID.



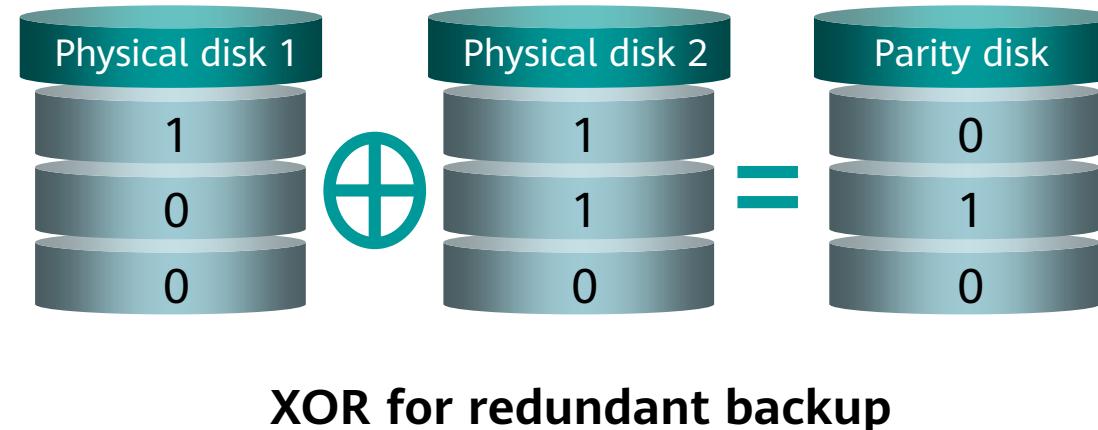
Data Organization Forms

- Disk striping: Space in each disk is divided into multiple strips of a specific size. Written data is also divided into blocks based on the strip size.
- Strip: A strip consists of one or more consecutive sectors in a disk, and multiple strips form a stripe.
- Stripe: A stripe consists of strips of the same location or ID on multiple disks in the same array.



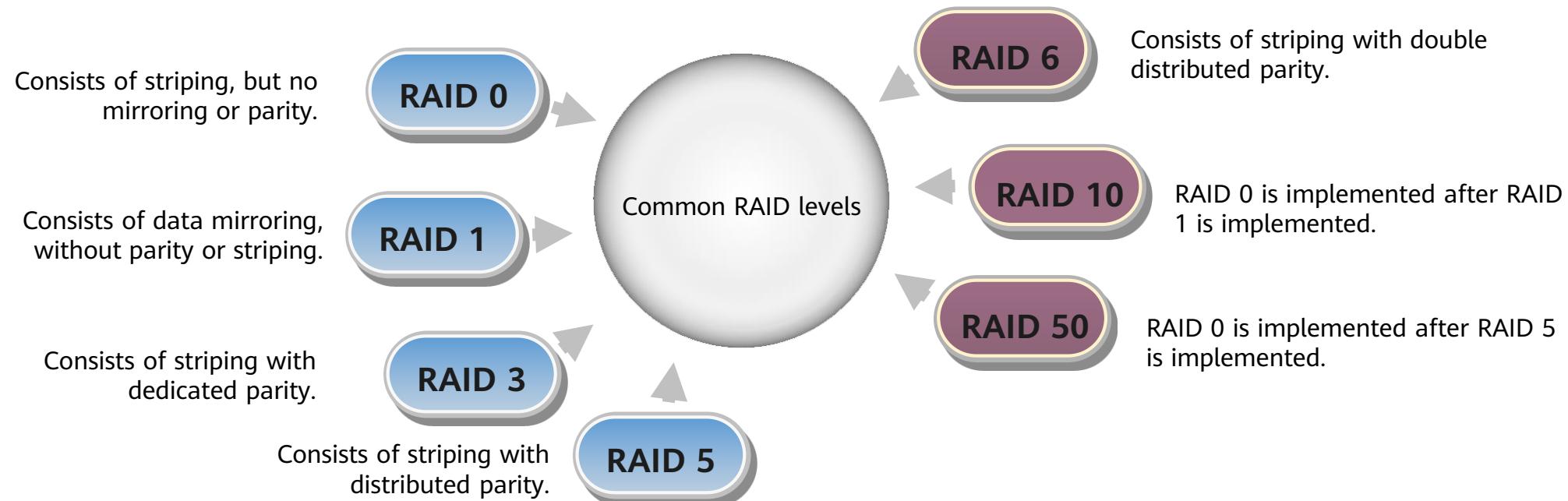
Data Protection Techniques

- Mirroring: Data copies are stored on another redundant disk.
- Exclusive or (XOR)
 - XOR is widely used in digital electronics and computer science.
 - XOR is a logical operation that outputs true only when inputs differ (one is true, the other is false).
 - $0 \oplus 0 = 0, 0 \oplus 1 = 1, 1 \oplus 0 = 1, 1 \oplus 1 = 0$

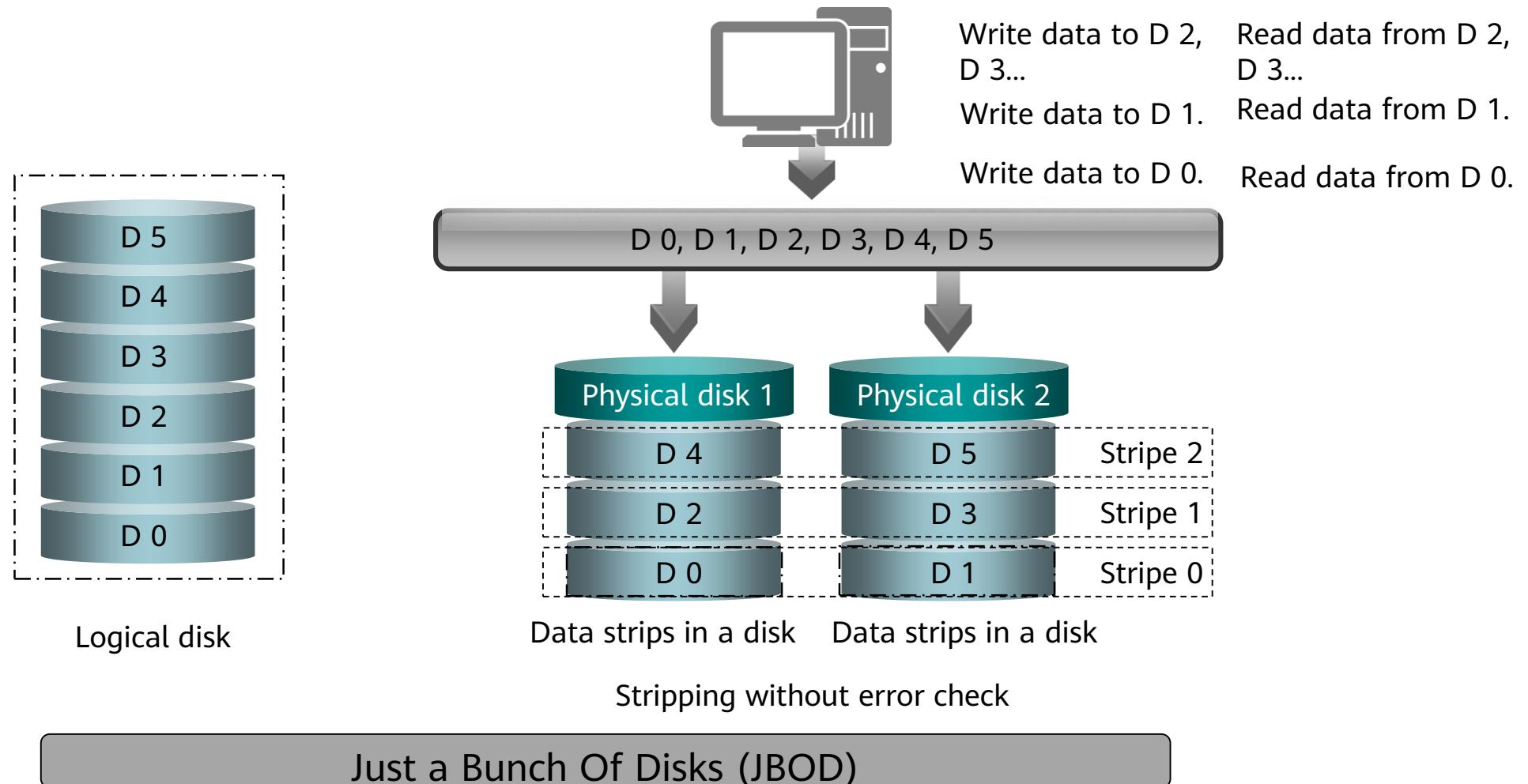


Common RAID Levels and Classification Criteria

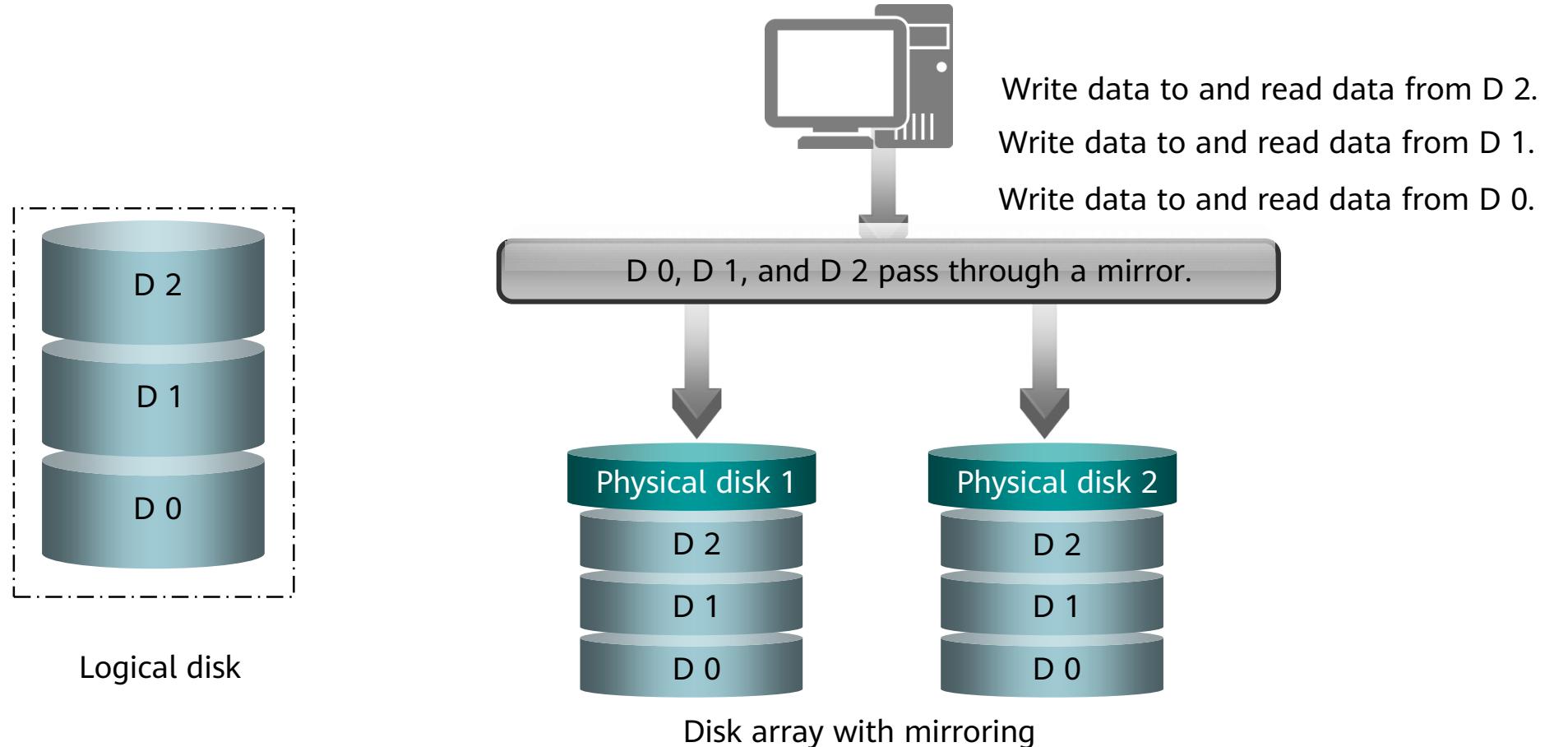
- RAID levels use different combinations of data organization forms and data protection techniques.



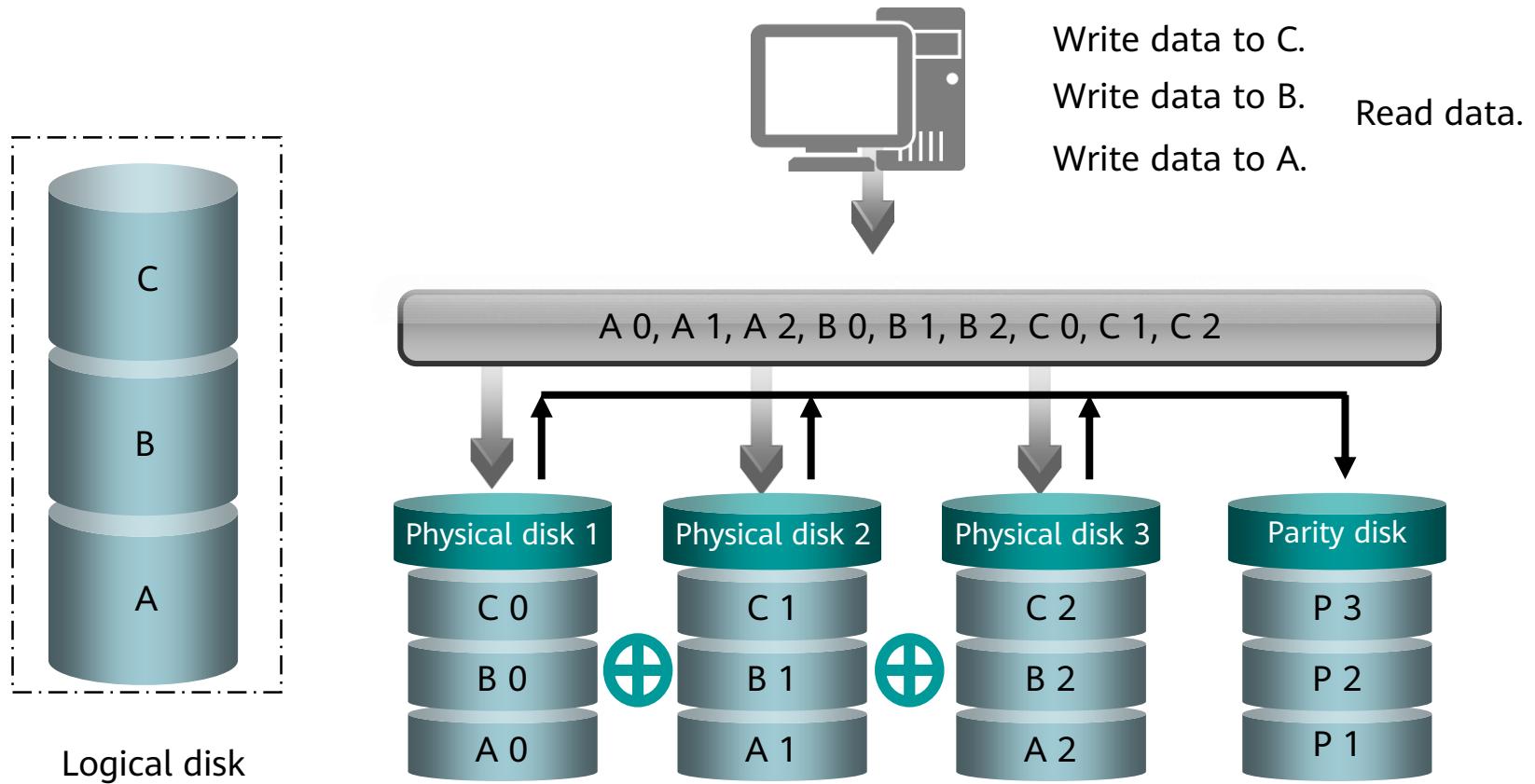
How Does RAID 0 Work?



How Does RAID 1 Work?

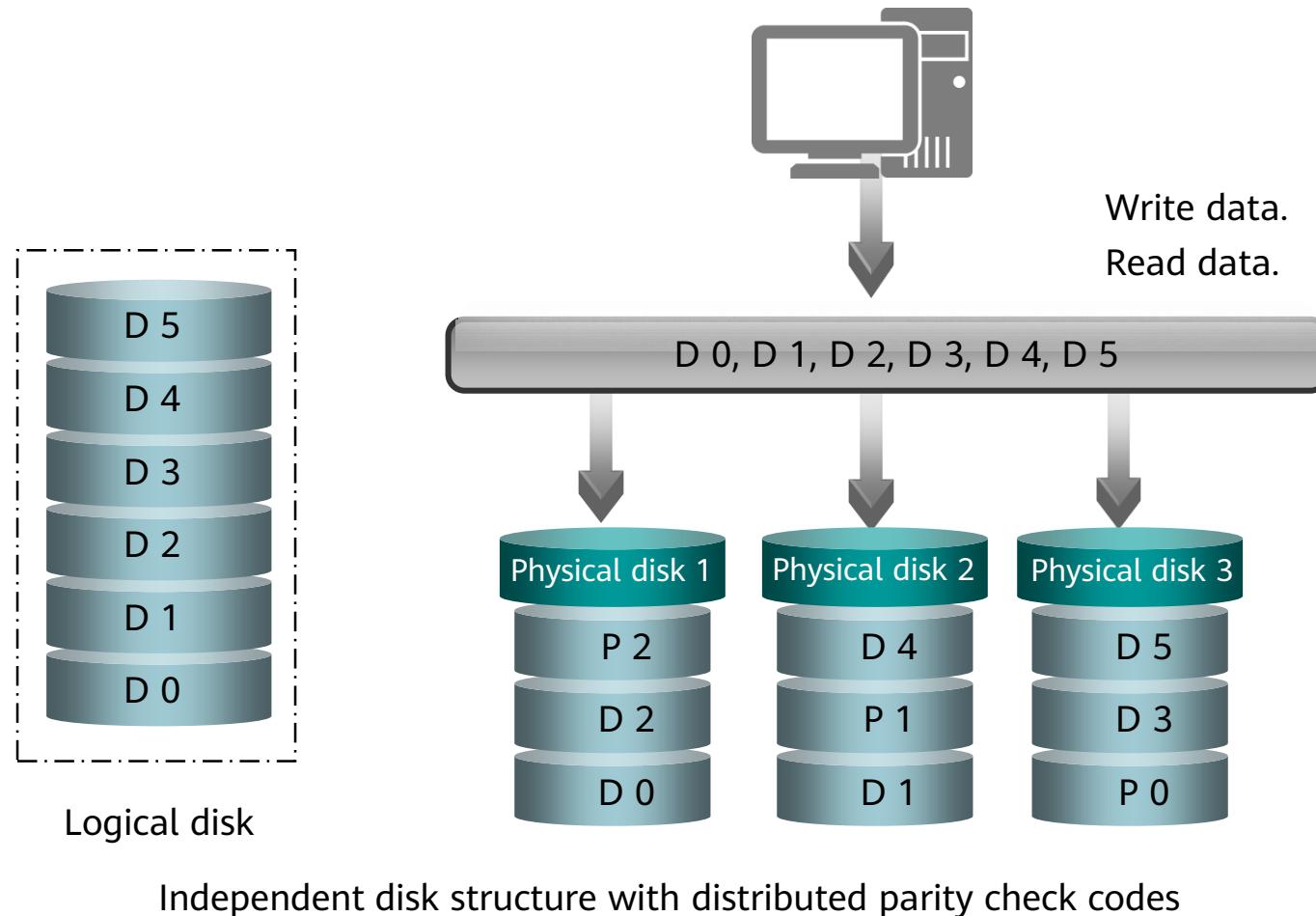


How Does RAID 3 Work?



Note: A write penalty occurs when just a small amount of new data needs to be written to one or two disks.

How Does RAID 5 Work?

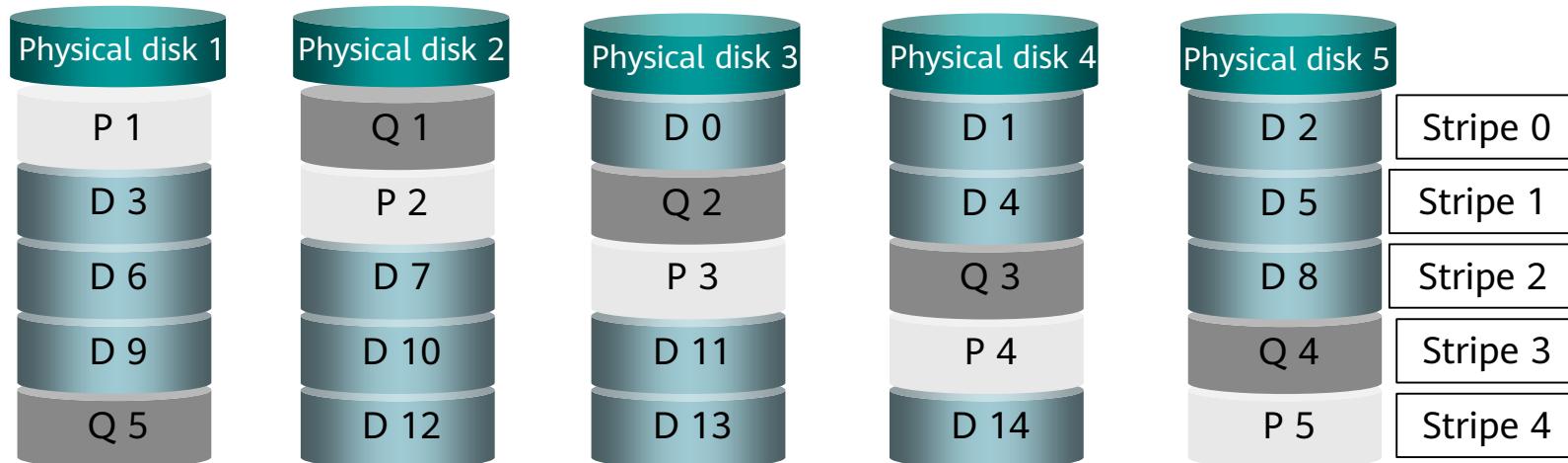


RAID 6

- RAID 6
 - Requires at least $N + 2$ ($N > 2$) disks and provides extremely high data reliability and availability.
- Common RAID 6 technologies:
 - RAID 6 P+Q
 - RAID6 DP

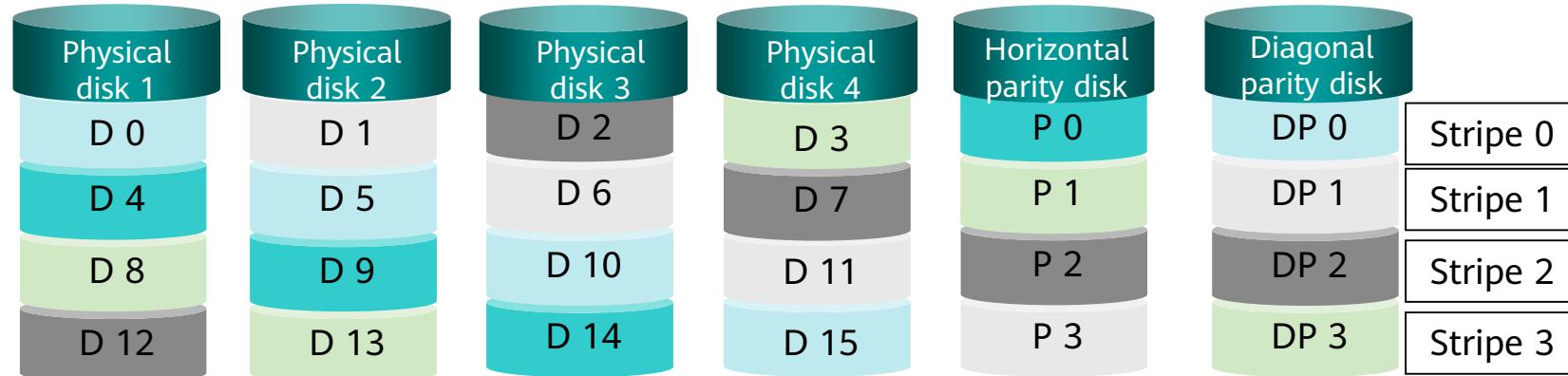
How Does RAID 6 P+Q Work?

- P and Q parity data is calculated. A maximum of two data blocks that are lost can be recovered using P and Q parity data. Formulas for calculating P and Q parity data are as follows:
 - $P = D_0 \oplus D_1 \oplus D_2 \dots$
 - $Q = (\alpha * D_0) \oplus (\beta * D_1) \oplus (\gamma * D_2) \dots$



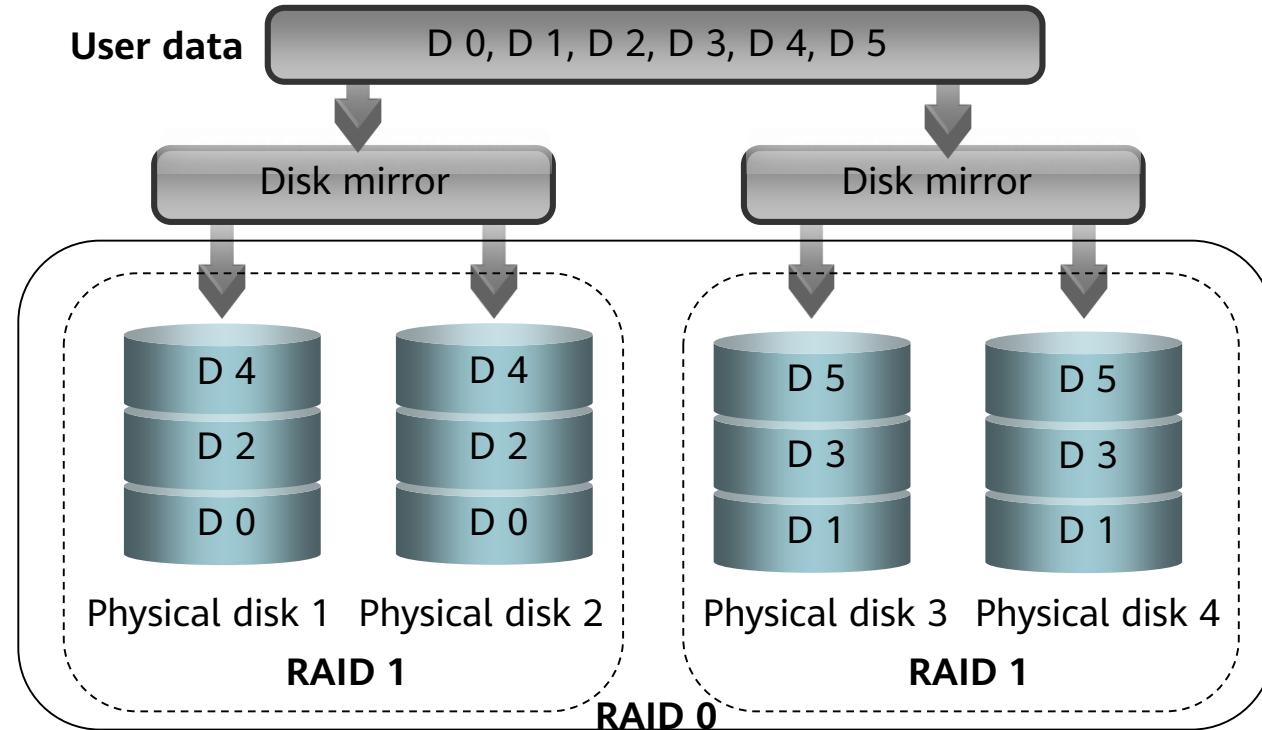
How Does RAID 6 DP Work?

- Double parity (DP) provides fault tolerance up to two failed drives by adding another disk in addition to the horizontal XOR parity disk used in RAID 4 to store diagonal XOR parity data.
- P 0 to P 3 in the horizontal parity disk represent the horizontal parity data for respective disks.
- For example, $P 0 = D 0 \text{ XOR } D 1 \text{ XOR } D 2 \text{ XOR } D 3$
- DP 0 to DP 3 in the diagonal parity disk represent the diagonal parity data for respective data disks and the horizontal parity disk.
- For example, $DP 0 = D 0 \text{ XOR } D 5 \text{ XOR } D 10 \text{ XOR } D 15$



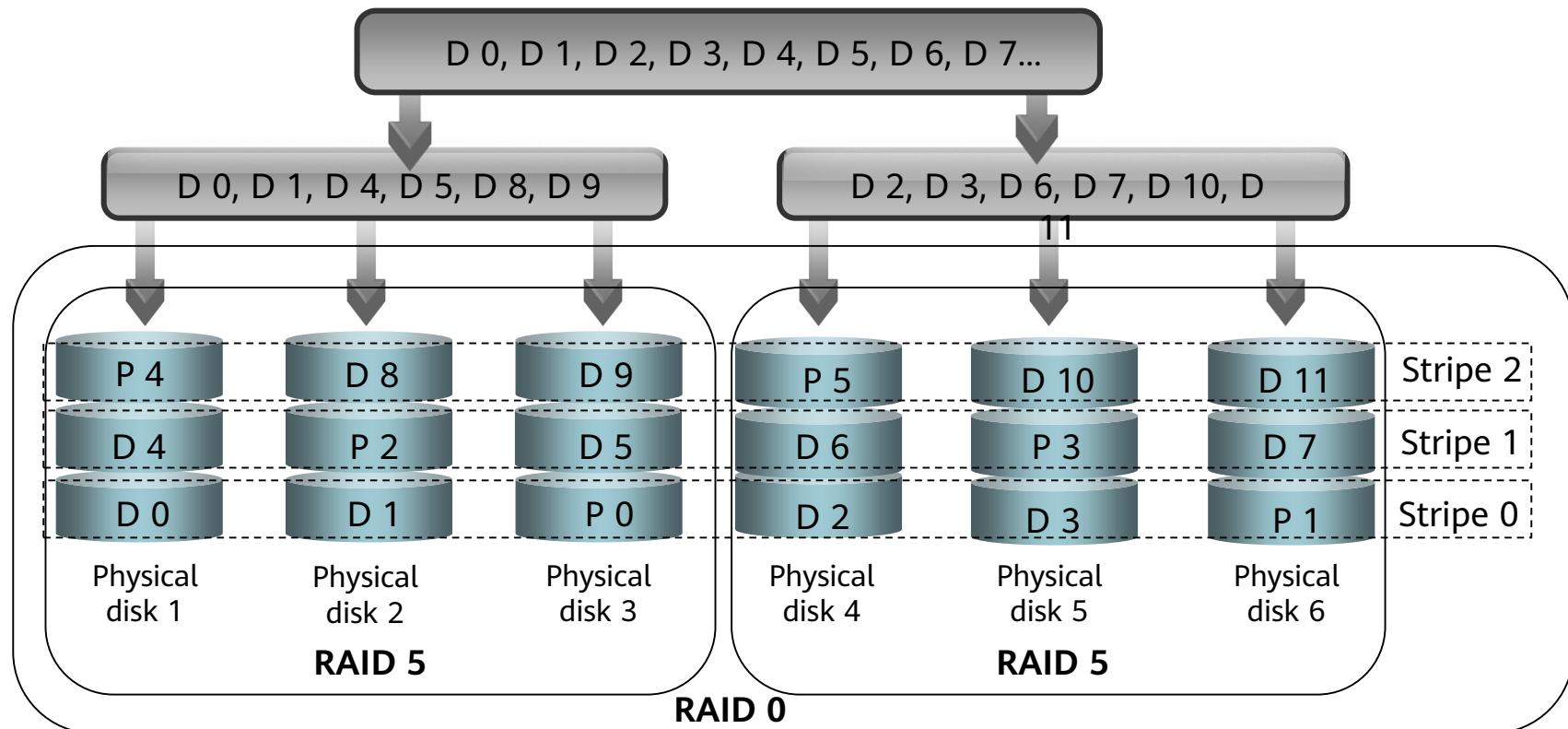
How Does RAID 10 Work?

- RAID 10 consists of nested RAID 1 + RAID 0 levels and allows disks to be mirrored (RAID 1) and then striped (RAID 0). RAID 10 is also a widely used RAID level.



How Does RAID 50 Work?

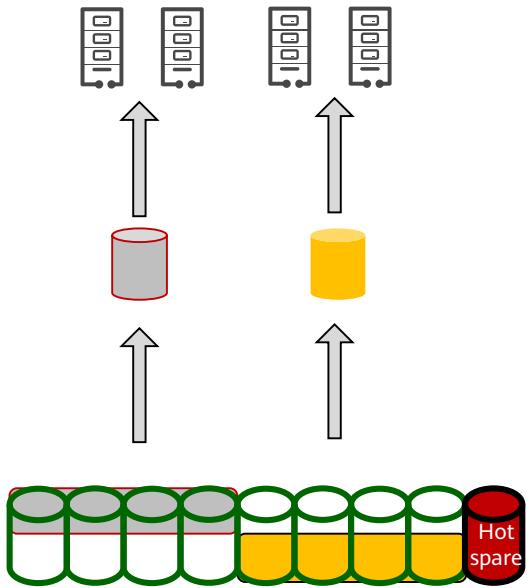
- RAID 50 consists of nested RAID 5 + RAID 0 levels. RAID 0 is implemented after RAID 5 is implemented.



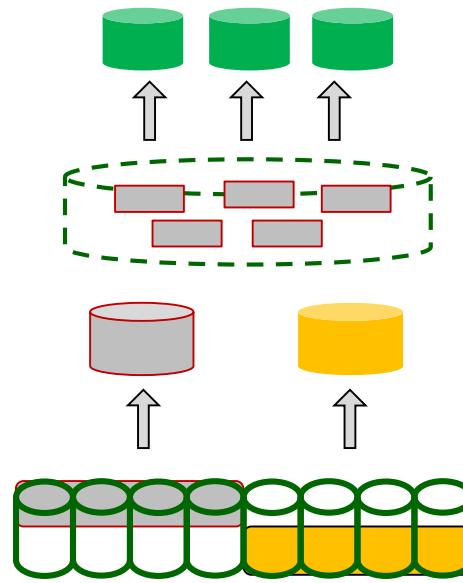
Contents

1. Traditional RAID
- 2. RAID 2.0+**
3. Other RAID Technologies

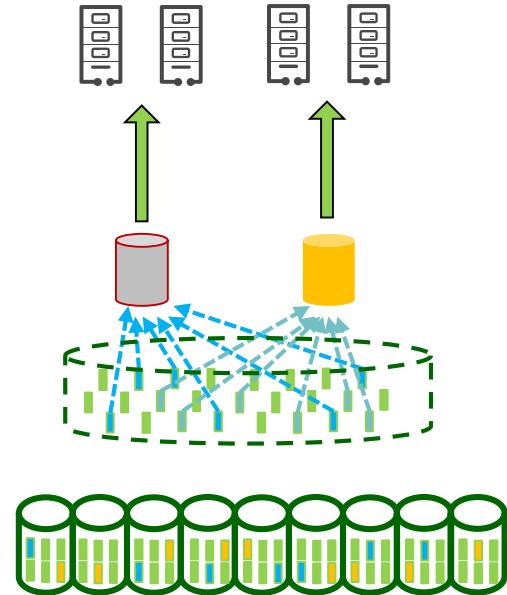
RAID Evolution



Traditional RAID

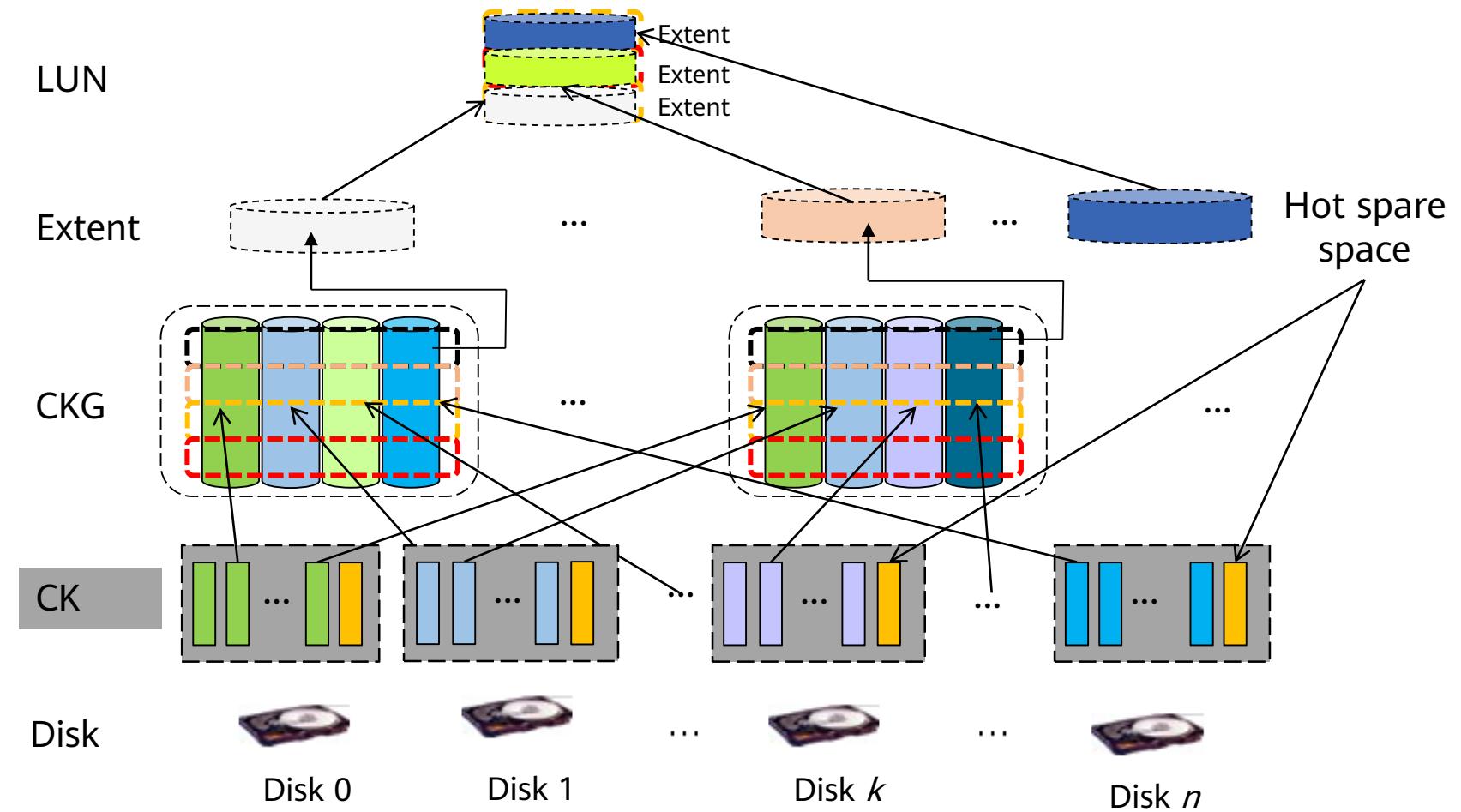


LUN virtualization

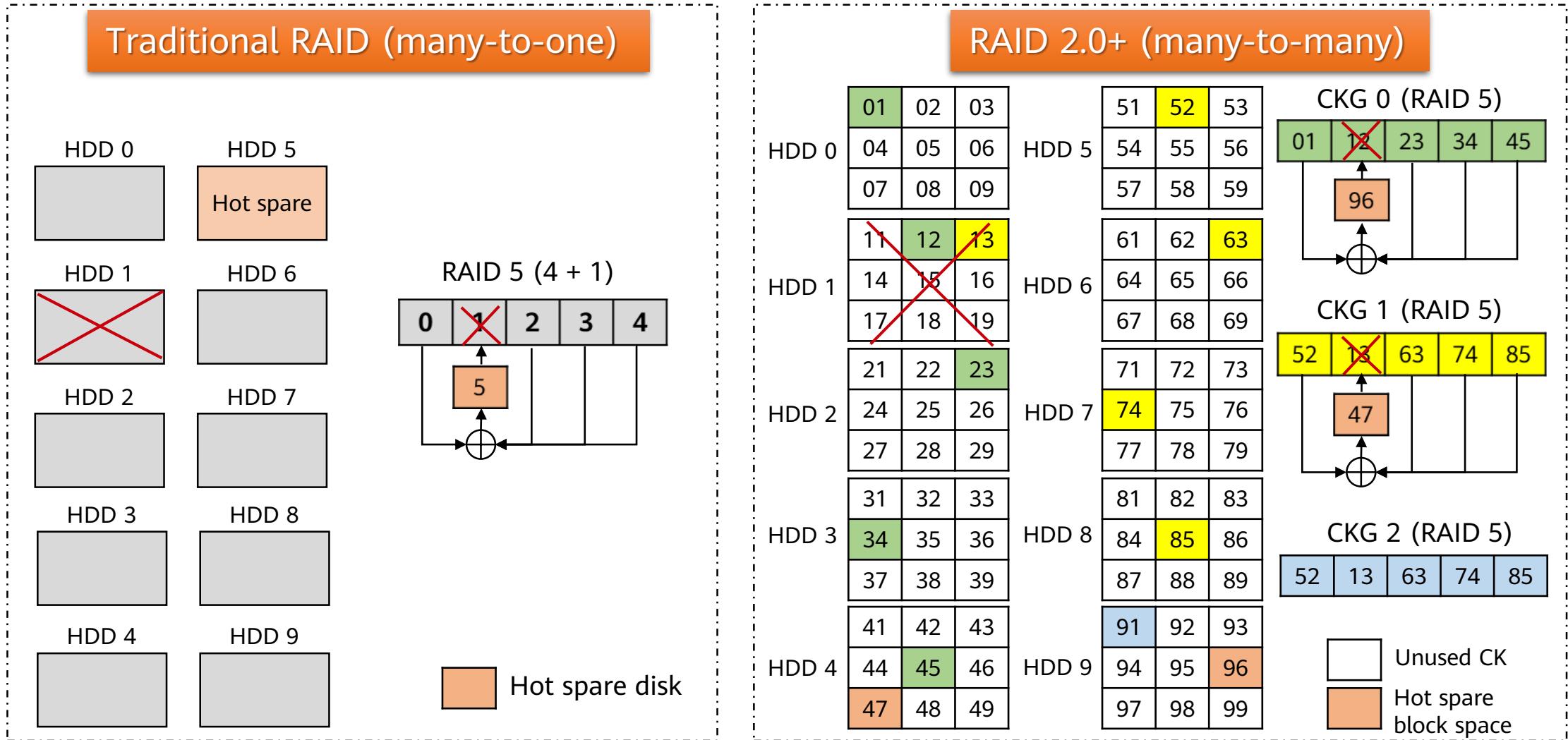


Block virtualization

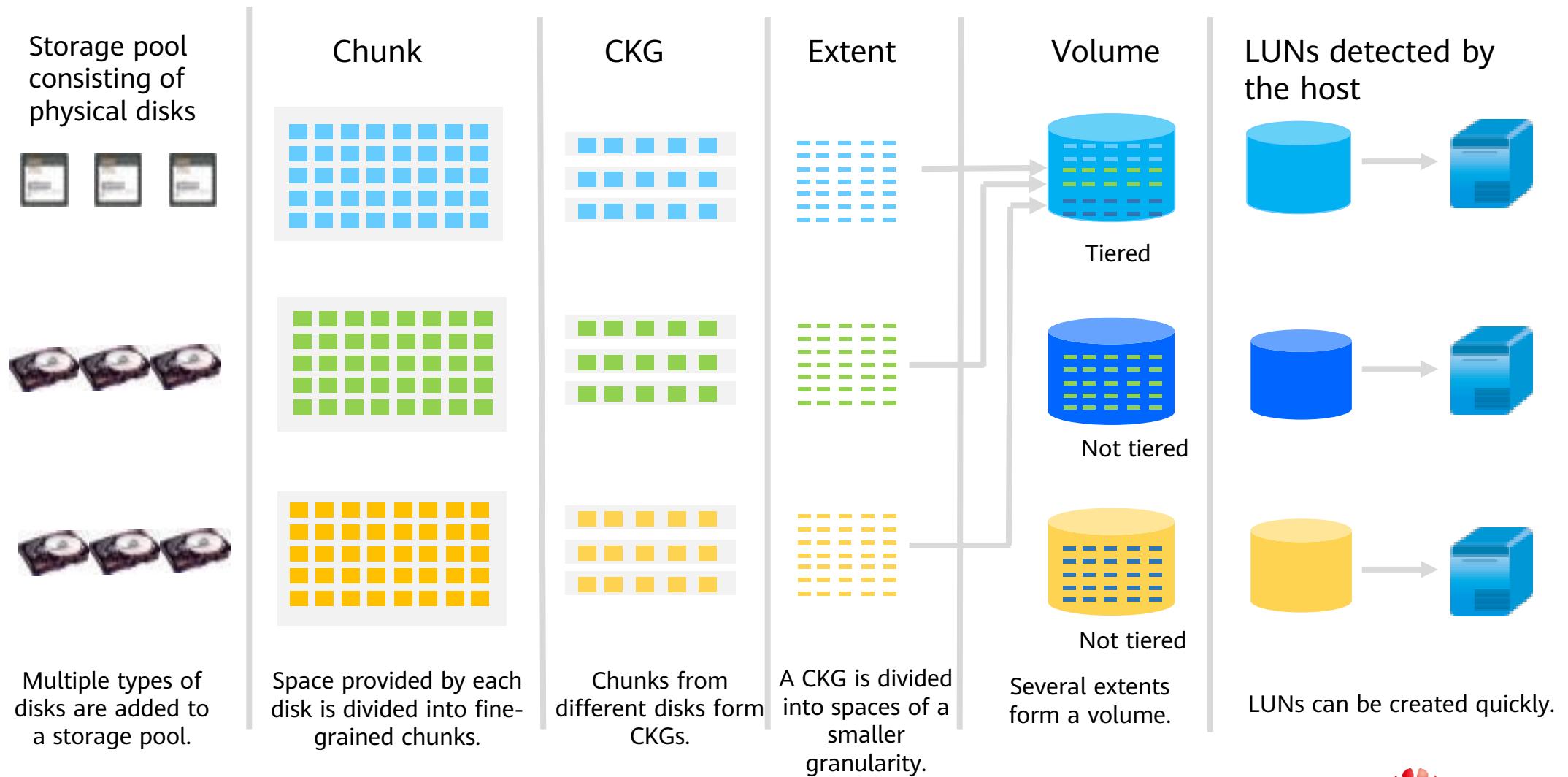
How Does RAID 2.0+ Work?



Reconstruction

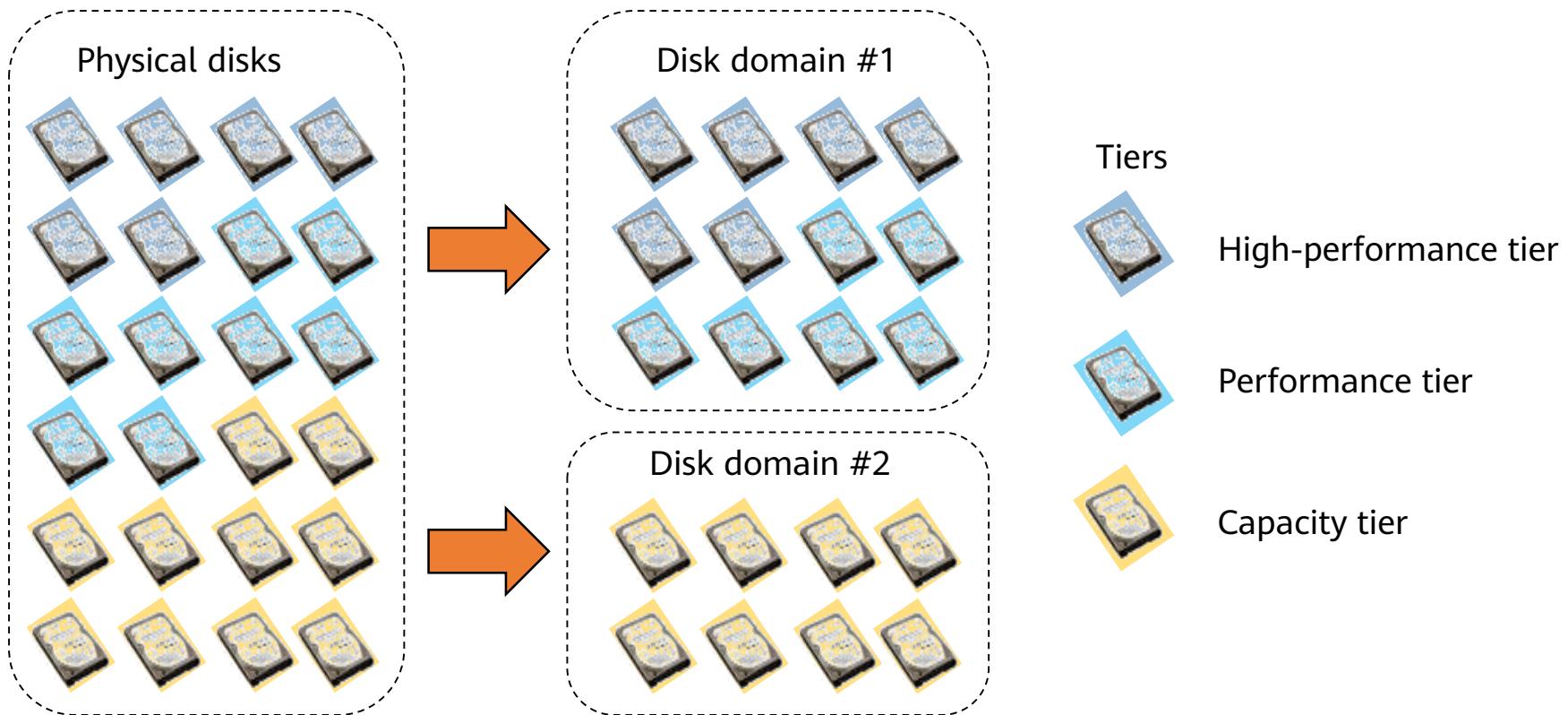


Logical Objects



Disk Domain

- A disk domain is a combination of disks (which can be all disks in the array). After the disks are combined and reserved for hot spare capacity, it provides storage resources for the storage pool.



Storage Pool and Tier

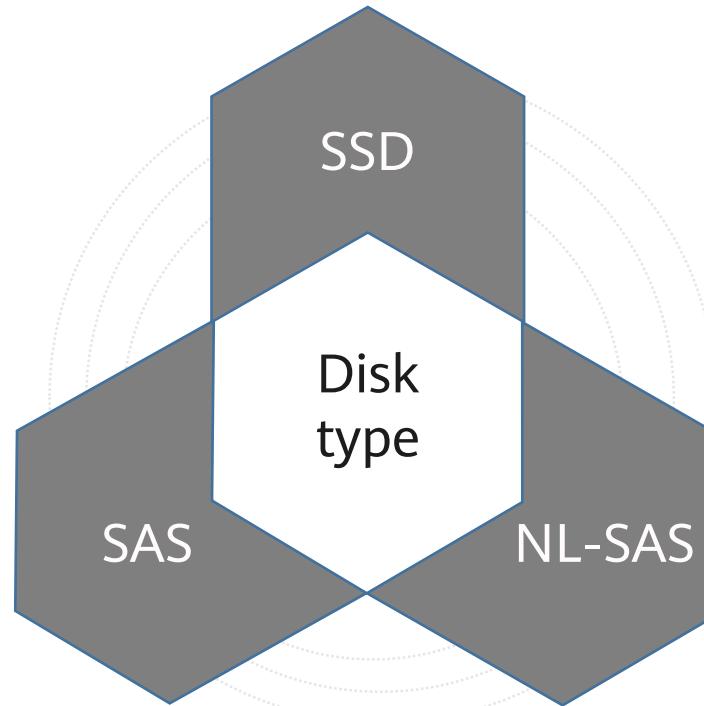
- A storage pool is a storage resource container. The storage resources used by application servers are all from storage pools.
- A storage tier is a collection of storage media providing the same performance level in a storage pool. Different storage tiers manage storage media of different performance levels and provide storage space for applications that have different performance requirements.

Storage Tier	Tier Type	Supported Disk Type	Application
Tier 0	High-performance tier	SSD	Best for storage of data that is frequently accessed with high performance and price.
Tier 1	Performance tier	SAS	Best for storage of data that is less frequently accessed with relatively high performance and moderate price.
Tier 2	Capacity tier	NL-SAS	Best for storage of mass data that is infrequently accessed with low performance and price, and large capacity per disk.

RAID Level	RAID Policy
RAID 1	1D + 1D, 1D + 1D + 1D + 1D
RAID 10	2D + 2D or 4D + 4D, which is automatically selected by a storage system
RAID 3	2D + 1P, 4D + 1P, 8D + 1P
RAID 5	2D + 1P, 4D + 1P, 8D + 1P
RAID 50	(2D + 1P) x 2, (4D + 1P) x 2, or (8D + 1P) x 2
RAID 6	2D + 2P, 4D + 2P, 8D + 2P, 16D + 2P

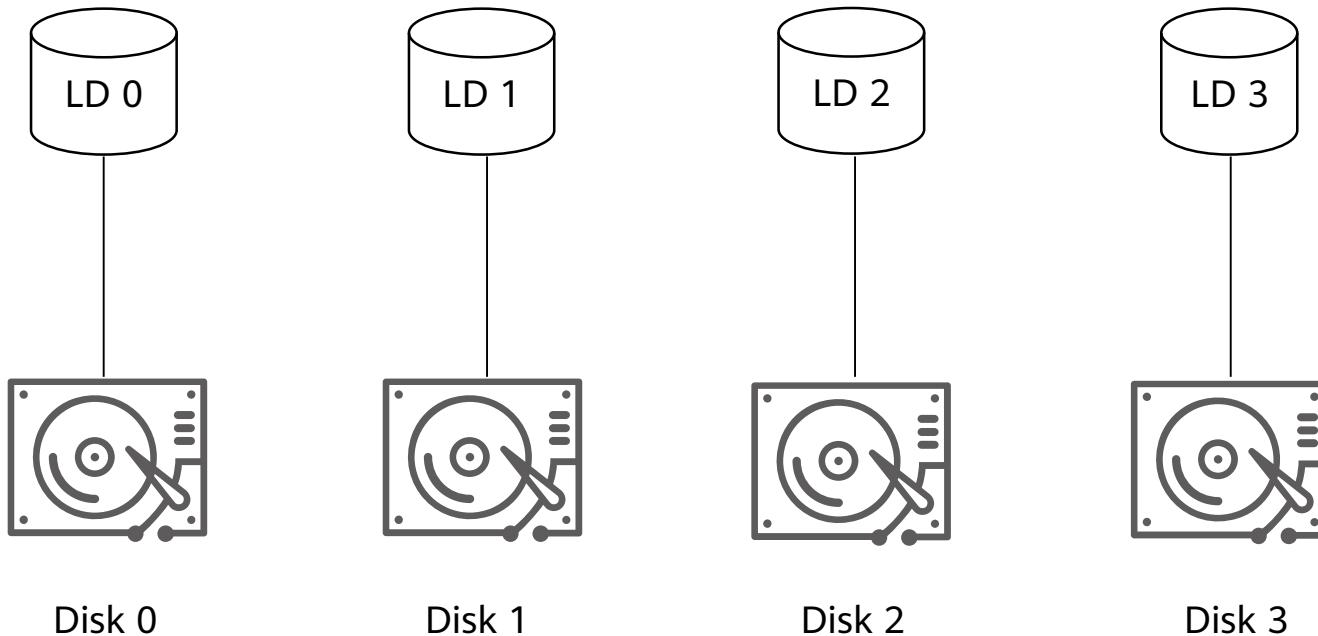
Disk Group

- A disk group (DG) is a set of disks of the same type in a disk domain. The disk type can be SSD, SAS, or NL-SAS.



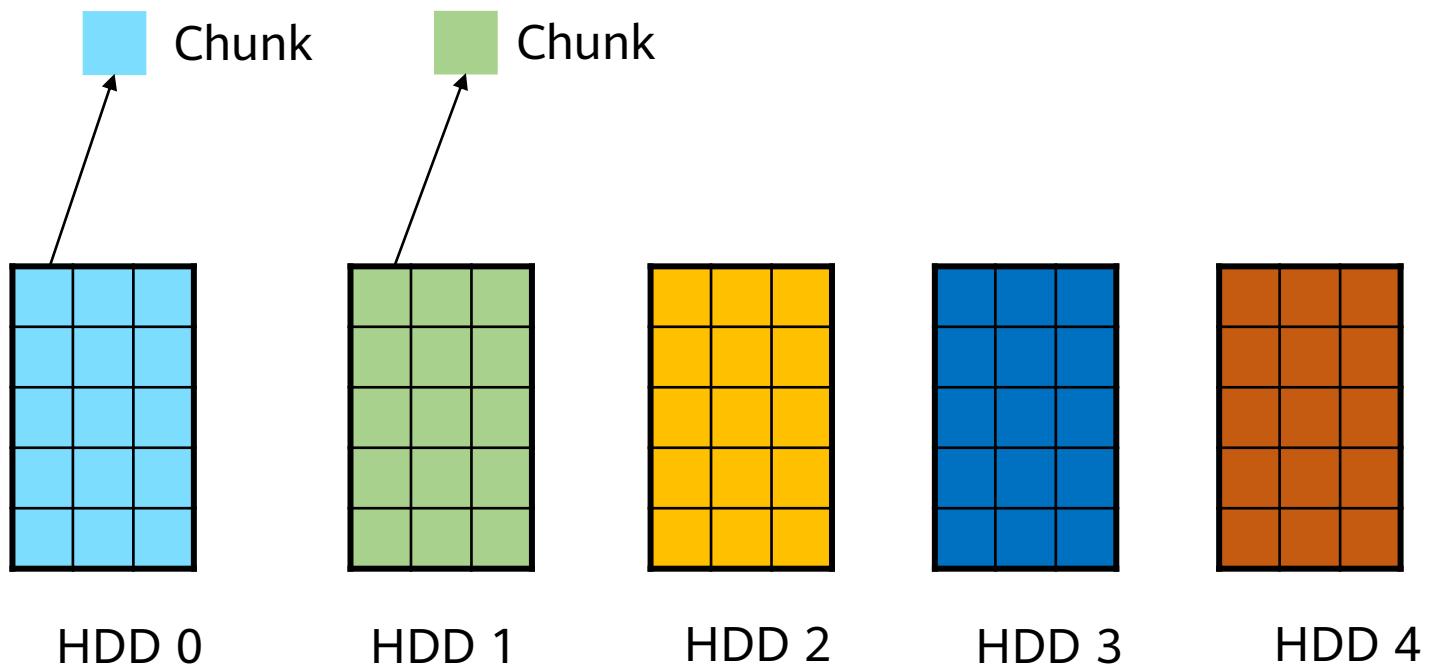
Logical Drive

- A logical drive (LD) is a disk that is managed by a storage system and corresponds to a physical disk.



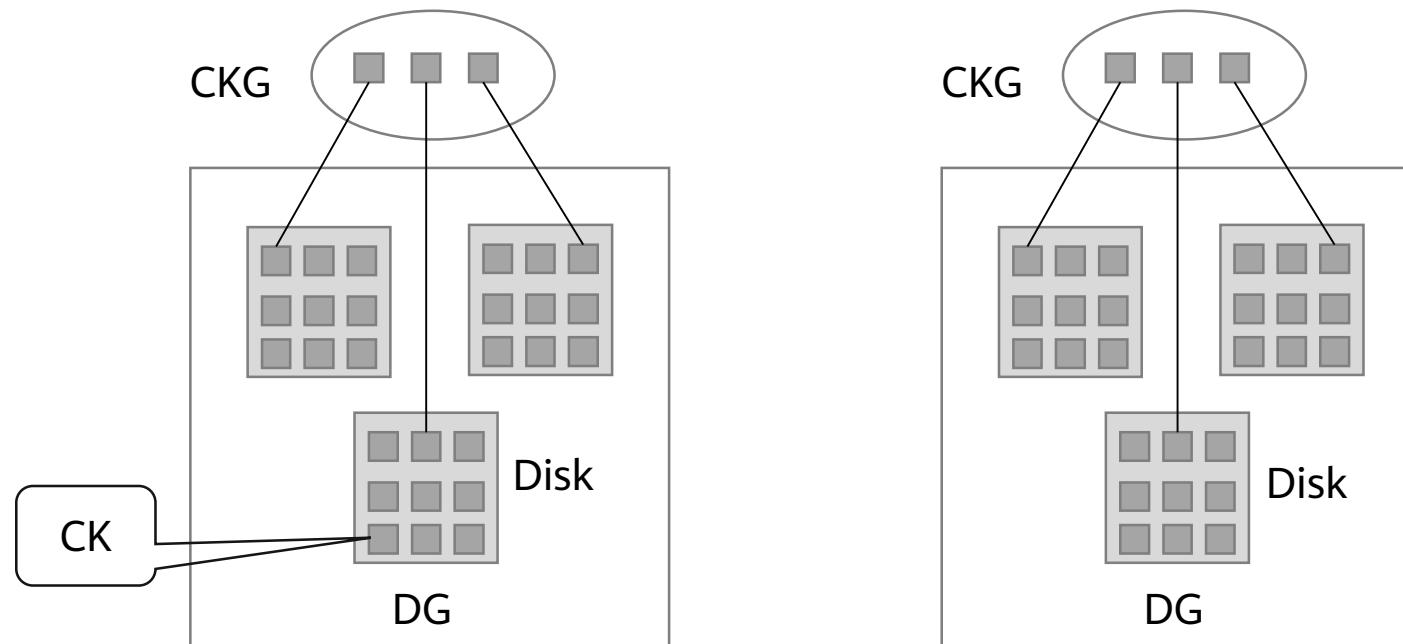
Chunk

- A chunk (CK) is a disk space of a specified size allocated from a storage pool. It is the basic unit of a RAID array.



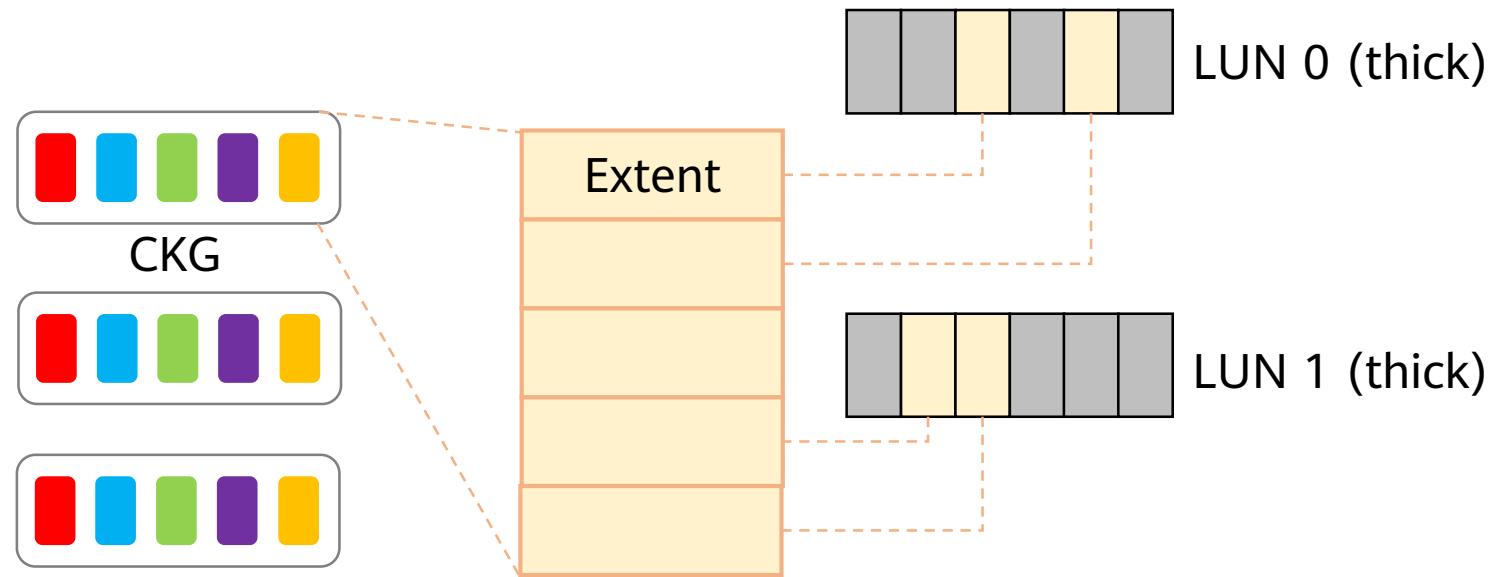
CKG

- A chunk group (CKG) is a logical storage unit that consists of CKs from different disks in the same DG based on the RAID algorithm. It is the minimum unit for allocating resources from a disk domain to a storage pool.



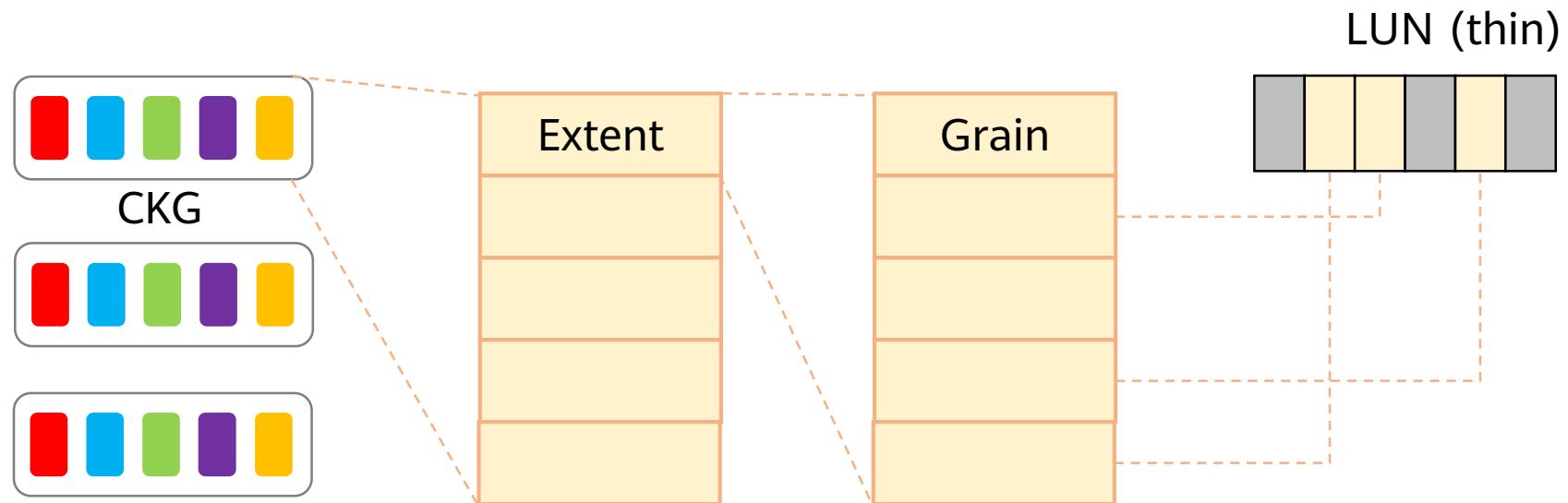
Extent

- Each CKG is divided into logical storage spaces of a specific and adjustable size called extents. Extent is the minimum unit (granularity) for migration and statistics of hot data. It is also the minimum unit for space application and release in a storage pool.



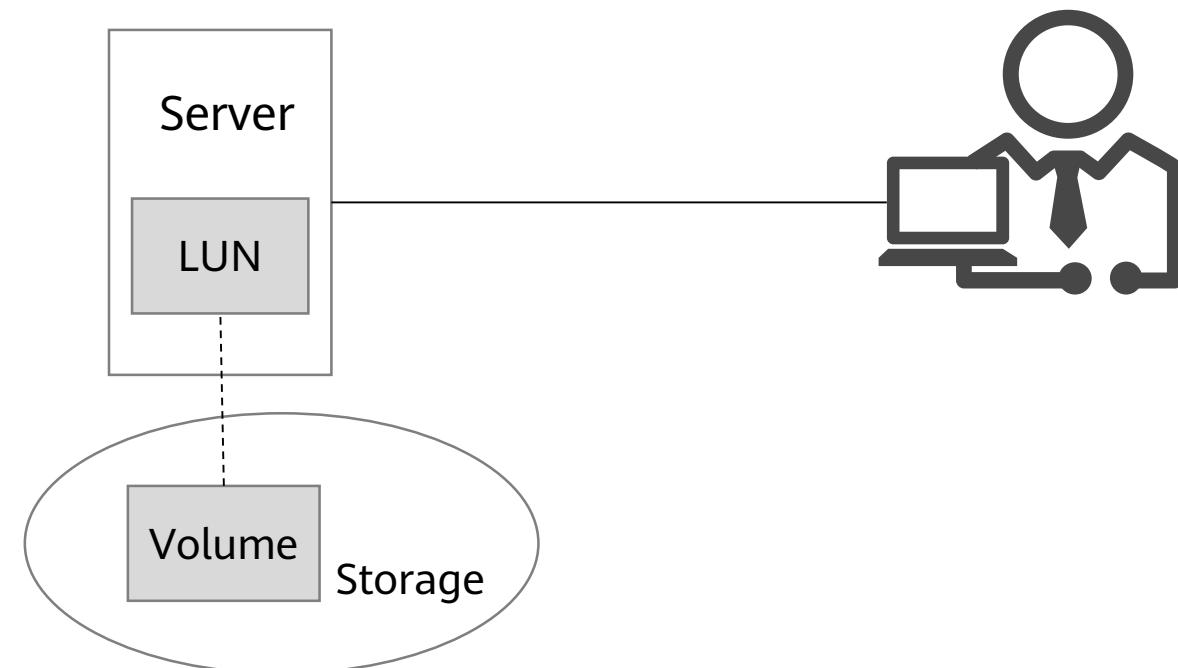
Grain

- When a thin LUN is created, extents are divided into grains of a fixed size. A thin LUN allocates storage space by grains. Logical block addresses (LBAs) in a grain are consecutive.



Volume and LUN

- A volume is an internal management object in a storage system.
- A LUN is a storage unit that can be directly mapped to a host for data reads and writes. A LUN is the external embodiment of a volume.



Contents

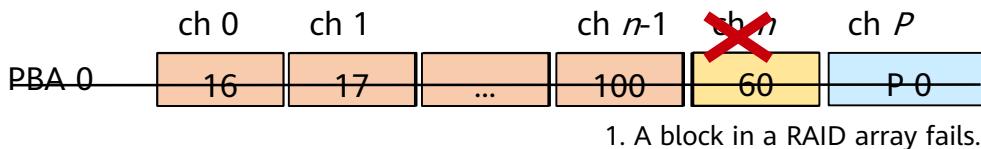
1. Traditional RAID
2. RAID 2.0+
- 3. Other RAID Technologies**

Huawei Dynamic RAID Algorithm

Common RAID algorithm

- When a block in a RAID array fails, recover the data in the faulty block, migrate all the data in the RAID array, and then shield the RAID array.
- Result: A large amount of available flash memory space is wasted.

4. Shield and obsolete the faulty RAID array, which wastes space.



1. A block in a RAID array fails.

2. Create a new RAID array to store the data of the RAID array where a block fails.

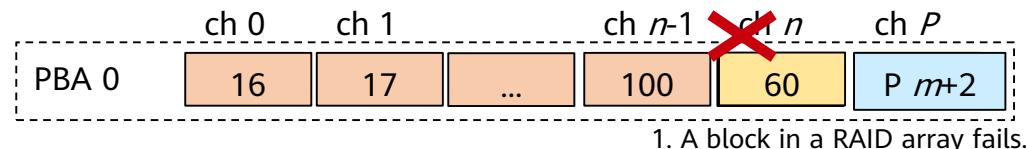


3. Recover the data in the faulty block using the RAID algorithm and migrate all the data in the RAID array.

Huawei dynamic RAID algorithm

- When a block in a RAID array fails, recover and migrate the data in the faulty block, shield the faulty block, and reconstruct a new RAID array using remaining blocks.
- Benefit: The flash memory space is fully and effectively used.

4. Reconstruct a new RAID array using remaining blocks to store data.



1. A block in a RAID array fails.



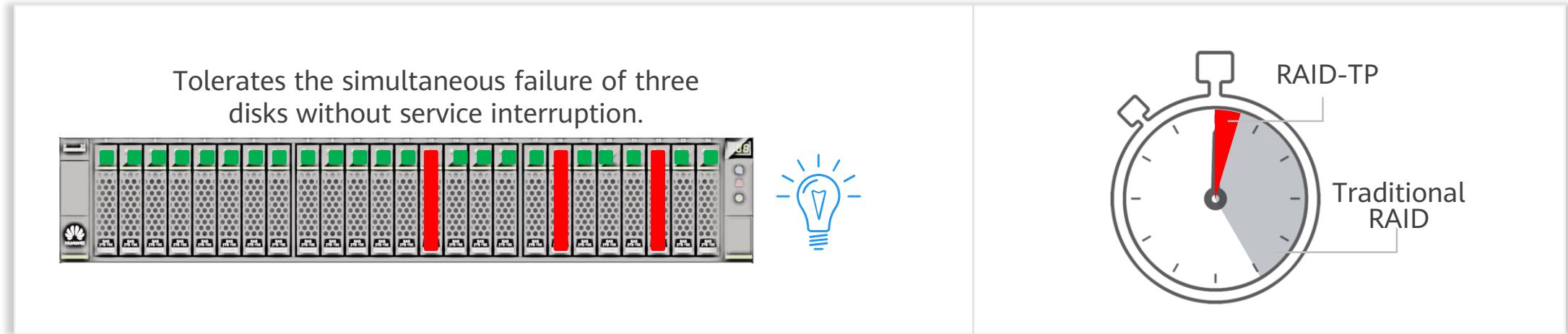
2. Create a new RAID array to store the data in the faulty block.



3. Recover the data in the faulty block using the RAID algorithm and migrate the data.

RAID-TP

- RAID protection is essential to a storage system for consistently high reliability and performance. However, the reliability of RAID protection is challenged by uncontrollable RAID array construction time due to drastic increase in capacity.
- RAID-TP achieves optimal performance, reliability, and capacity utilization.



Huawei RAID-TP:

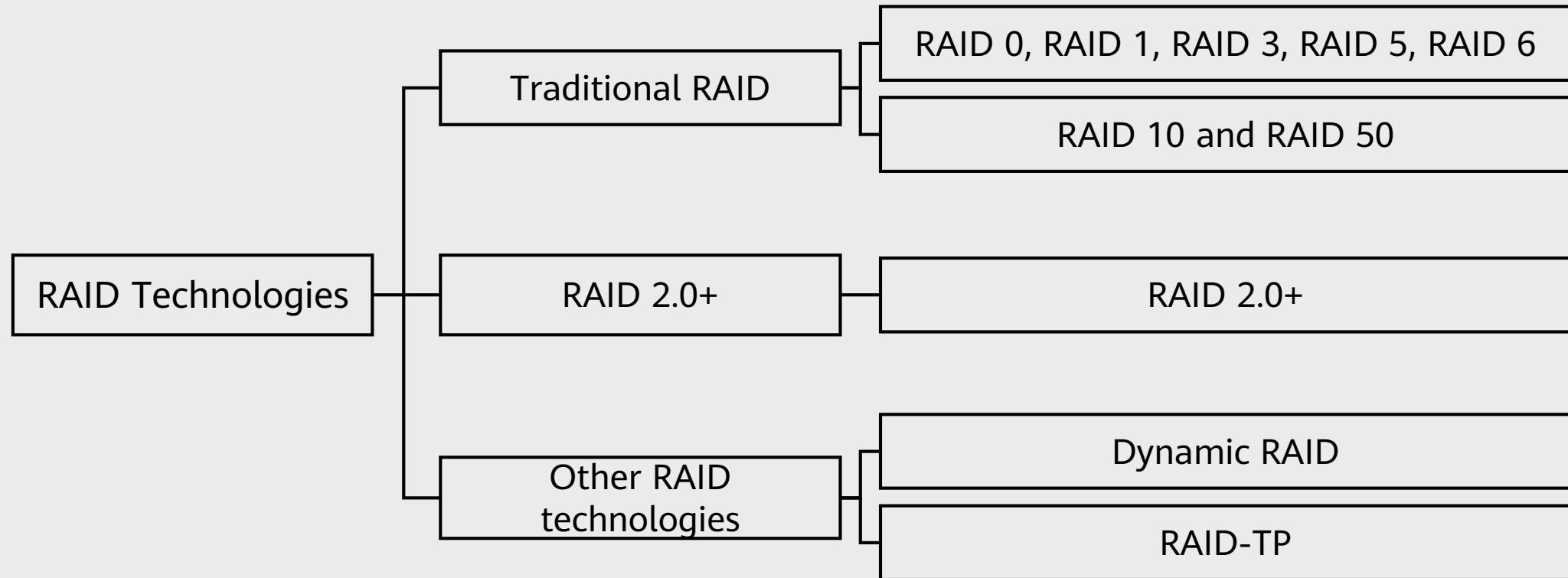
- Tolerates the simultaneous failure of three disks

- Greatly reduces reconstruction time.
- Effectively copes with data protection challenges in the era of large-capacity disks.

Quiz

1. What is the difference between a strip and a stripe?
2. Which RAID level would you recommend if a user focuses on reliability and random write performance?
3. Is it true or false that data access will remain unaffected when any disk in a RAID 10 array fails?

Summary



Recommendations

- Huawei official websites
 - Enterprise business: <http://enterprise.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/en/index.html>
 - Online learning: <https://learning.huawei.com/en/>
- Popular tools
 - HedEx Lite
 - Network Documentation Tool Center
 - Information Query Assistant

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

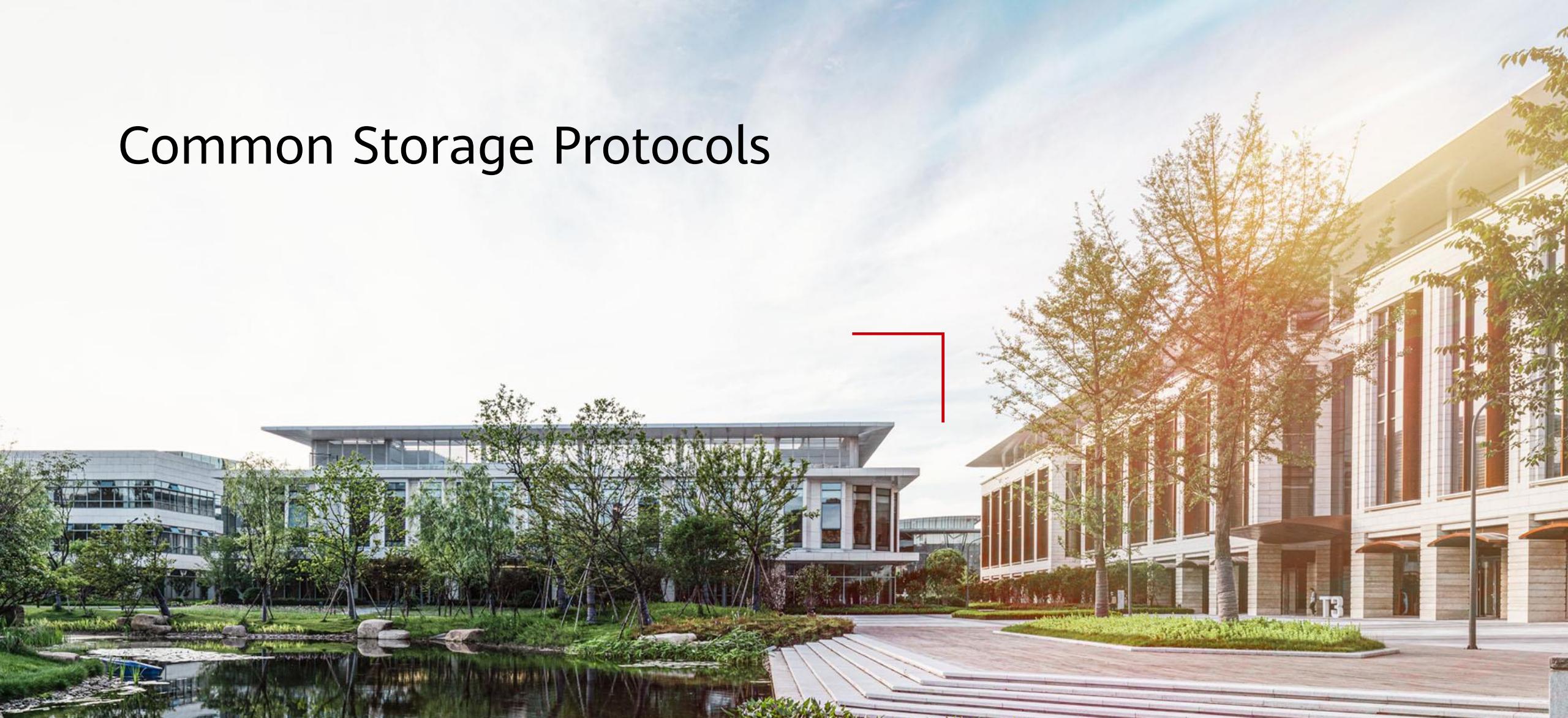
Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.



Common Storage Protocols



Foreword

- A protocol is a set of conventions that both computers must comply with to communicate. For example, the protocol may establish conventions to set up connections or identify each other.
- A protocol not only defines the language used for communication, but also specifies the hardware, transmission medium, transmission protocol, and interface technology. This course describes the definitions and principles of different storage protocols.

Objectives

Upon completion of this course, you will learn:

- Common protocols used in storage systems; and
- Working principles and characteristics of these protocols

Contents

1. SAN Protocols

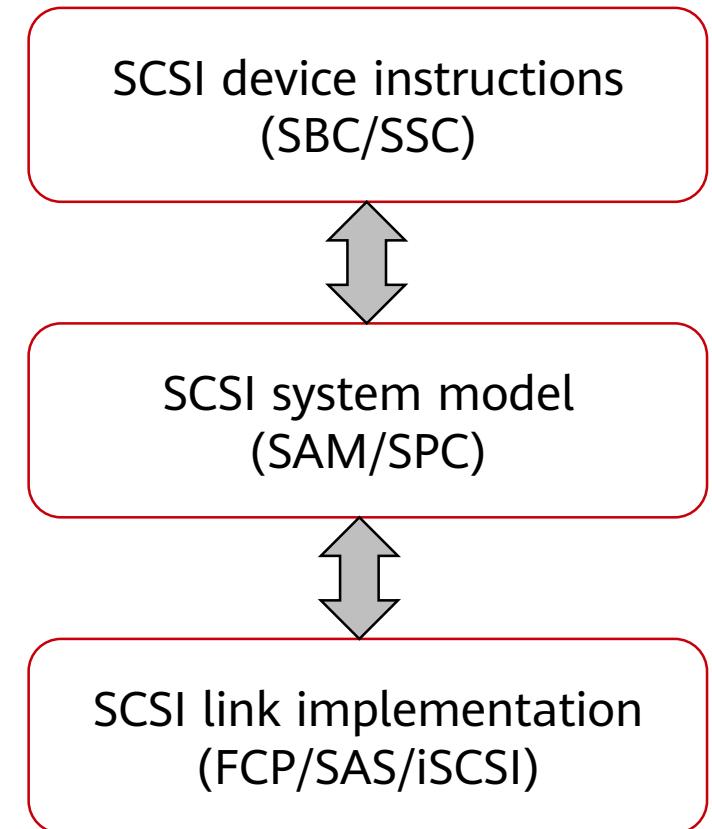
- SCSI and SAS
 - iSCSI and FC
 - PCIe and NVMe
 - RDMA and RoCE

2. NAS Protocols

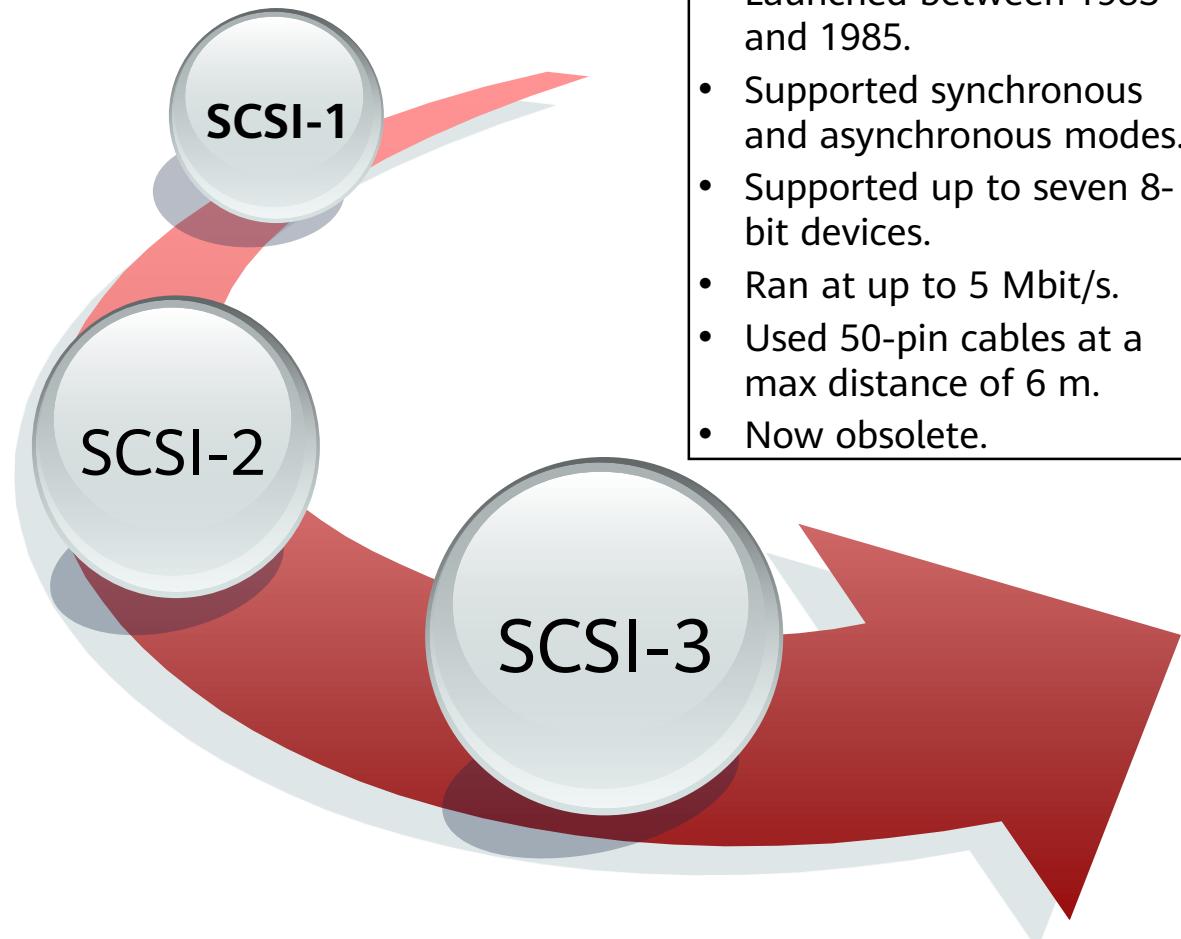
3. Object and HDFS Storage Protocols

SCSI Protocol

- Small Computer System Interface (SCSI) is a vast protocol system evolved from SCSI-1 to SCSI-2 and to its current version, SCSI-3.
- It defines a model and necessary command set for different devices to exchange information using the framework.
- Transmission media has no bearing on the SCSI protocol. It can be used across various media types, including virtual media.



SCSI Evolution



SCSI-1

- Launched between 1983 and 1985.
- Supported synchronous and asynchronous modes.
- Supported up to seven 8-bit devices.
- Ran at up to 5 Mbit/s.
- Used 50-pin cables at a max distance of 6 m.
- Now obsolete.

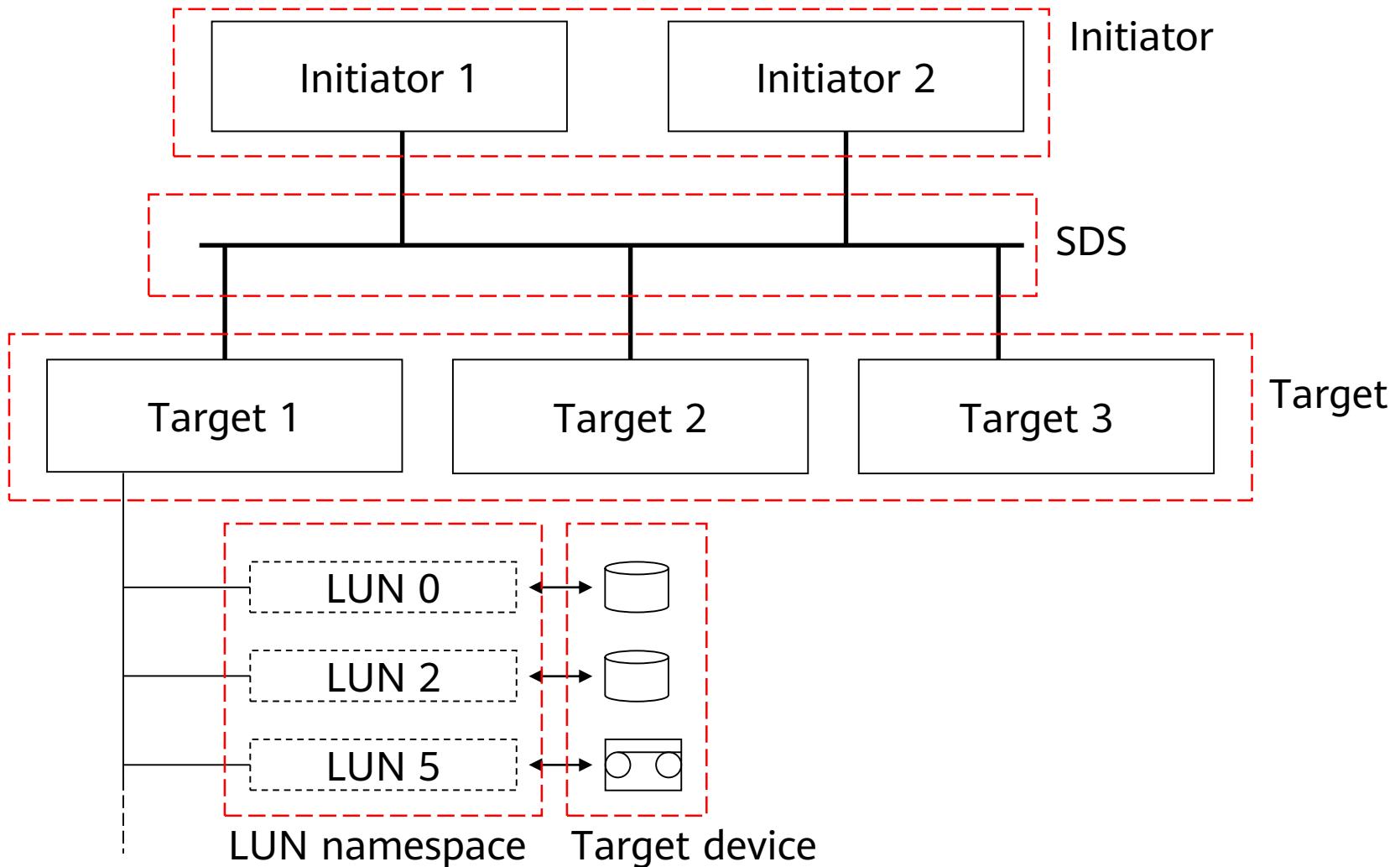
SCSI-2

- Prevalent between 1988 and 1994.
- Compatible with SCSI-1.
- Supports 16-bit bandwidth.
- Runs up to 20 Mbit/s.

SCSI-3

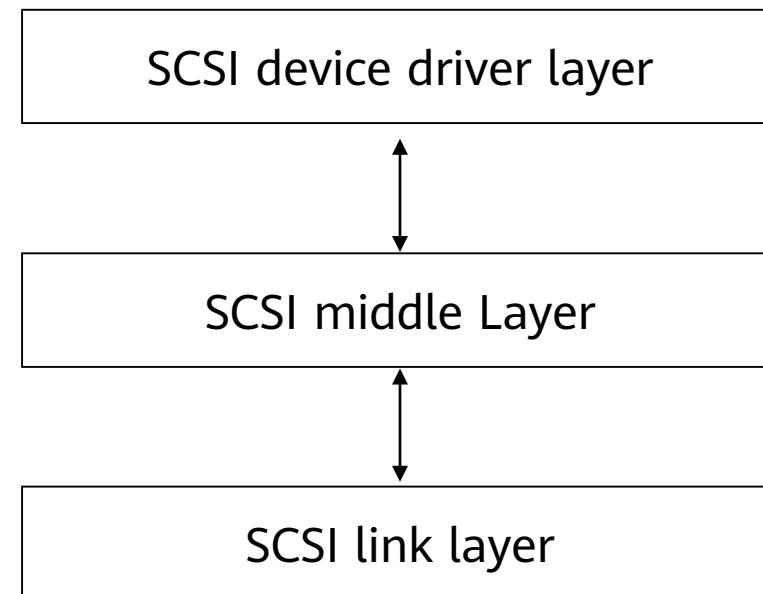
- Standardized in 1993.
- Compatible with SCSI-1 and SCSI-2.
- Is a standard system.
- Supports various media such as FCP and IEEE1394.

SCSI Logical Topology



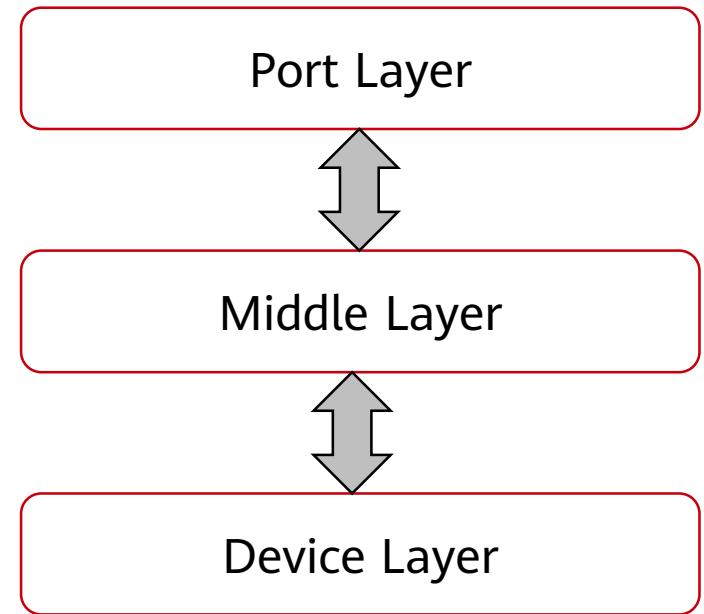
SCSI Initiator Model

- A host's SCSI system typically works in the initiator mode. The SCSI architecture on Windows, Linux, AIX, Solaris, and BSD contains the architecture (middle layer), device, and transport layers.



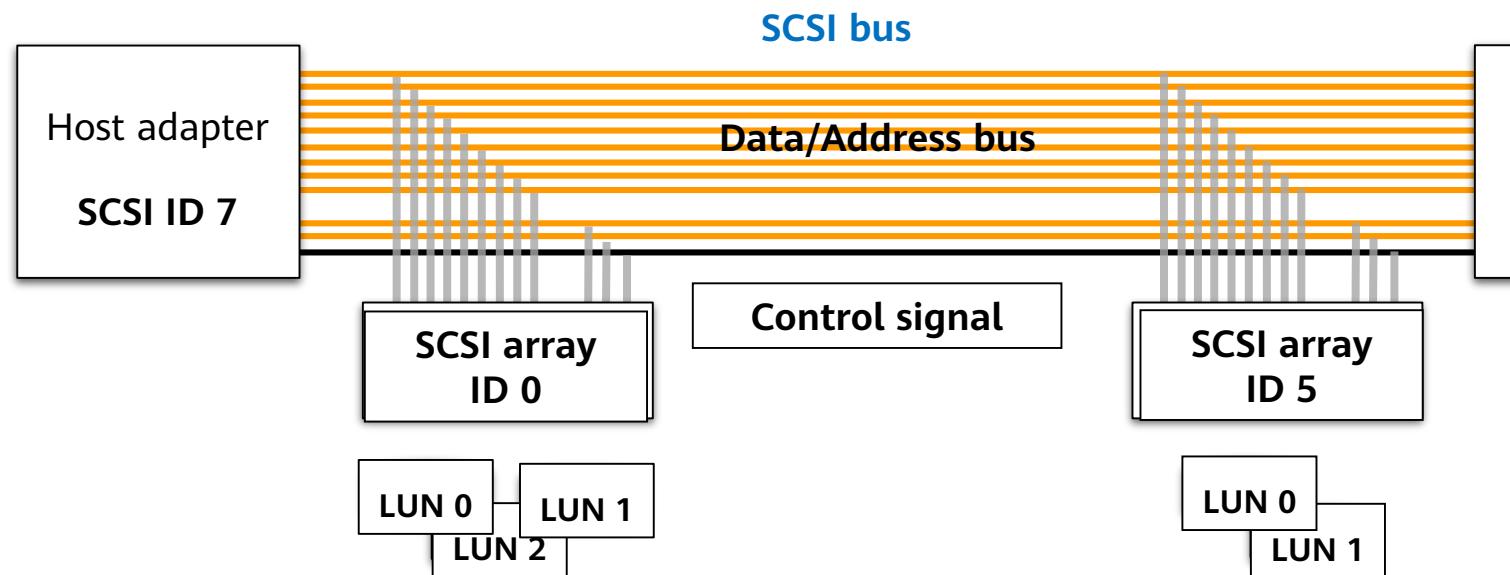
SCSI Target Model

- Based on the SCSI architecture, a target is divided into three layers: port layer, middle layer, and device layer.
- The middle layer, the most important, manages LUN namespaces, link ports, target devices, tasks, task sets, and sessions based on SAM/SPC specifications.
- Drivers at the port layer are dynamically loaded via registration, while those at the device layer are dynamically loaded.

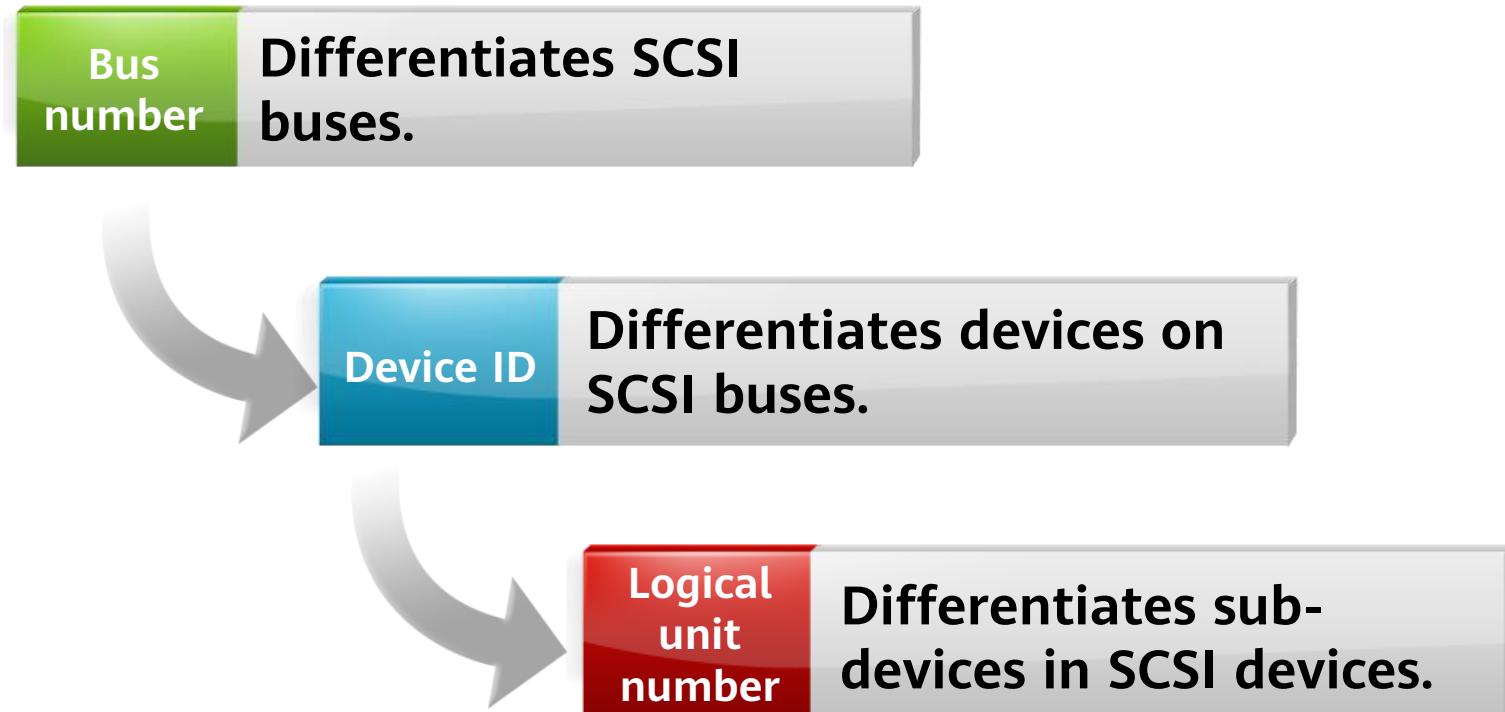


SCSI Protocol and Storage System

- The SCSI protocol is the basic protocol used for communication between hosts and storage devices.
- DAS uses the SCSI protocol to achieve interconnection between hosts and storage devices.

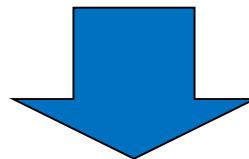


SCSI Protocol Addressing



Background of SAS

- The performance of the parallel SCSI technology has plateaued and the bandwidth cannot be improved.
- Serial bus technologies, such as Fibre Channel, InfiniBand (IB), and Ethernet, have drawbacks in storage applications to some extent:
 - a) Fibre Channel applies to complex networking and long-distance scenarios but requires a high budget.
 - b) InfiniBand requires a high budget and complex networking.
 - c) iSCSI has high latency and low transfer speeds.

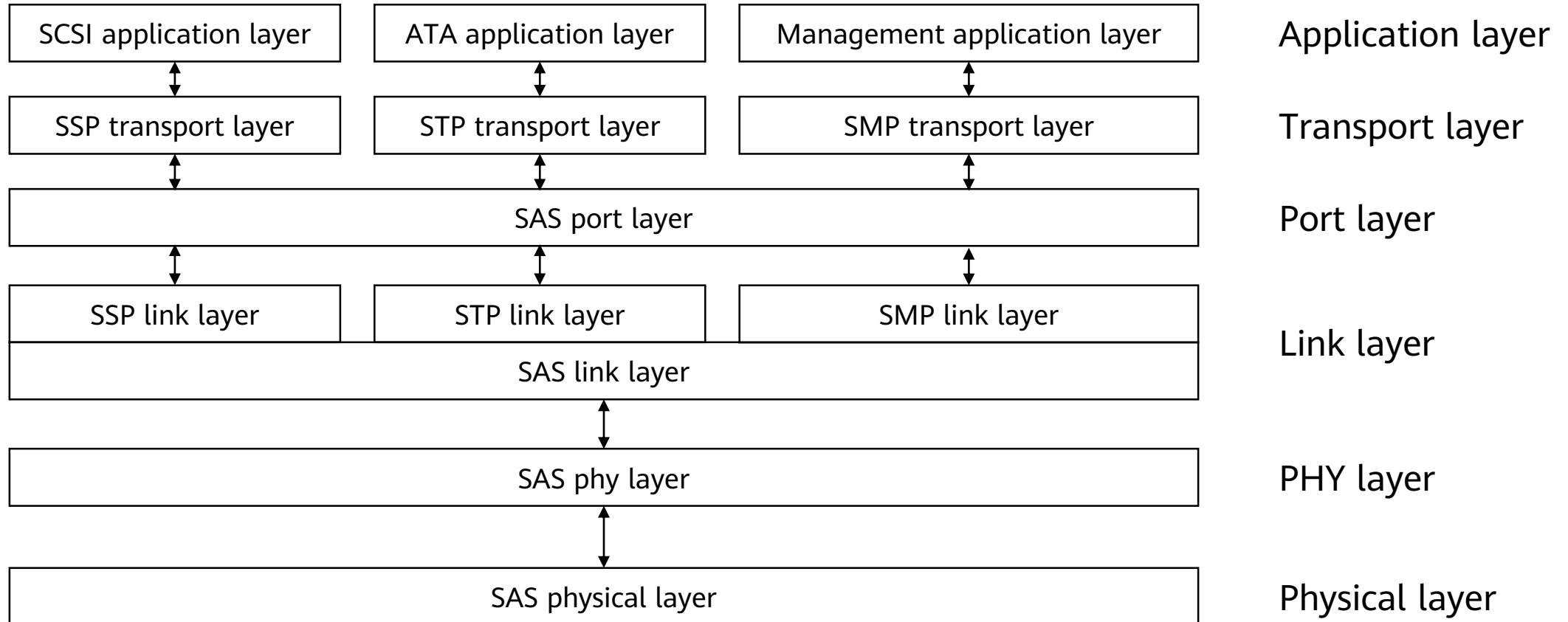


SCSI in serial mode: SAS

What Is SAS

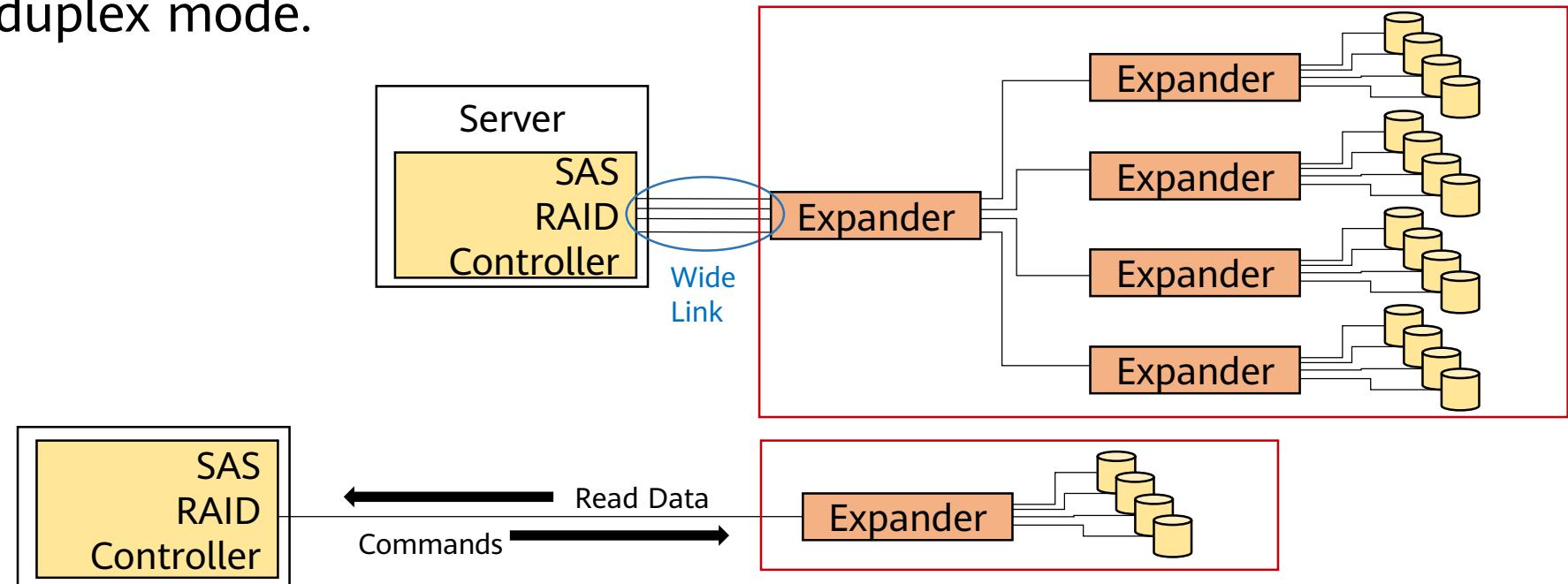
- Serial Attached SCSI (SAS) is the serial standard of the SCSI bus protocol.
- SAS uses serial technology to achieve higher transmission rate and better scalability, and is compatible with SATA disks.
- SAS adopts the point-to-point architecture to achieve a transmission rate of up to 3 Gbit/s, 6 Gbit/s, 12 Gbit/s, or higher. The full-duplex mode is supported.

SAS Protocol Layers



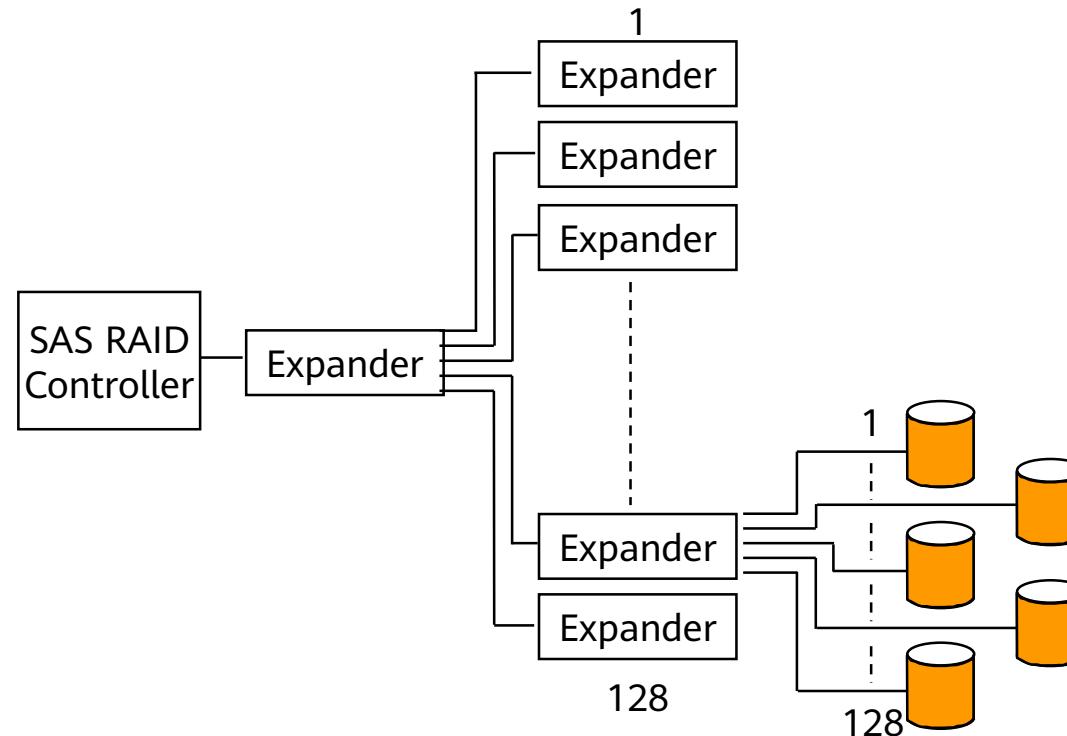
Highlights of SAS

- Provides the serial communication mode to allow multiple data channels to communicate at full speed with devices.
- Binds multiple narrow ports to form a wide port.
- Uses expanders to expand interfaces, providing excellent scalability.
- Works in full-duplex mode.



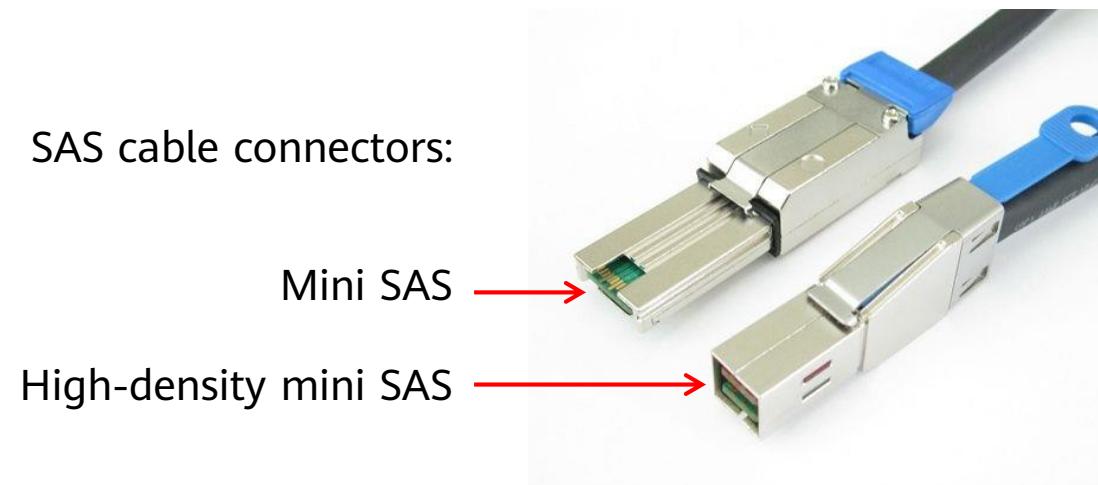
Scalability of SAS

- SAS uses expanders to expand interfaces. One SAS domain supports up to 16,384 disk devices.



Cable Connection Principles of SAS

- Generally, a SAS cable has four channels, each of which supports 12 Gbit/s bandwidth.
- SAS devices are connected in the form of a loop (also called a chain).
- The cable bandwidth is 4×12 Gbit/s, which limits the number of disks that can be included in a loop.
- A maximum of 168 disks are supported in a loop. That is, a loop consists of a maximum of seven disk enclosures with 24 disk slots each.



Contents

1. SAN Protocols

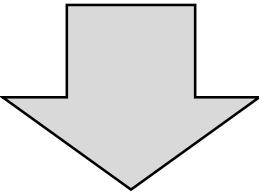
- SCSI and SAS
- iSCSI and FC
 - PCIe and NVMe
 - RDMA and RoCE

2. NAS Protocols

3. Object and HDFS Storage Protocols

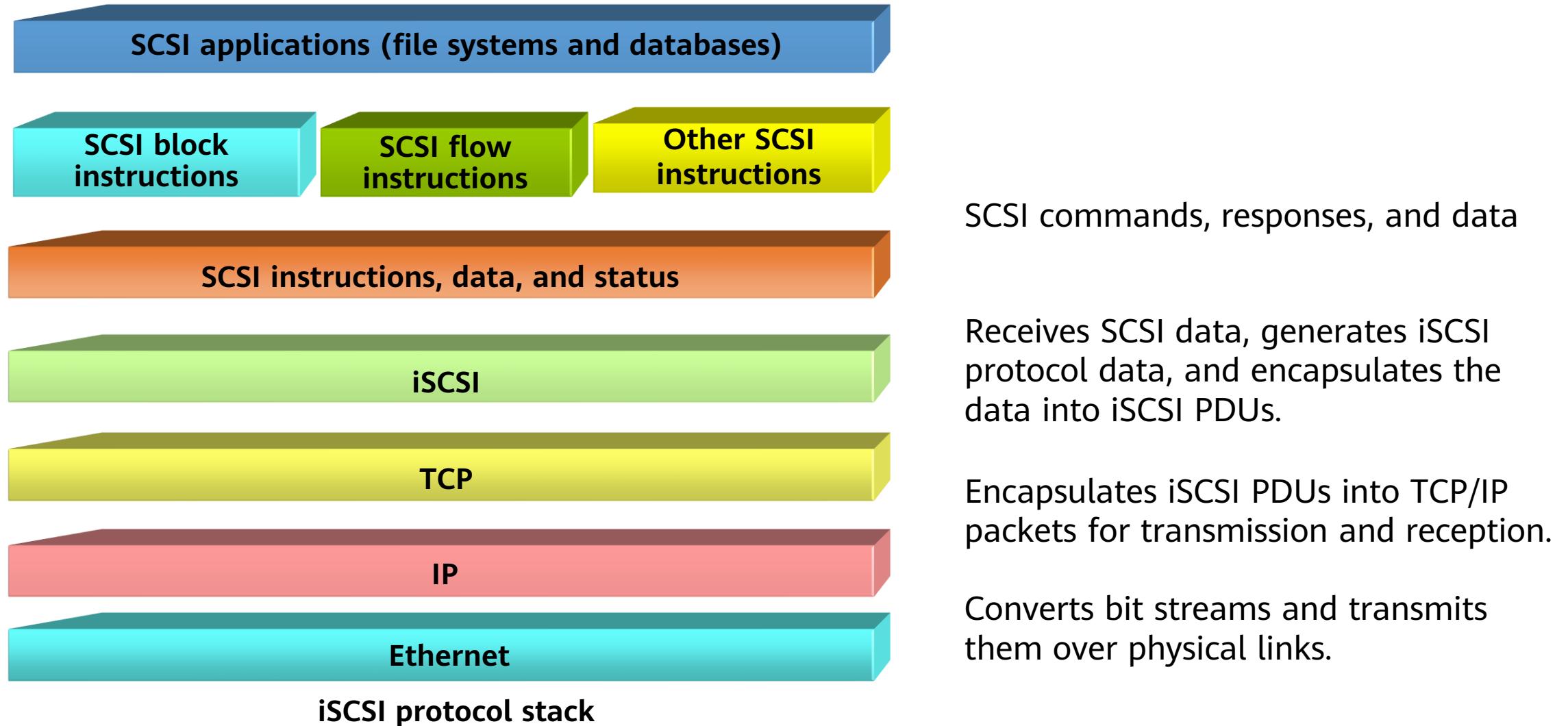
Emergence of iSCSI

SCSI is used to connect a small number of devices at a limited distance.



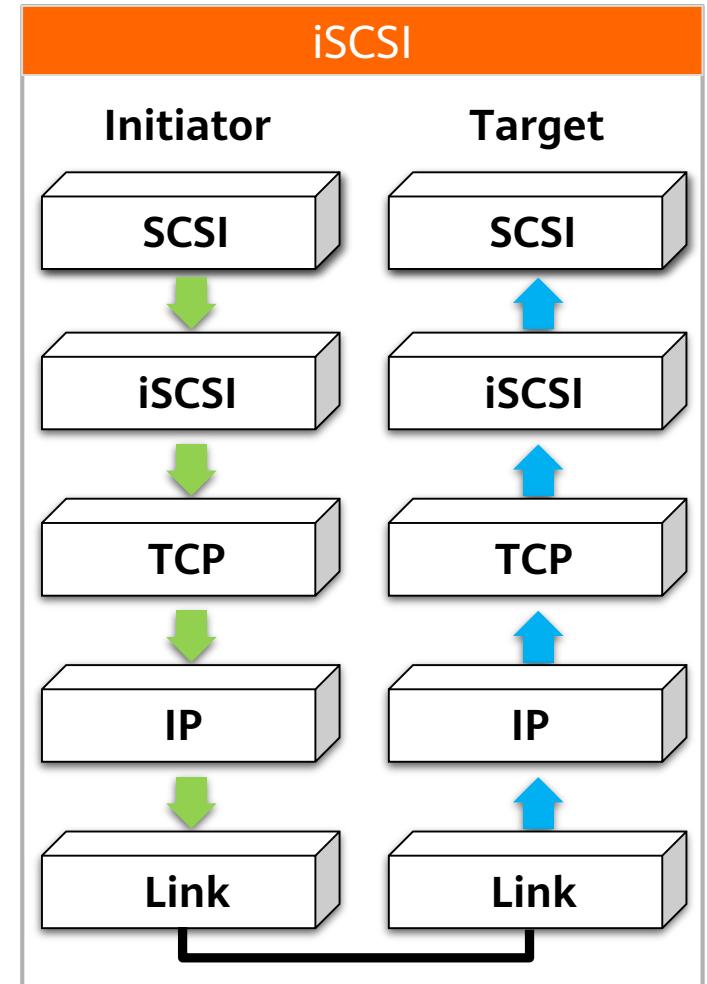
IP-network-based SCSI: iSCSI

Introduction to iSCSI



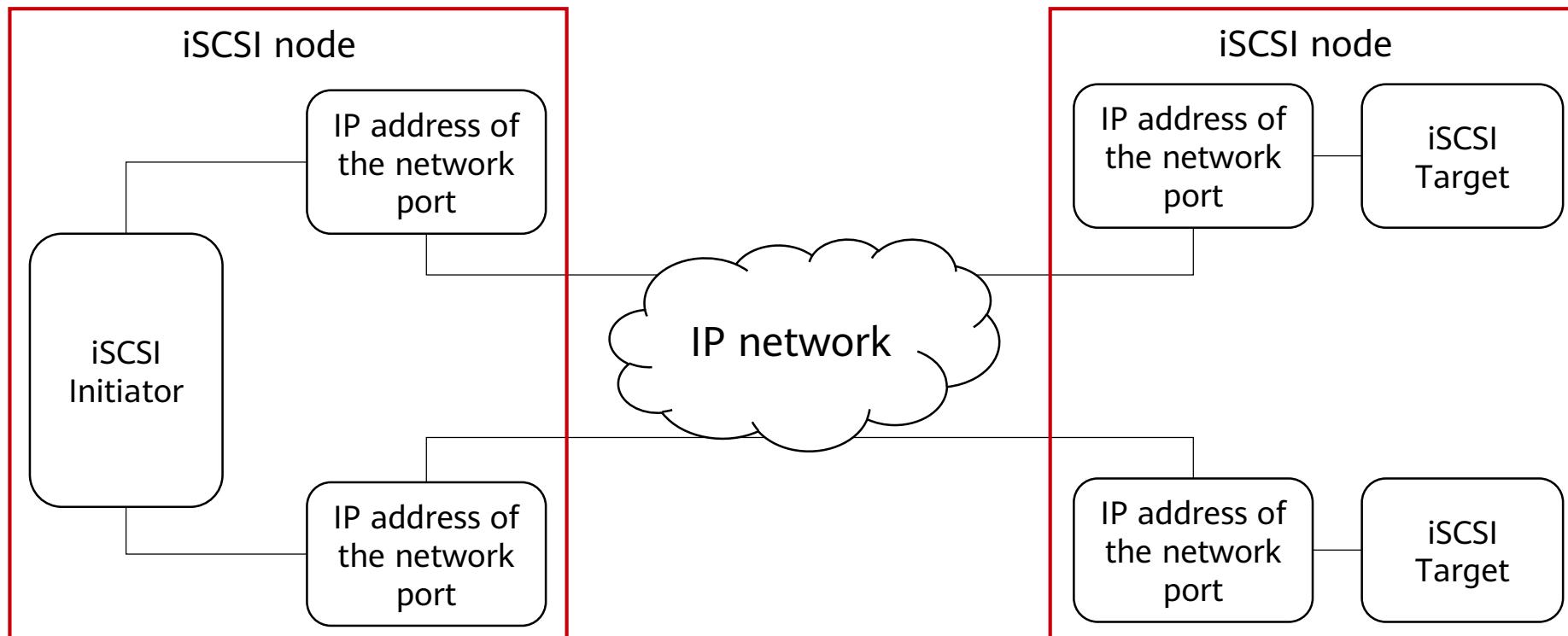
iSCSI Initiator and Target

- Initiator
 - The SCSI layer generates command descriptor blocks (CDBs) and transfers them to the iSCSI layer.
 - The iSCSI layer generates iSCSI protocol data units (PDUs) and sends them to the target over an IP network.
- Target
 - The iSCSI layer receives PDUs and sends CDBs to the SCSI layer.
 - The SCSI layer interprets CDBs and gives responses when necessary.

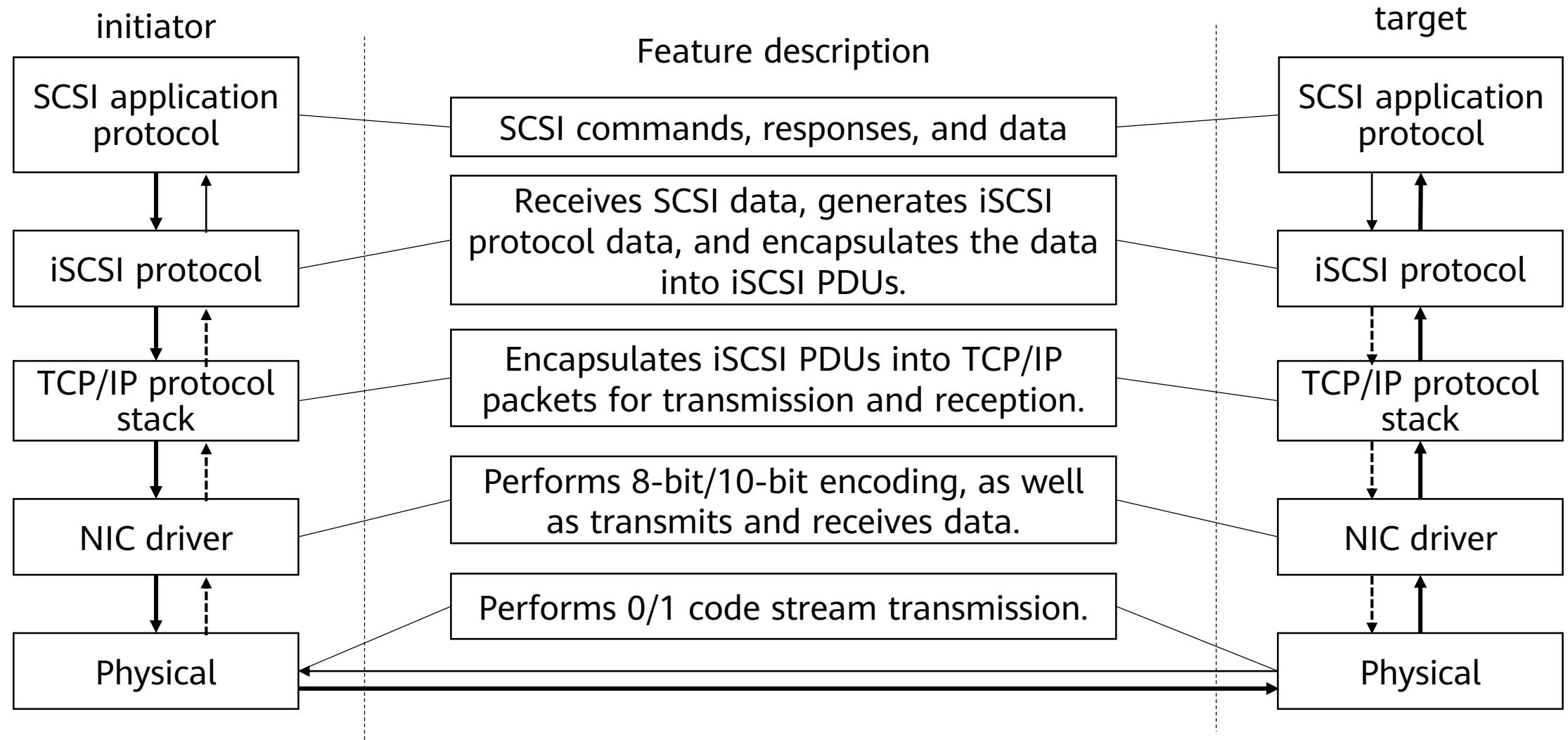


iSCSI Architecture

- iSCSI nodes encapsulate SCSI instructions and data into iSCSI packets and send the packets to the TCP/IP layer, where the packets are encapsulated into IP packets to be transmitted over an IP network.

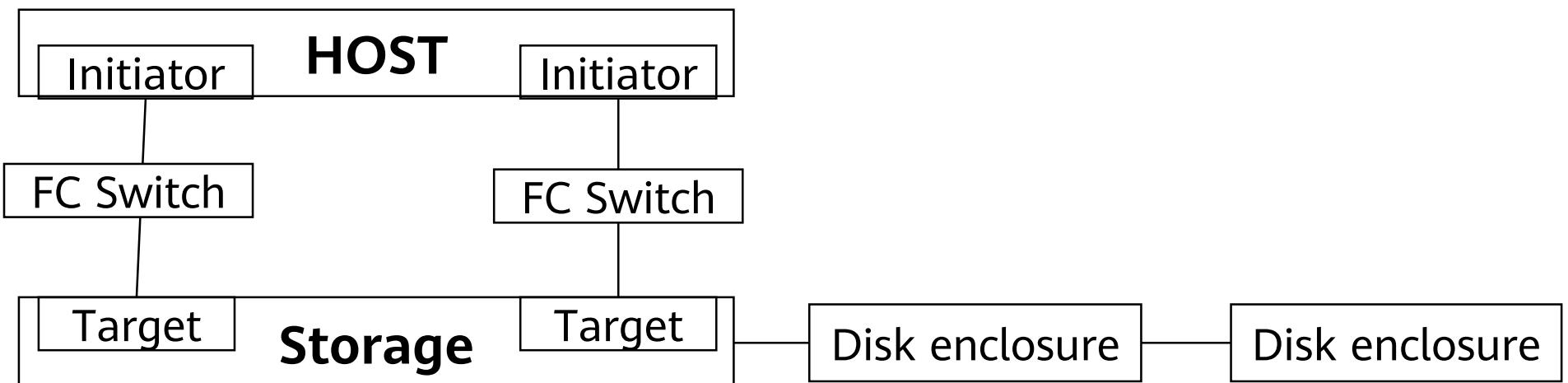


Relationships Between iSCSI and SCSI, TCP and IP

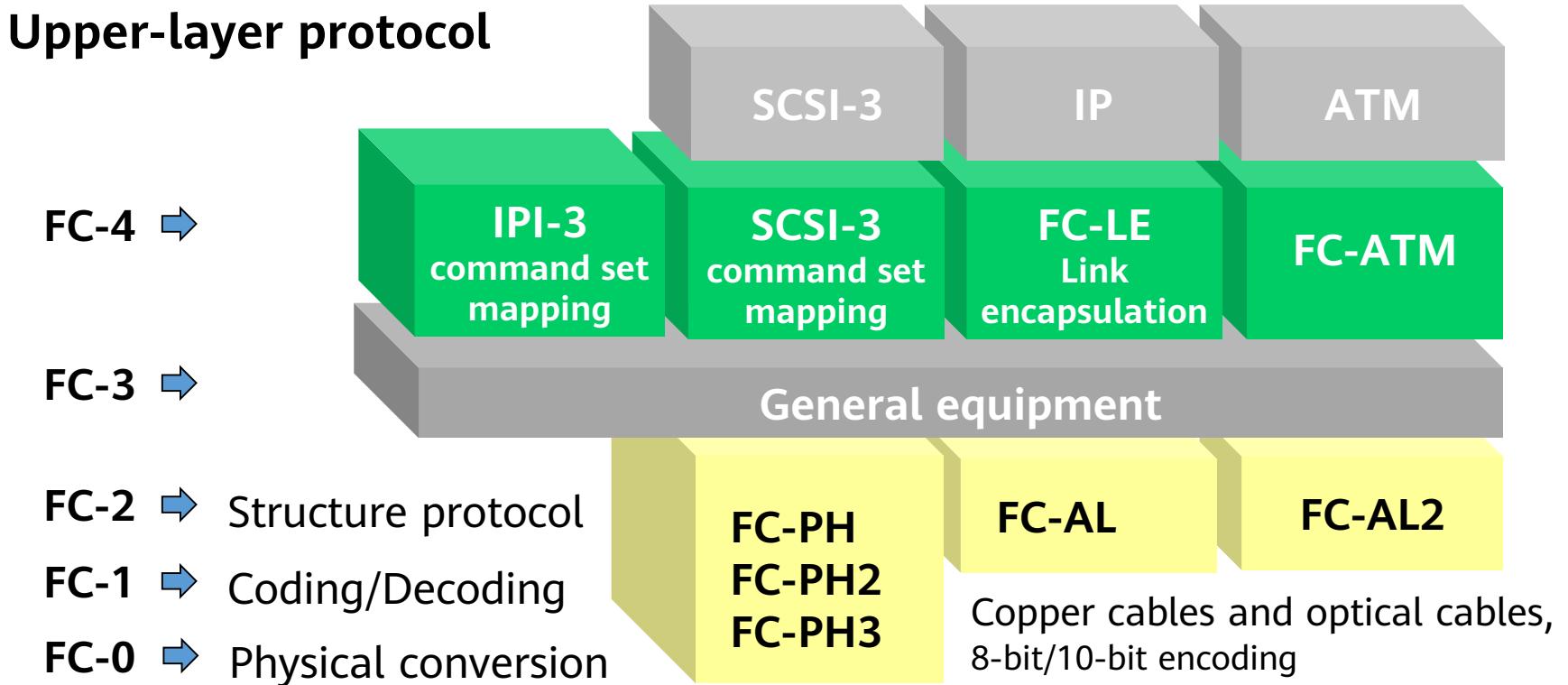


FC in Storage

- Fibre Channel (FC), also referred to as the FC protocol, FC network, or FC interconnection, delivers high performance for front-end host access on point-to-point and switch-based networks.
- FC brings the following advantages to the storage network:
 - Improved scalability
 - Increased transmission distance
 - High security

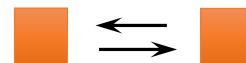


FC Protocol Structure



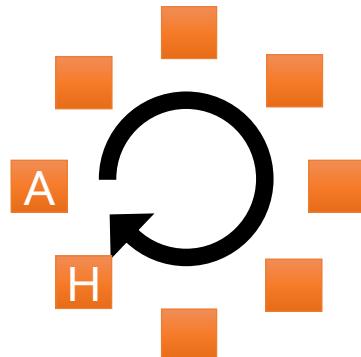
FC Topology

Point-to-point



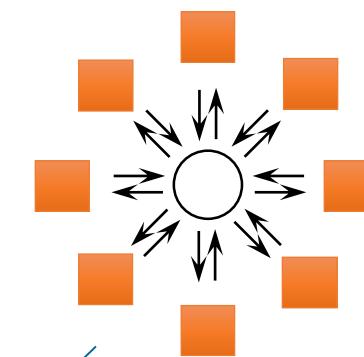
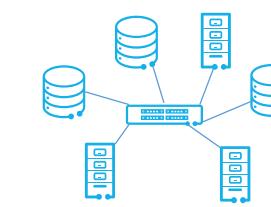
Only two devices can be connected.
(Direct connection)

FC-AL



Up to 127 devices can
be connected.

FC switching network

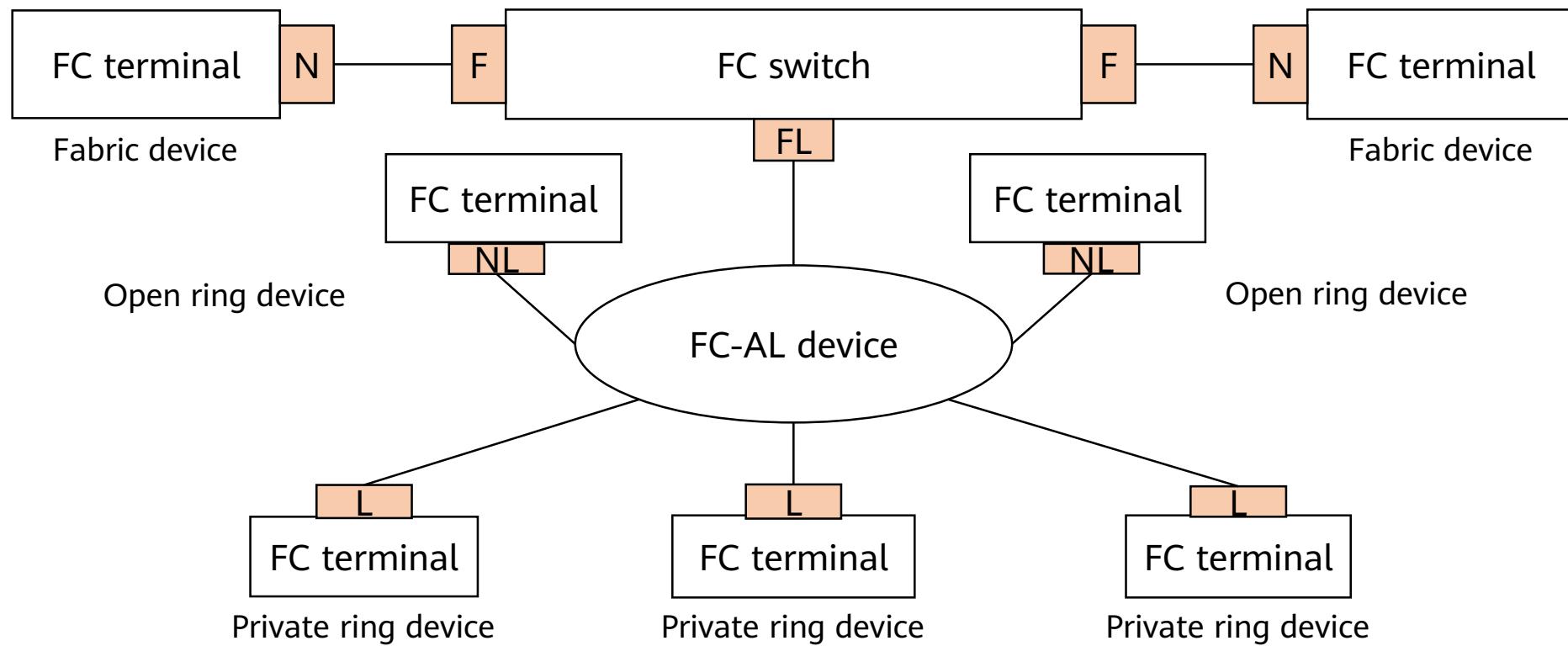


Most widely used technology

Up to 16 million devices
can be connected.

Seven Port Types of the FC Protocol

- There are seven types of ports in FC networks.



FC Adapter

- The FC host bus adapter (HBA) supports FC network applications and provides high-bandwidth and -performance storage network solutions.



Contents

1. SAN Protocols

- SCSI and SAS
- iSCSI and FC
- PCIe and NVMe
 - RDMA and RoCE

2. NAS Protocols

3. Object and HDFS Storage Protocols

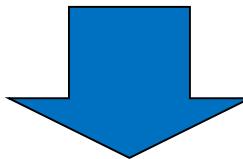
PCIe

- PCIe is short for PCI Express, which is a high-performance and high-bandwidth serial communication interconnection standard. It was first proposed by Intel and then developed by the Peripheral Component Interconnect Special Interest Group (PCI-SIG) to replace the bus-based communication architecture, such as PCI, PCI Extended (PCI-X), and Accelerated Graphics Port (AGP).



Why PCIe?

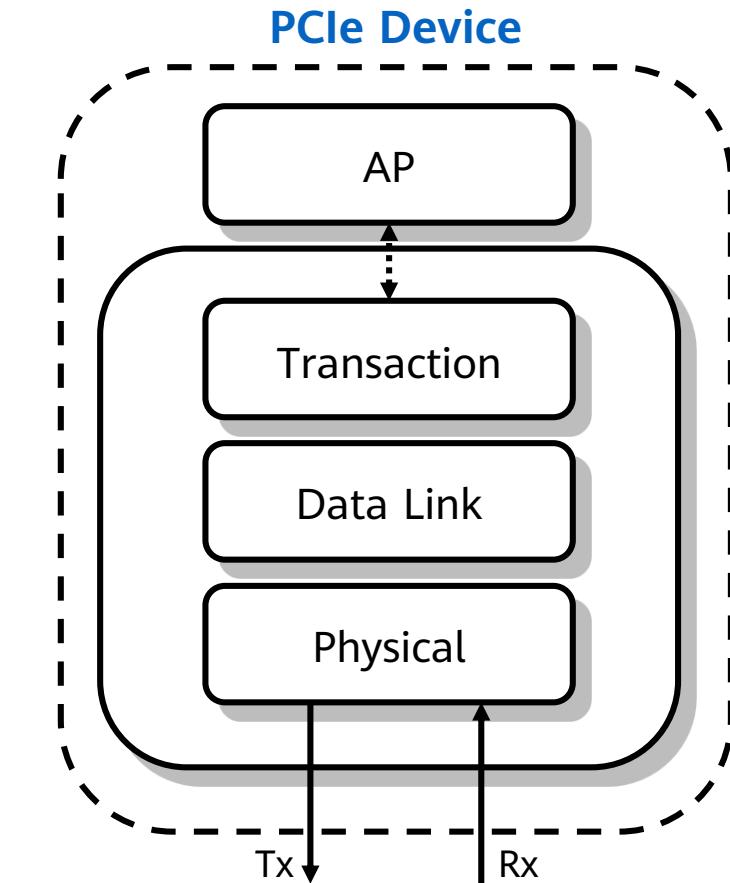
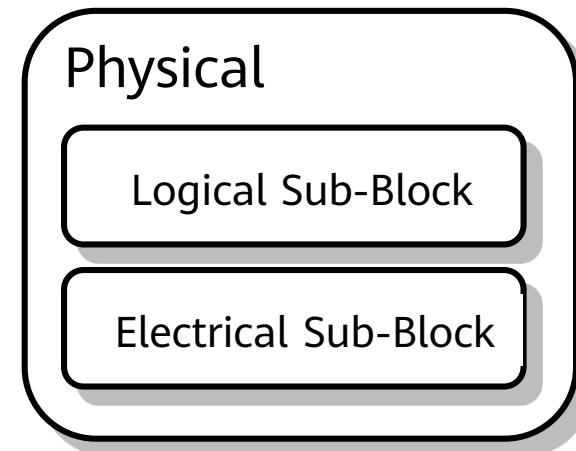
PCIe is used to significantly improve system throughput, scalability, and flexibility at lower production costs, which are almost impossible to achieve using the traditional bus-based interconnection.



High-performance and high-bandwidth serial interconnection standard: PCIe

PCIe Protocol Structure

- PCIe device layers comprise the physical, data link, transaction, and application layers.
 - Physical layer
 - Data link layer
 - Transaction layer
 - Application layer

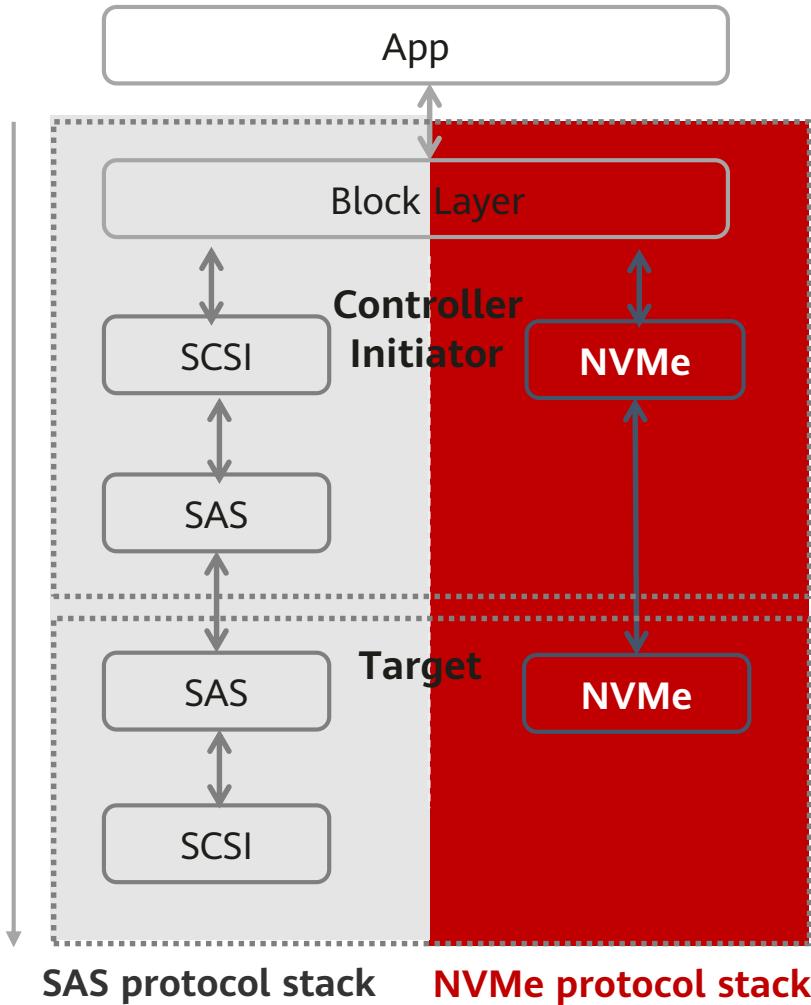


NVMe

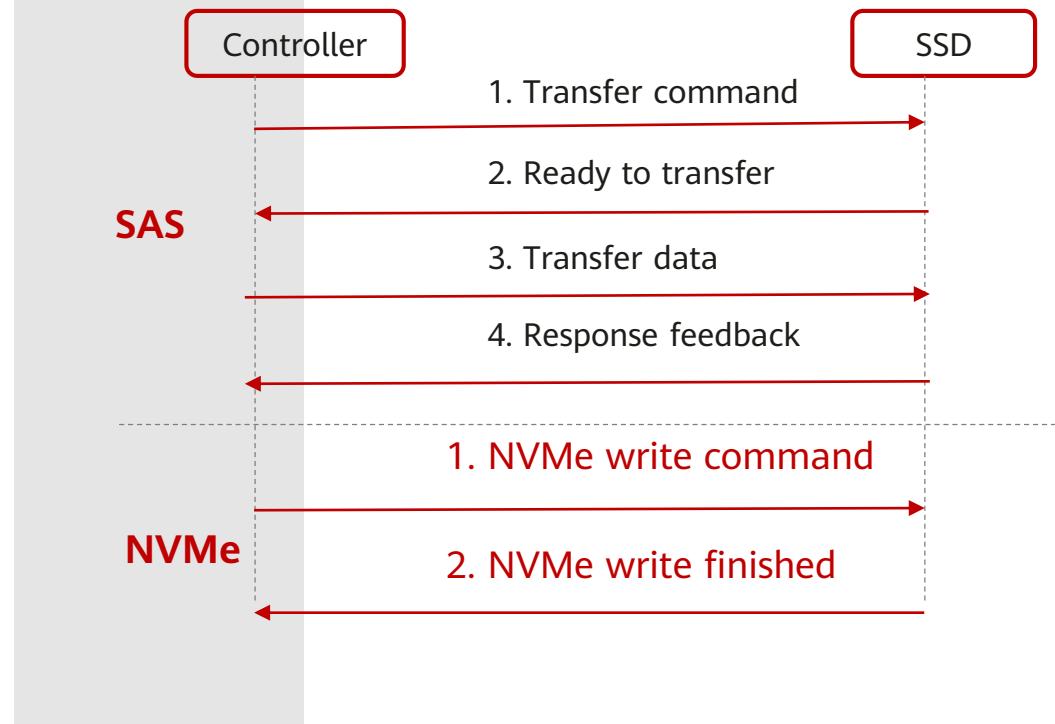
- NVMe is short for Non-Volatile Memory Express.
- The NVMe standard is oriented to PCIe SSDs. Direct connection from the native PCIe channel to the CPU can avoid the latency caused by communication between the external controller (PCH) of the SATA and SAS interface and the CPU.
- PCIe is an interface form and a bus standard, and NVMe is a standard interface protocol customized for PCIe SSDs.



NVMe Protocol Stack



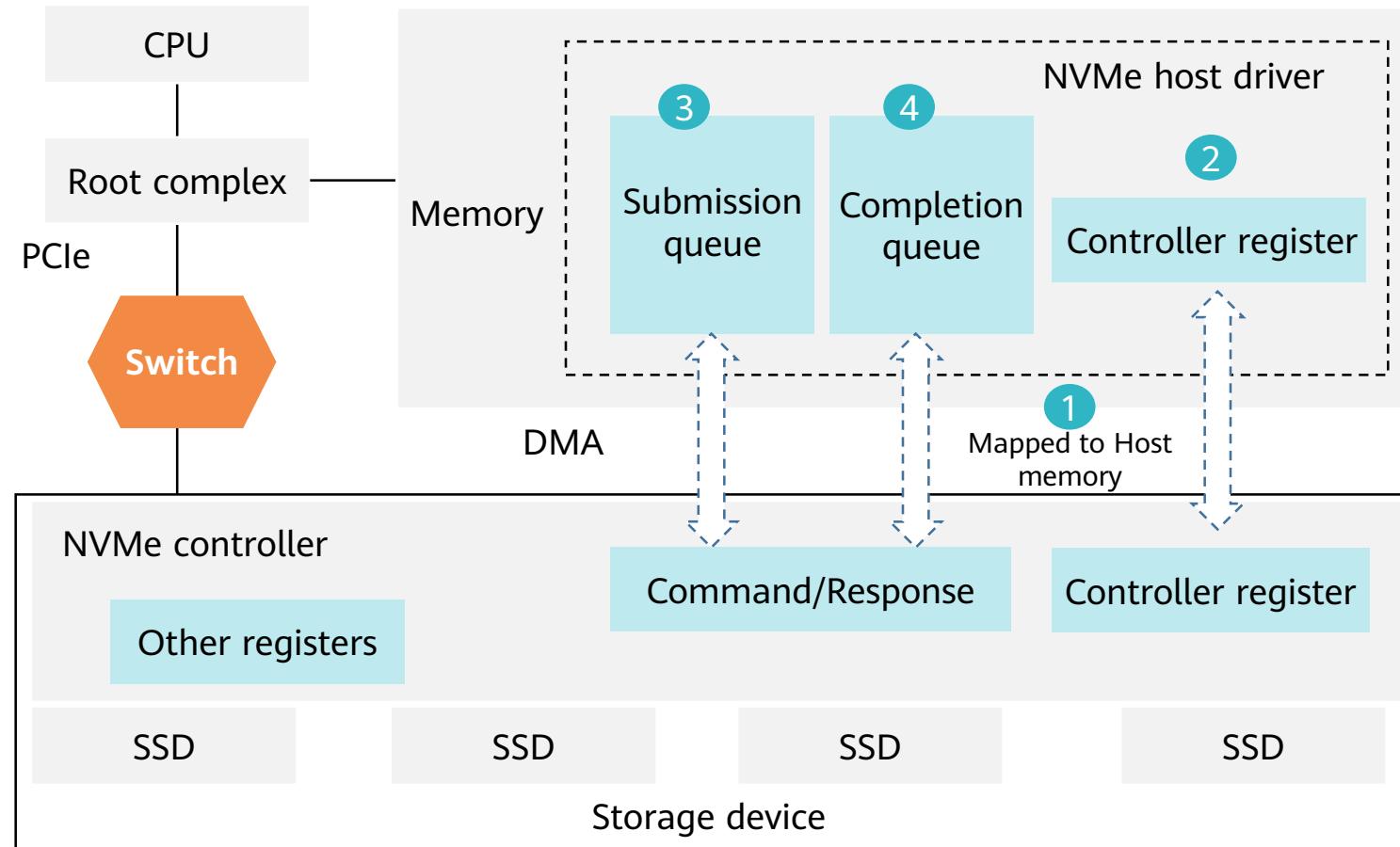
Reduced interaction: The number of communication interactions is reduced **from 4 to 2**, reducing the latency.



The average I/O latency when NVMe is used is less than that when SAS 3.0 is used.

NVMe Working Principle (DMA)

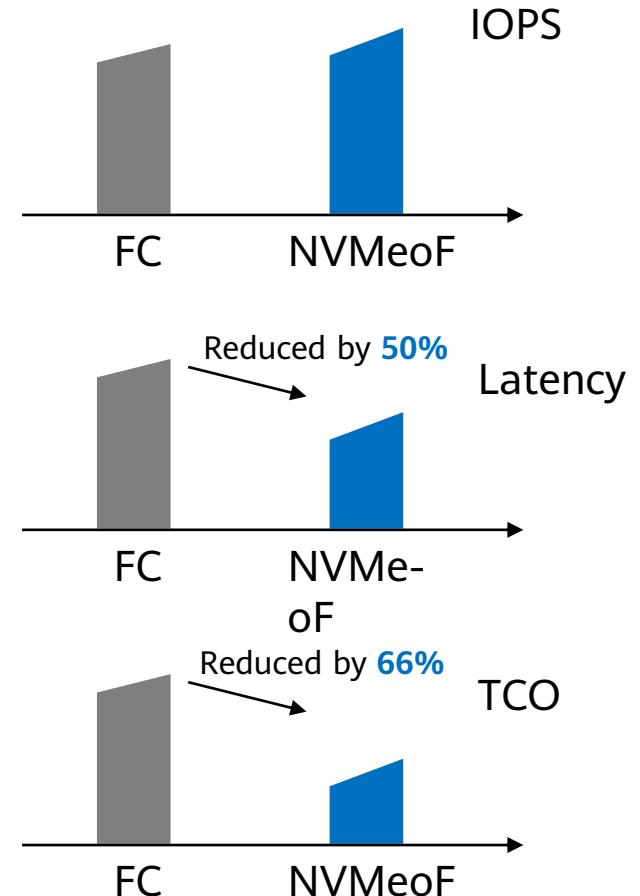
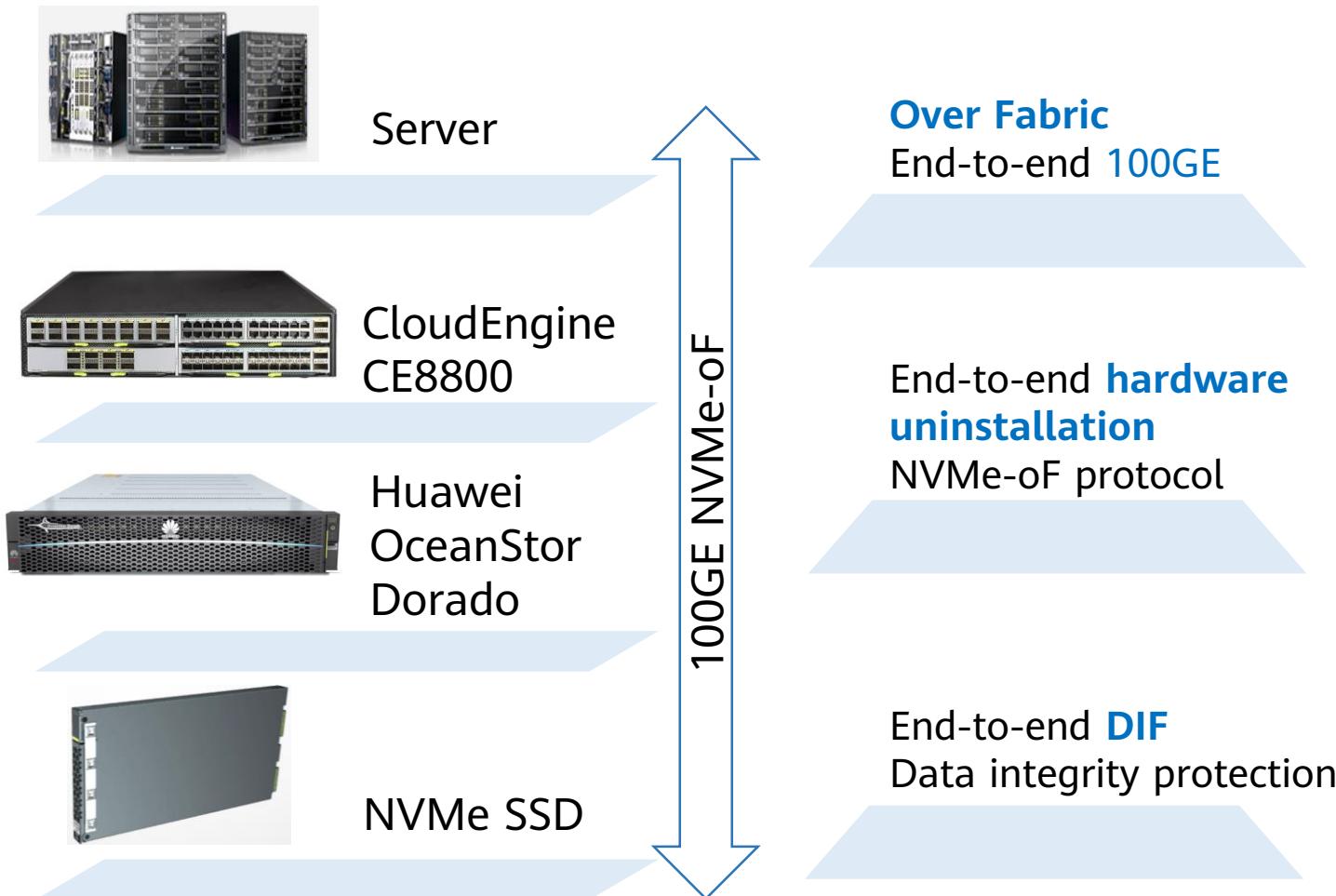
- NVMe uses Direct Memory Access (DMA) to transfer NVMe commands.



1. The NVMe driver maps the controller register of the storage device to the host memory.
2. The host controls the storage device and reads the device status using the register.
3. The host writes I/O requests into the submission queue for sending NVMe commands.
4. The NVMe device transfers data via DMA and writes completion to the host.

NVMe supports multi-queue data transmission with high concurrency and low latency.

Advantages and Application of NVMe



Contents

1. SAN Protocols

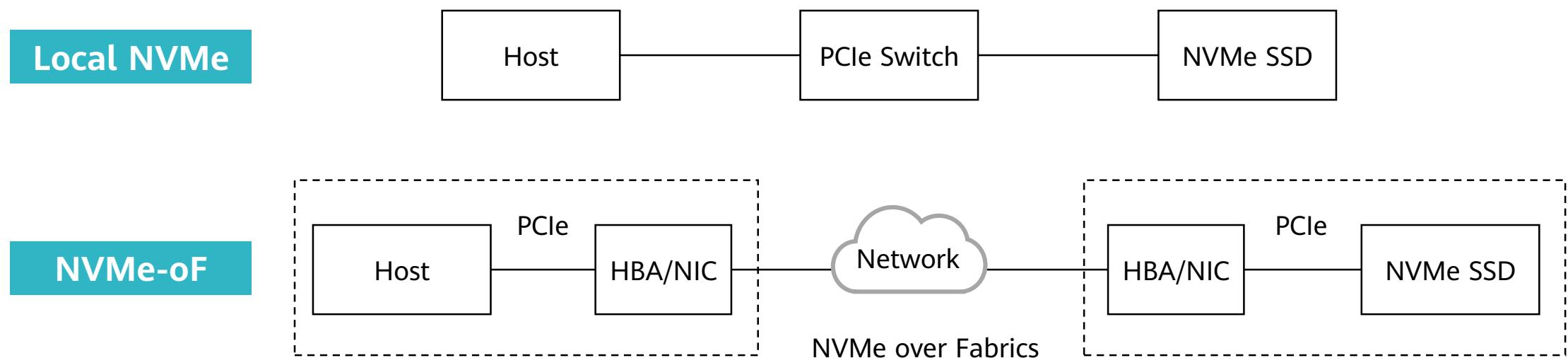
- SCSI and SAS
- iSCSI and FC
- PCIe and NVMe
- RDMA and RoCE

2. NAS Protocols

3. Object and HDFS Storage Protocols

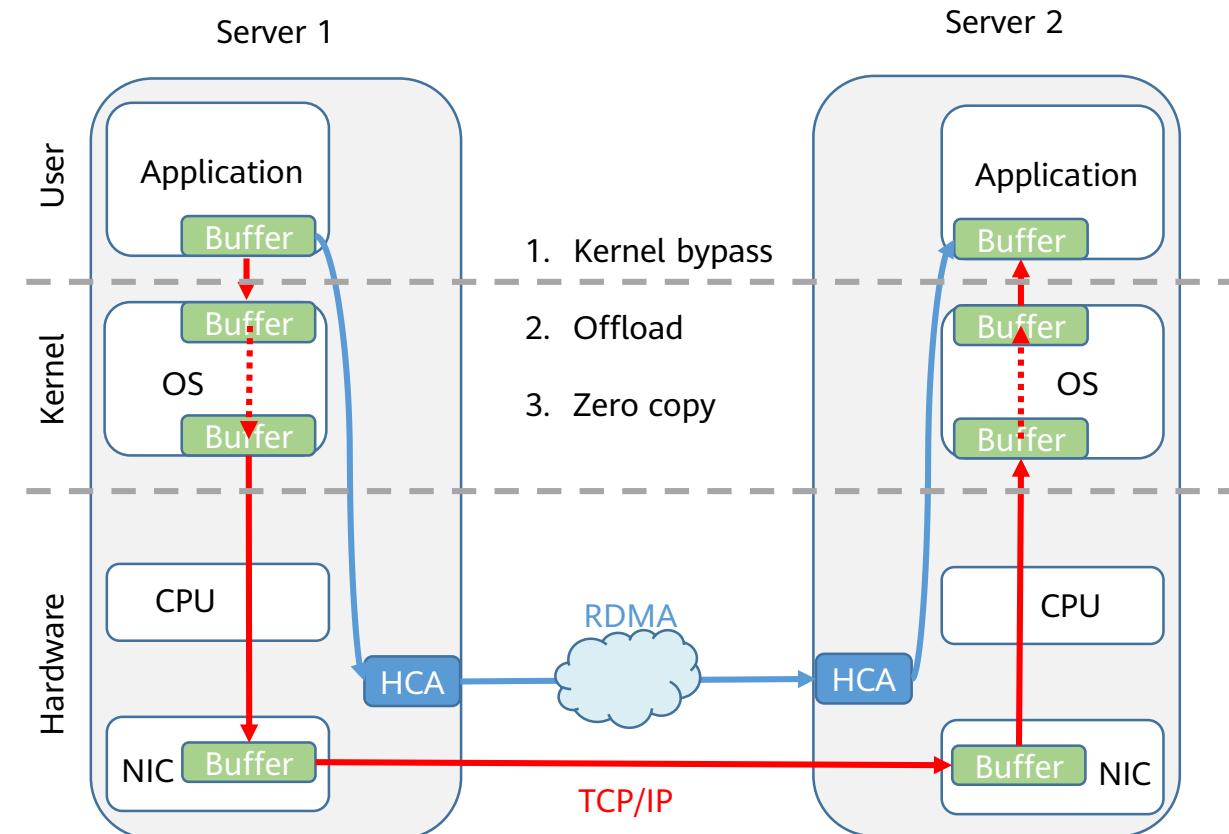
Introduction to NVMe over Fabrics

- NVMe over Fabrics (NVMe-oF) is an extension of the NVMe protocol to Ethernet and Fibre Channel. It allows accessing remote NVMe devices as if accessing local NVMe devices.
- To address the PCIe scalability problem, NVMe released a specification in 2016, which defines two types of fabric transports, implementing the end-to-end NVMe protocol.
 - **NVMe over Fabrics using RDMA**
 - NVMe over Fabrics using Fibre Channel (FC-NVMe)

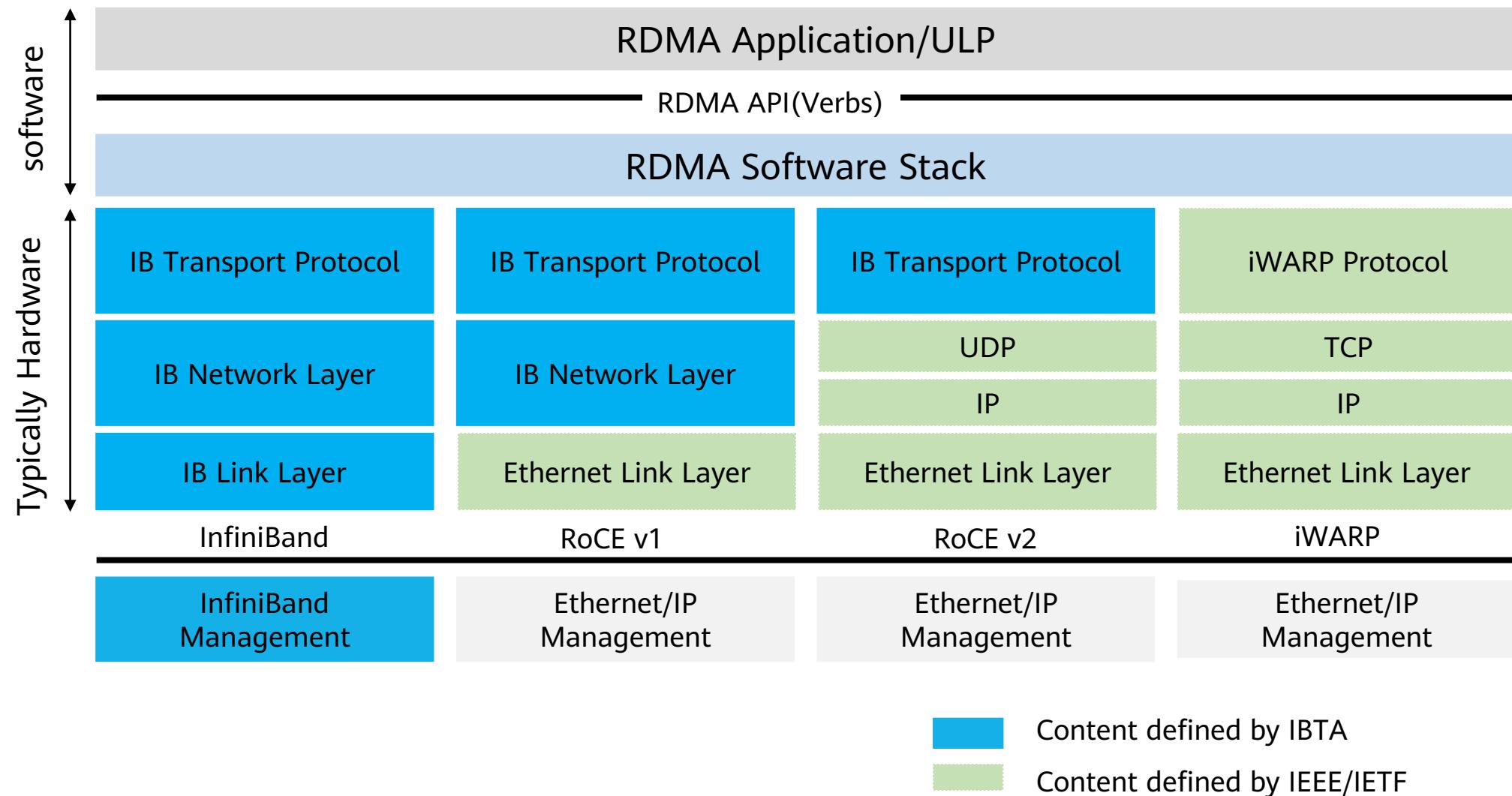


RDMA

- RDMA is short for Remote Direct Memory Access, which is a method of transferring data in a buffer between application software on two servers over a network.
 - Low latency
 - High throughput
 - Low CPU and OS resource occupancy

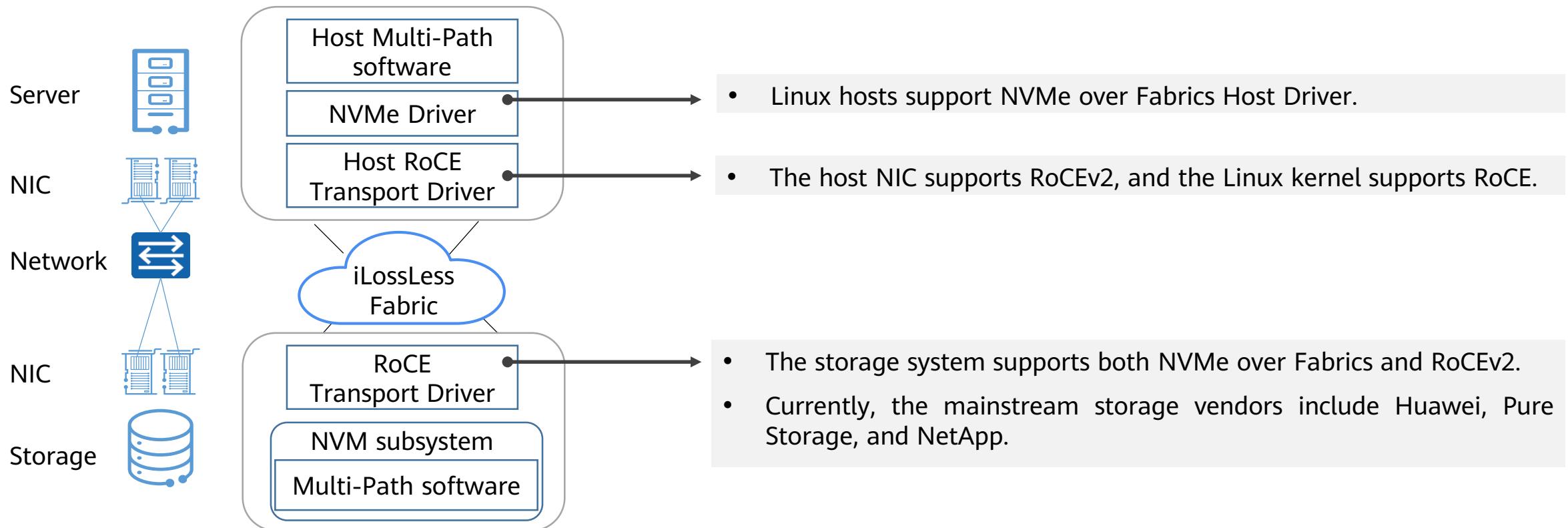


RDMA Bearer Network



Introduction to NVMe over RoCE

- NVMe over RoCE is a type of the NVMe over Fabrics protocol based on RDMA. It has been significantly optimized in terms of performance, cost, network management, and technology development, and is gradually becoming an optimal application of NVMe over Fabrics.



Contents

1. SAN Protocols

2. NAS Protocols

- File System

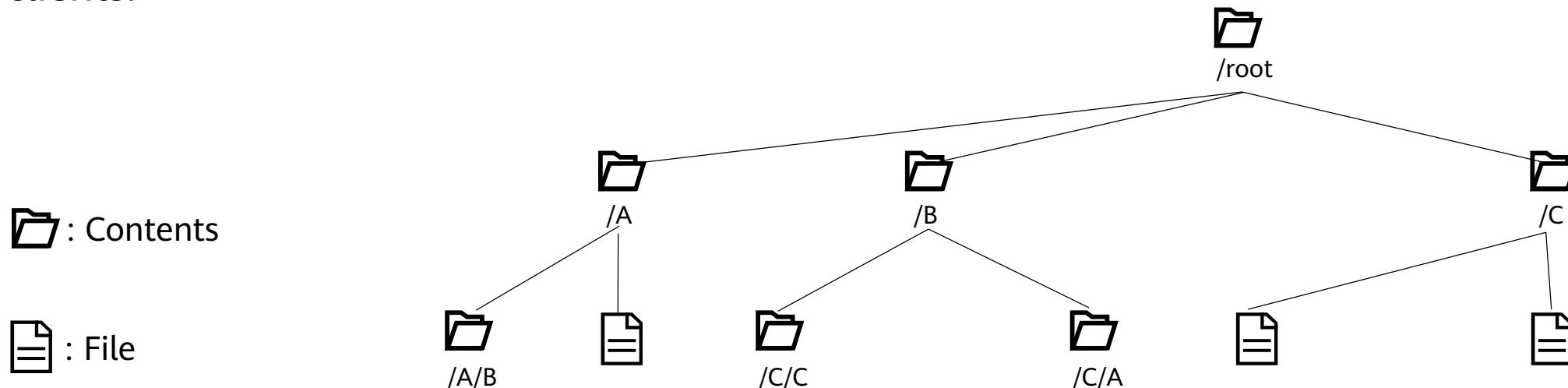
- CIFS, NFS and Cross-Protocol Access

- HTTP, FTP and NDMP

3. Object and HDFS Storage Protocols

File System

- A file system is used by a computer to manage and organize data in the form of files and directories. It forms a tree diagram, where the leaf node is a file, the intermediate node is a directory at each level, and the top level is the root directory.
- The file service is used to provide stable and reliable file sharing functions. It is a basic feature of the NAS storage system and supports shared file access using CIFS and NFS clients.



Contents

1. SAN Protocols

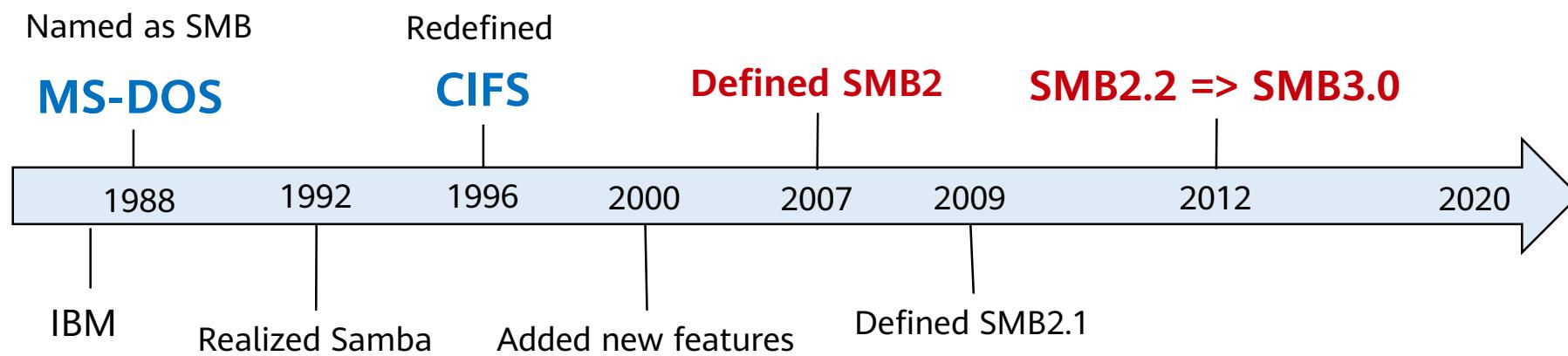
2. NAS Protocols

- File System
- CIFS, NFS and Cross-Protocol Access
- HTTP, FTP and NDMP

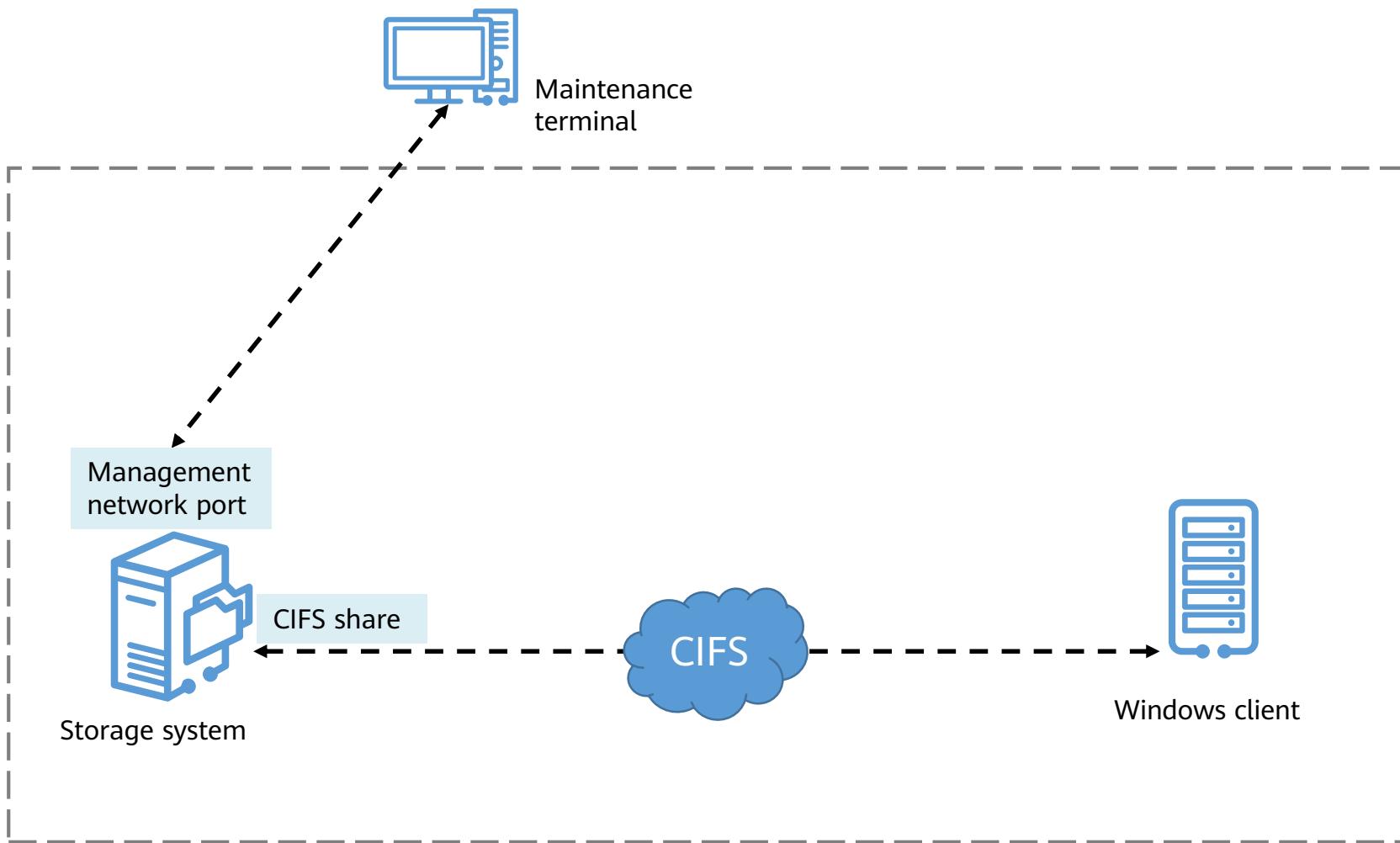
3. Object and HDFS Storage Protocols

CIFS Protocol

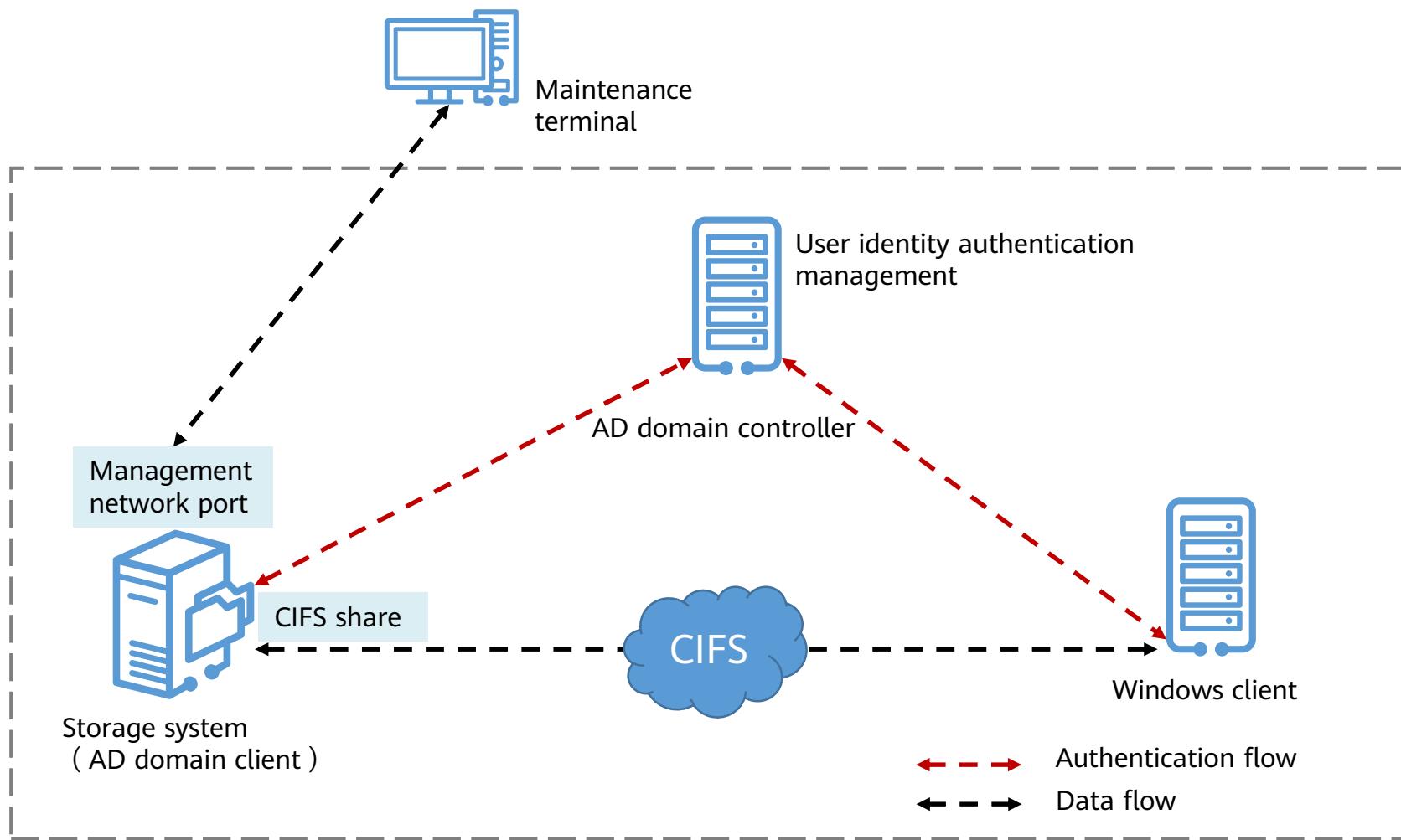
- In 1996, Microsoft renamed SMB to CIFS and added many new functions. Now, CIFS includes SMB1, SMB2, and SMB3.
- CIFS uses the client/server mode and TCP/IP and IPX/SPX basic network protocols. The CIFS share feature is primarily used by Windows-based clients to share files in a non-domain environment or an AD domain environment.



CIFS share in non-domain environments

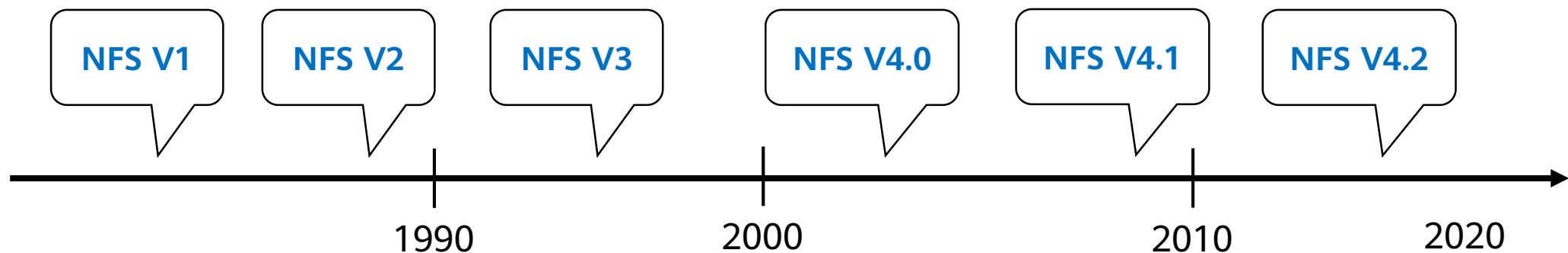


CIFS share in AD domain environments

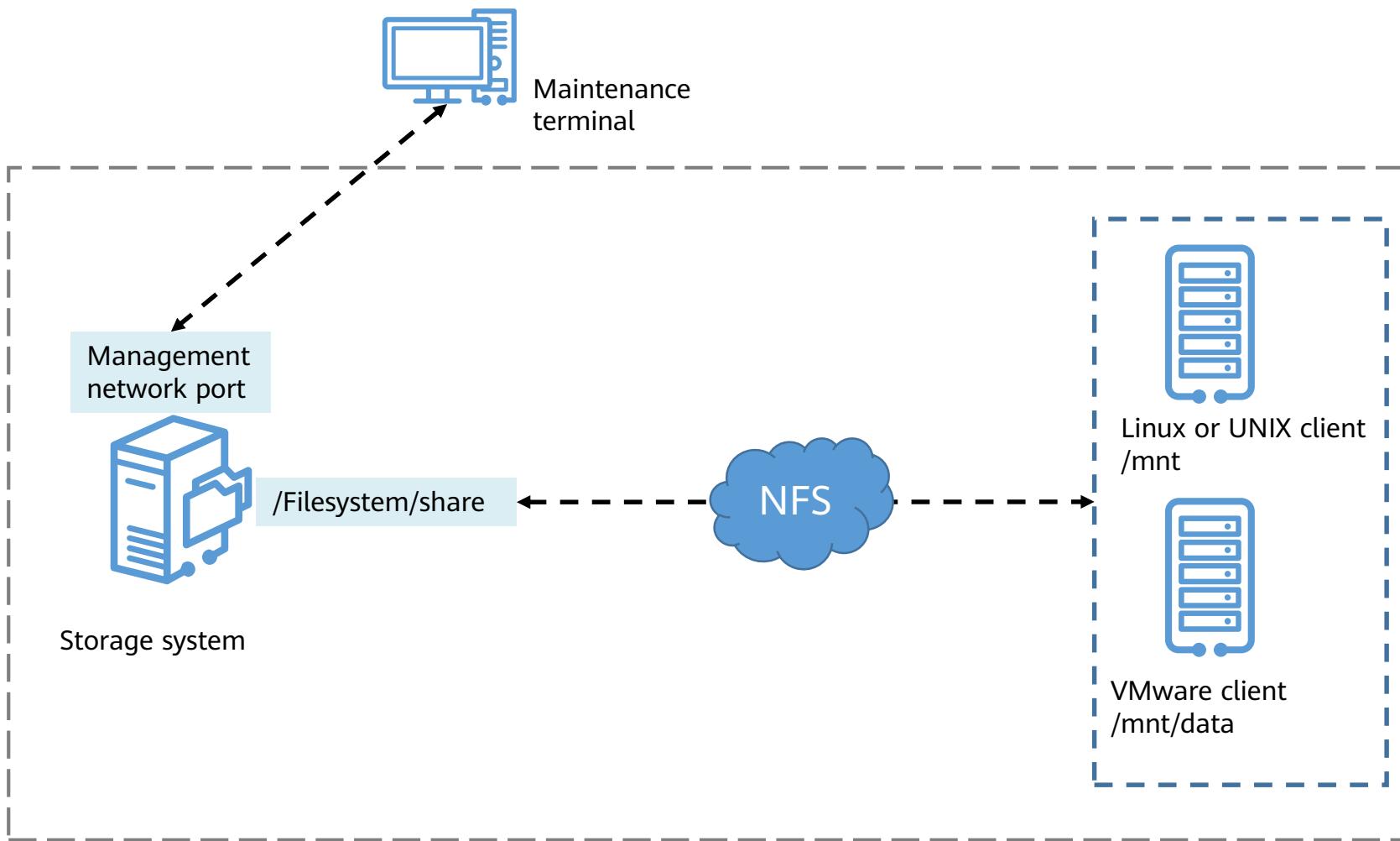


NFS Protocol

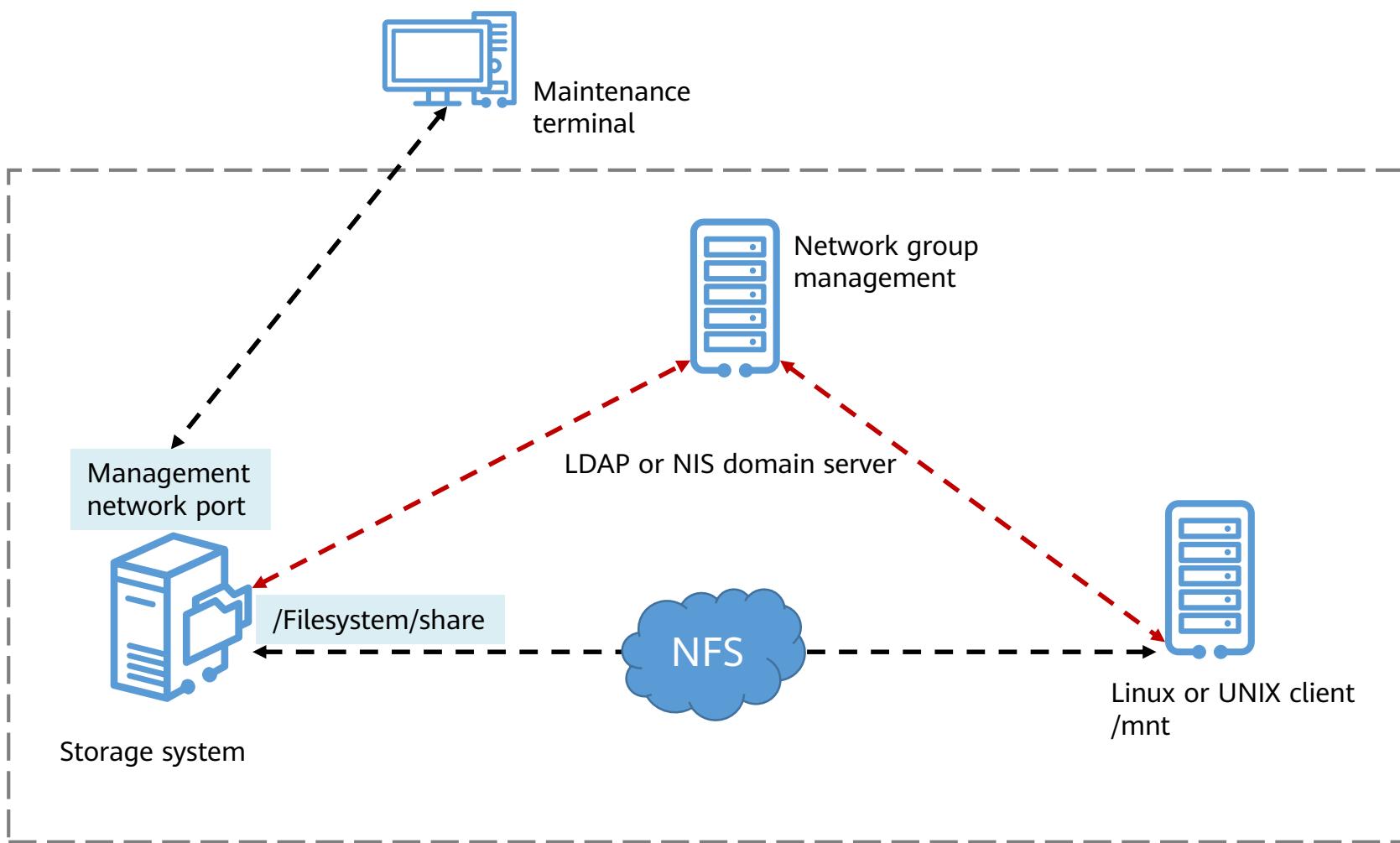
- NFS, or Network File System, is a protocol defined by the IETF and widely used in the Linux/Unix environment.
- NFS works based on the client/server architecture. A server provides clients with file system access, whereas clients access shared file systems. The NFS feature enables clients running a variety of operating systems to share files over a network. It applies to a wide range of network environments, including the non-domain environment, LDAP domain environment, and NIS domain environment.



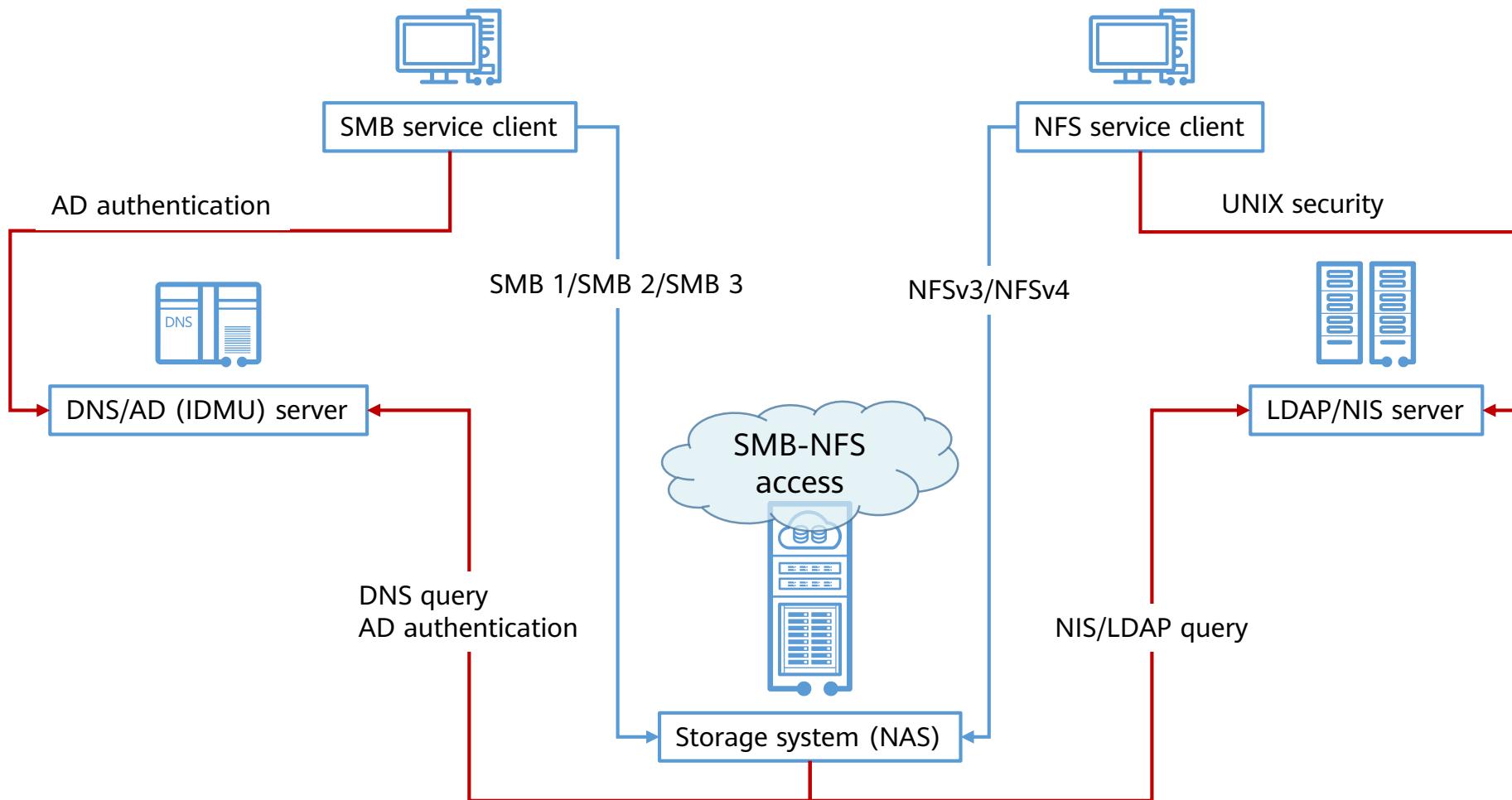
NFS share in a non-domain environment



NFS share in a domain environment



CIFS-NFS Cross-Protocol Access



Contents

1. SAN Protocols

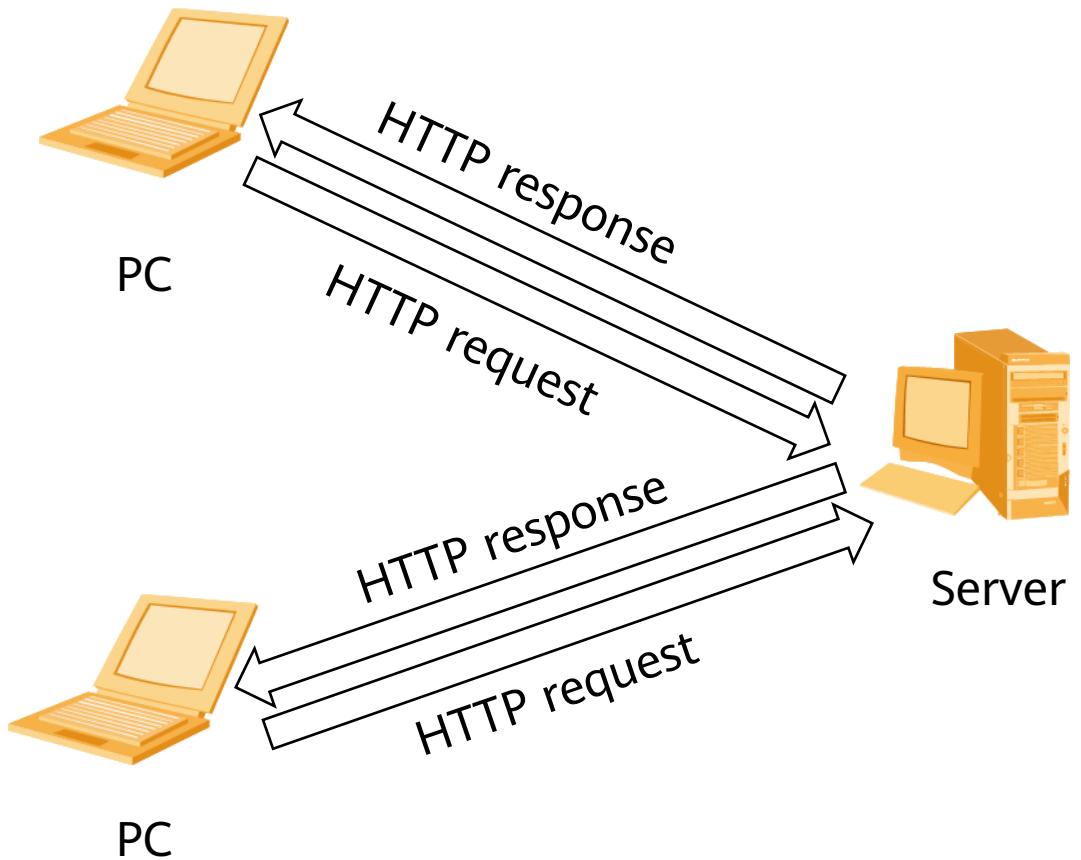
2. NAS Protocols

- File System
- CIFS, NFS and Cross-Protocol Access
- HTTP, FTP and NDMP

3. Object and HDFS Storage Protocols

HTTP Shared File System

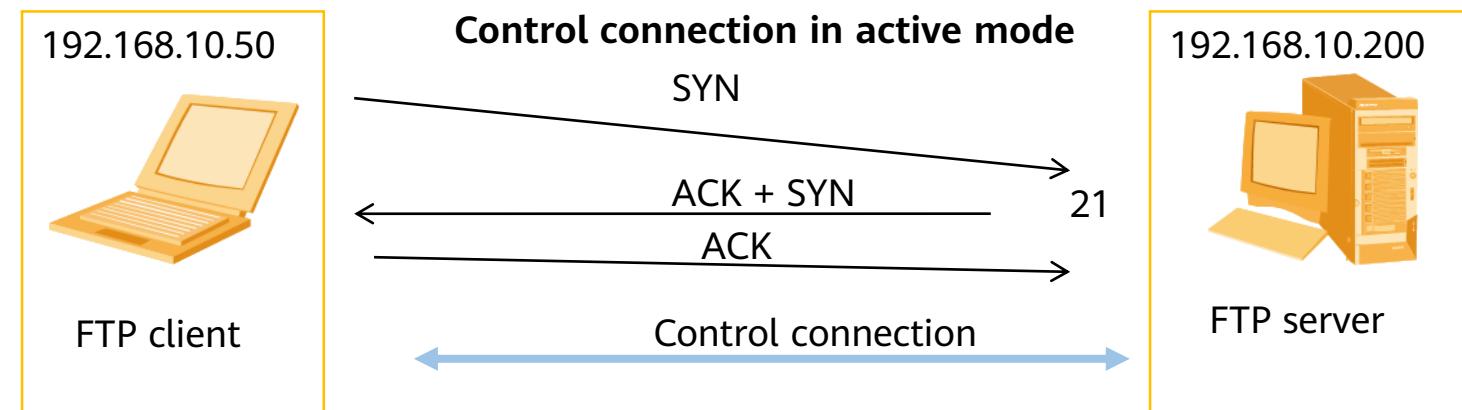
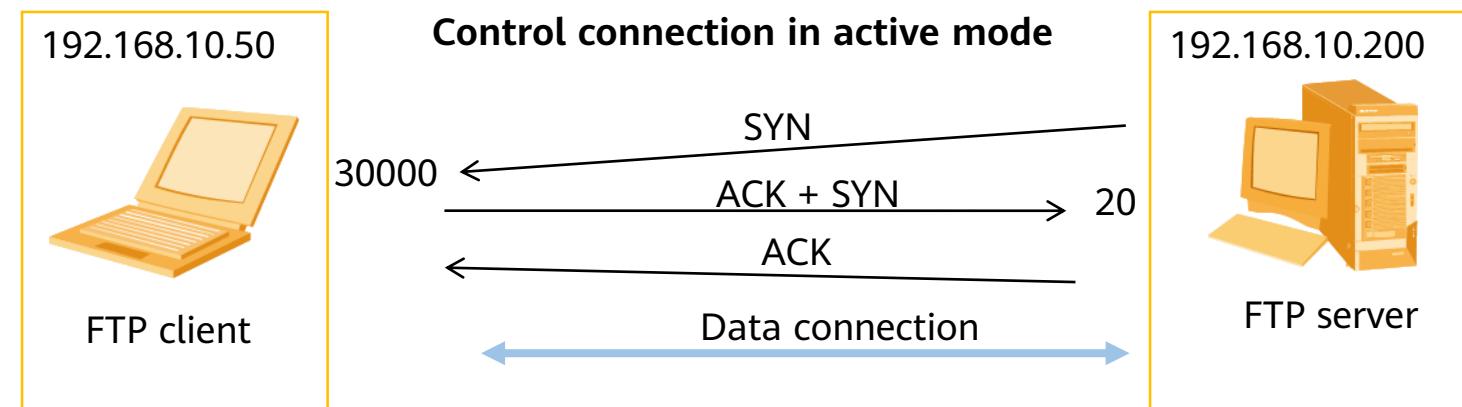
- The storage system supports the HTTP shared file system. With the HTTPS service enabled, you can share a file system in HTTPS mode.
- Shared resource management is implemented based on the WebDAV protocol. As an HTTP extension protocol, WebDAV allows clients to copy, move, modify, lock, unlock, and search for resources in shared directories on servers.



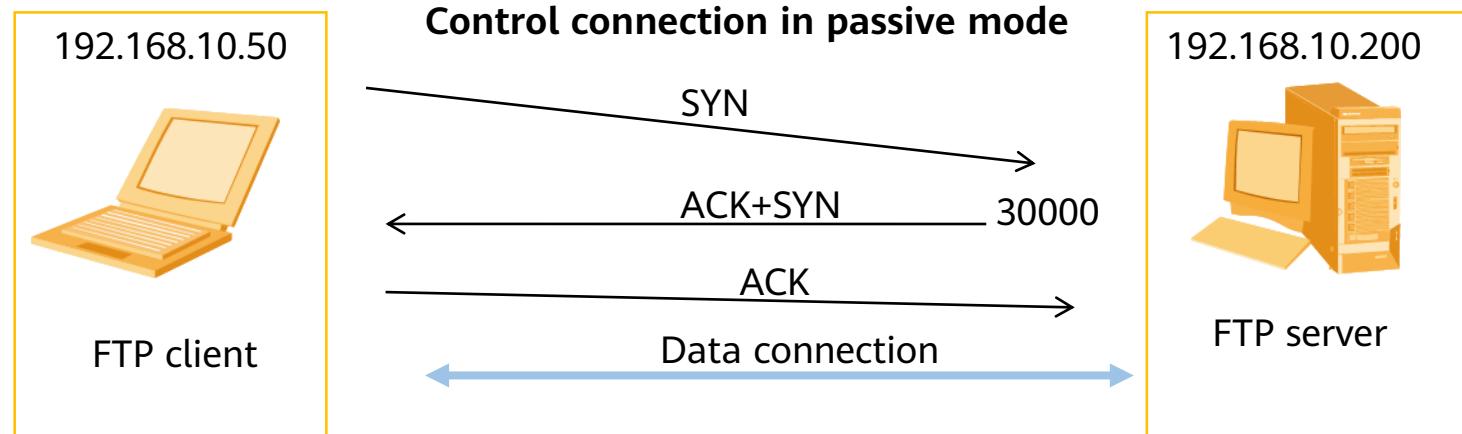
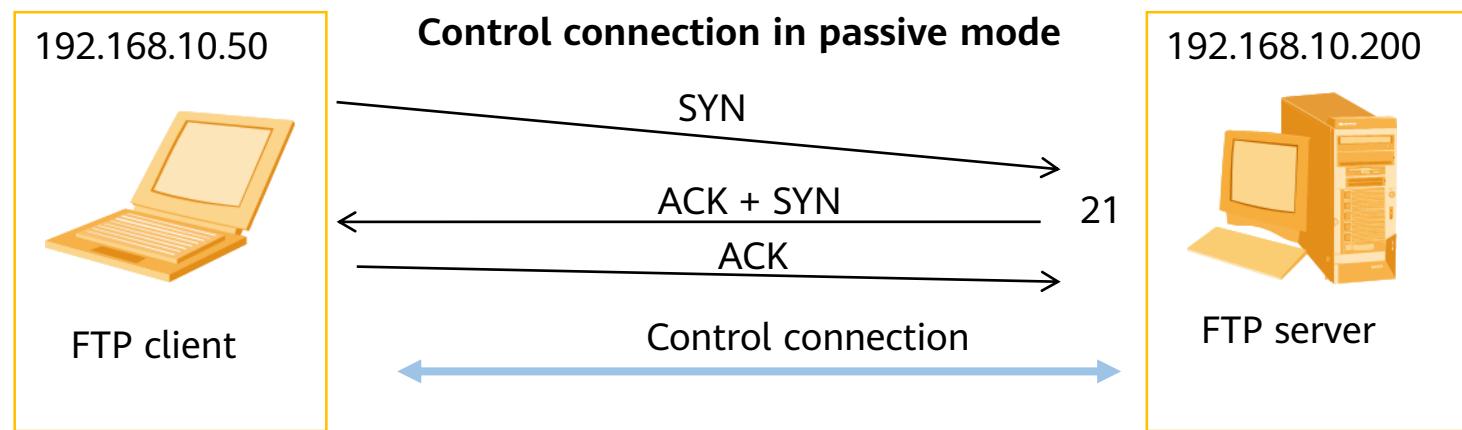
File Transfer Protocol (FTP)

- FTP is a universal protocol for transferring files between remote servers and local clients over an IP network.
- It belongs to the application layer in the TCP/IP protocol suite and employs TCP ports 20 and 21 to transfer files between remote servers and local clients. Port 20 is used to transfer data, and port 21 is used to transfer control messages. RFC 959 describes the basic FTP operations.
- The storage system supports the FTP service. The FTPS server function on a device allows users to securely access a remote device by using the FTP client.
- FTP works in either of the following modes:
 - Active mode (PORT): The FTP server creates a data connection request. This mode is inapplicable when the FTP client is behind a firewall, for example, the FTP client resides on a private network.
 - Passive mode (PASV): The FTP client creates a data connection request. This mode is inapplicable when the FTP server does not allow the FTP client to connect to its high-order ports (usually, the port IDs are larger than 1024).

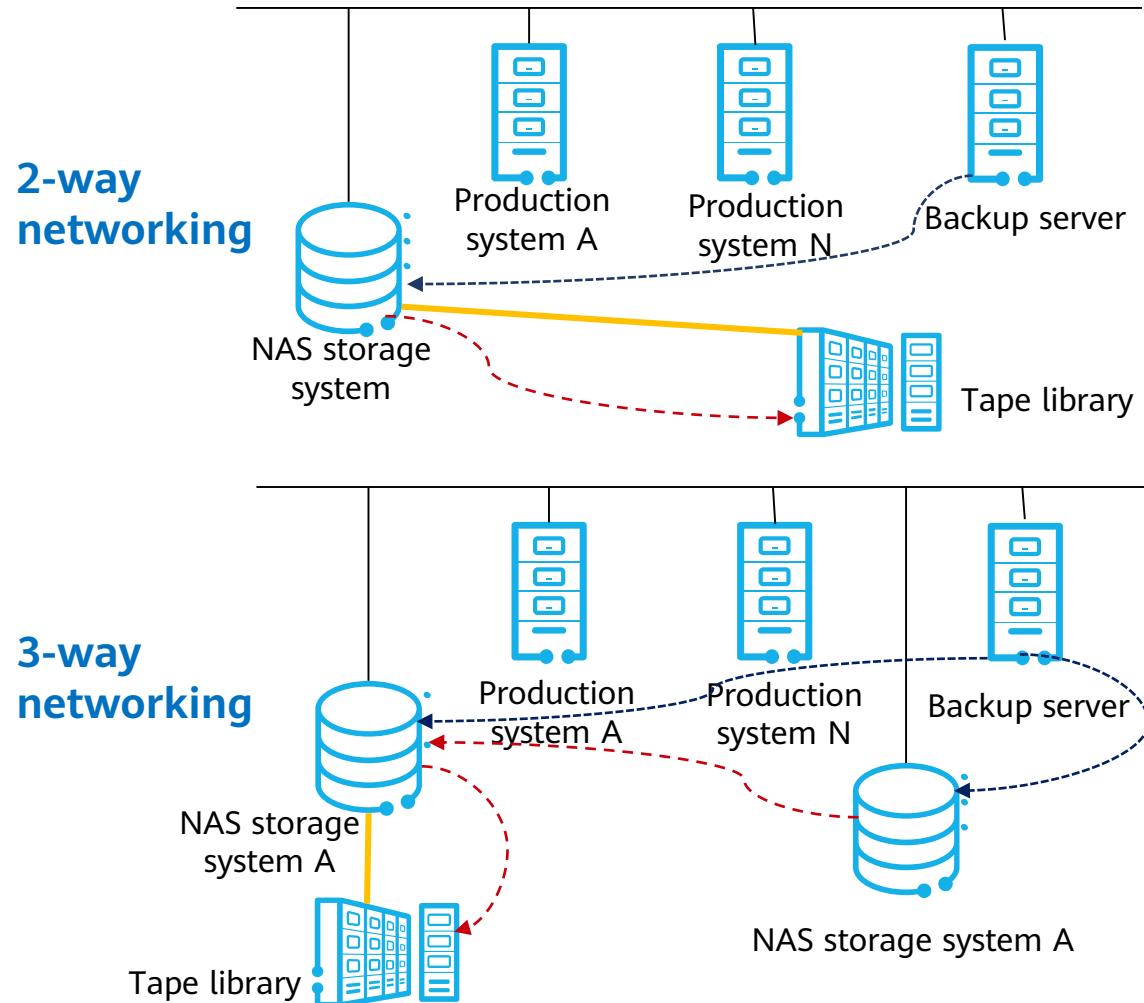
Active Mode of the FTP Server



Passive Mode of the FTP Server



NDMP Protocol



- The NDMP protocol is designed for the data backup system of NAS devices. It enables NAS devices, without any backup client agent, to send data directly to the connected disk devices or the backup servers on the network for backup.

- There are two networking modes for NDMP:
 - 2-way
 - 3-way

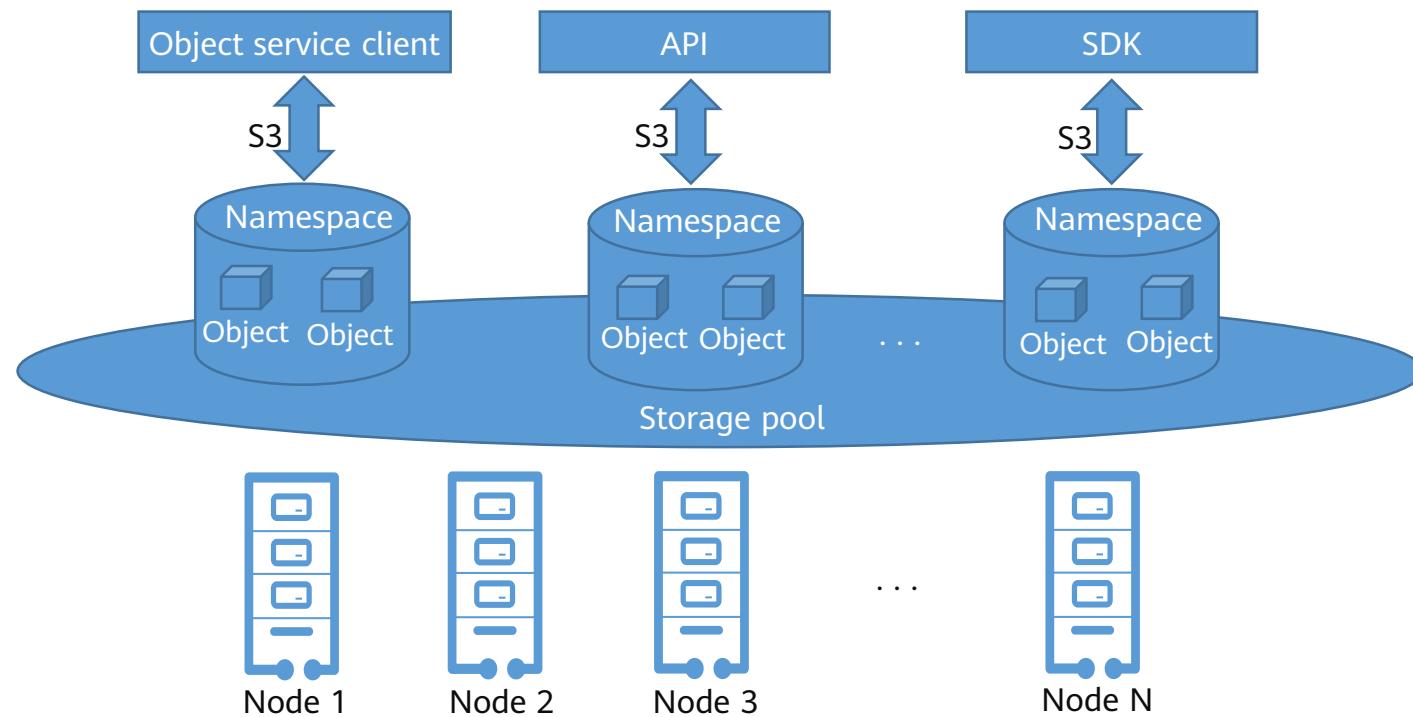
— Fiber Channel
— Ethernet
←— Backup data flow
←— Control flow

Contents

1. SAN Protocols
2. NAS Protocols
- 3. Object and HDFS Storage Protocols**
 - Object Storage Protocol
 - HDFS Storage Protocol

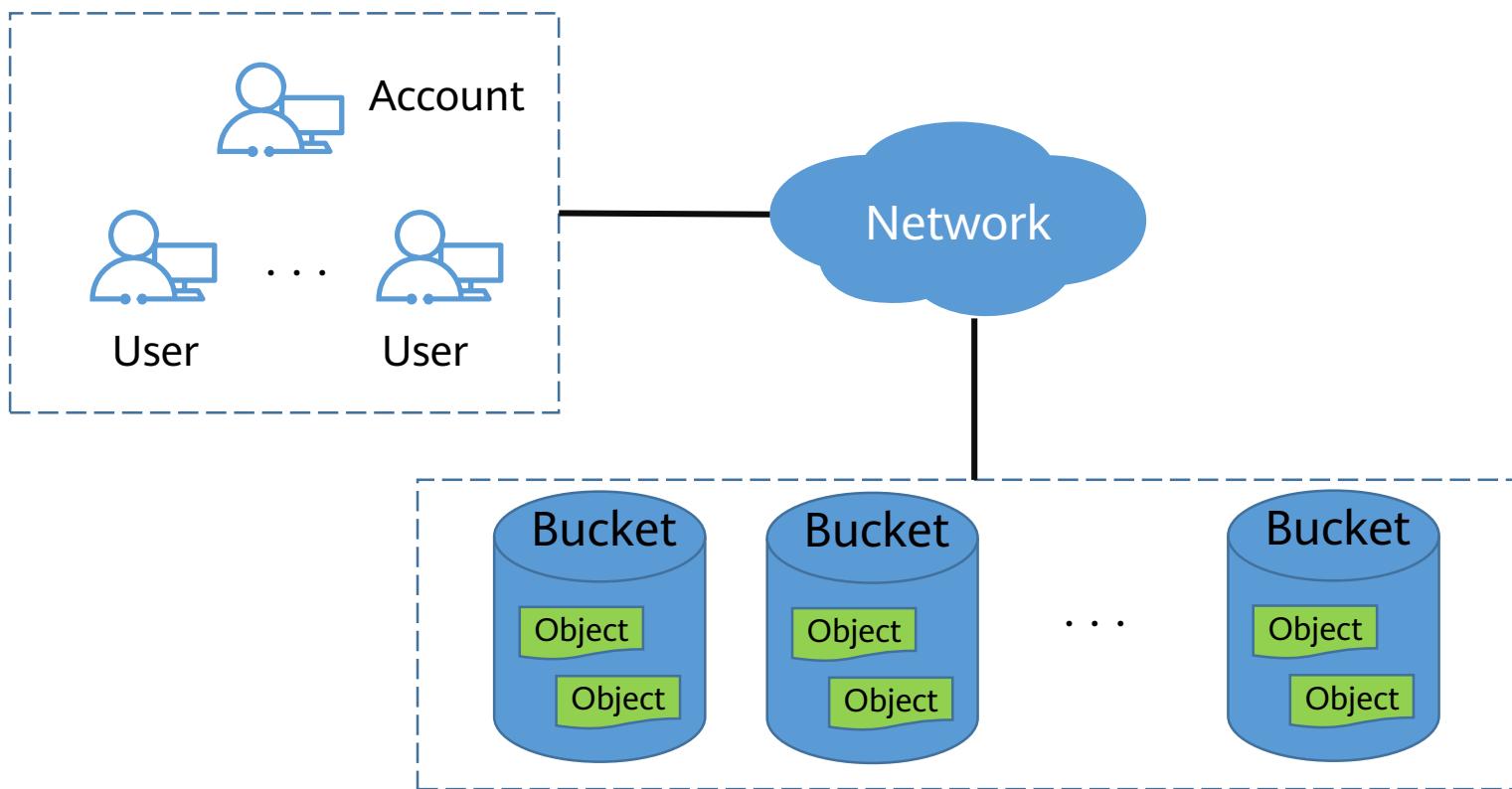
Object Service

- The object service is an object-based mass data storage service offering scalable, secure, reliable, and cost-effective data storage capabilities.
- The object service provides standard S3 APIs, which are HTTP/HTTPS-based REST APIs. Users can use object service clients, APIs, and SDKs to easily manage and use object service data and develop various types of upper-layer applications.



S3 Concepts

- The following figure shows the relationship among buckets, objects, accounts, and users.



RESTful

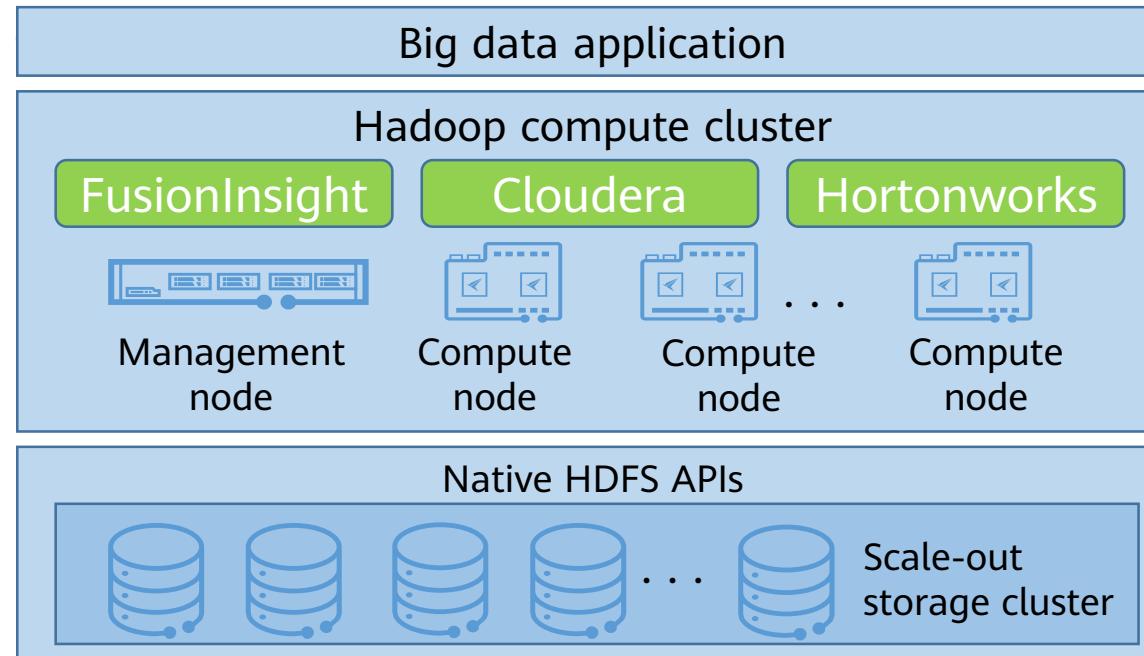
- REST
 - REST is short for REpresentational State Transfer. It indicates the state transfer of resources in a certain form on the network.
 - Resource: resource, that is, data. For example, newsfeed and friends.
 - Representational: a representation form, for example, an image or a video.
 - State Transfer: status change. This is implemented through HTTP verbs.
 - It uses the HTTP protocol and URI to add, delete, modify, and query resources using the client/server model.
 - REST is not a specification, but an architecture for network applications. It can be regarded as a design mode which is applied to the network application architecture.
- RESTful
 - An architecture complying with the REST principle is called a RESTful architecture.
 - It provides a set of software design guidelines and constraints for designing software for interaction between clients and servers. RESTful software is simpler and more hierarchical and facilitates the cache mechanism.

Contents

1. SAN Protocols
2. NAS Protocols
- 3. Object and HDFS Storage Protocols**
 - Object Storage Protocol
 - HDFS Storage Protocol

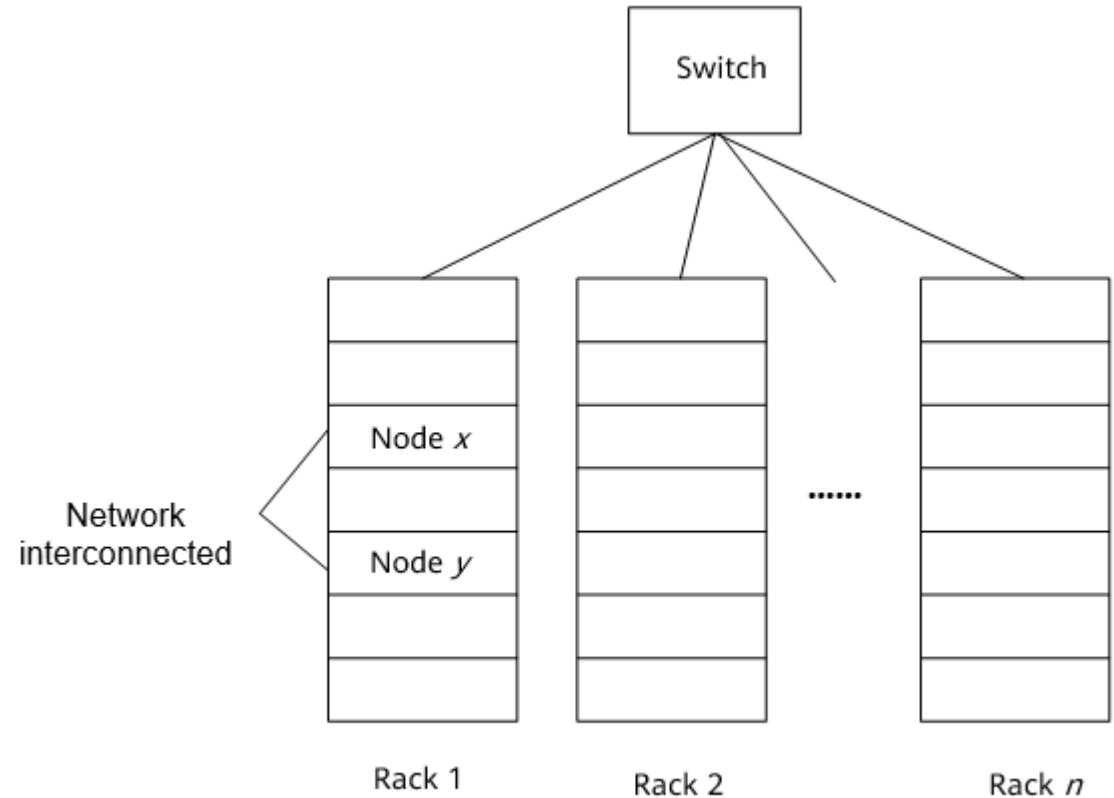
HDFS Service

- The HDFS service provides an HDFS decoupled storage-compute solution based on native HDFS. The solution implements on-demand configuration of storage and compute resources, provides consistent user experience, and helps reduce the total cost of ownership (TCO). It can coexist with the legacy coupled storage-compute architecture.
- The HDFS service provides native HDFS interfaces to interconnect with big data platforms, such as FusionInsight, Cloudera CDH, and Hortonworks HDP, to implement big data storage and computing and provide big data analysis services for upper-layer big

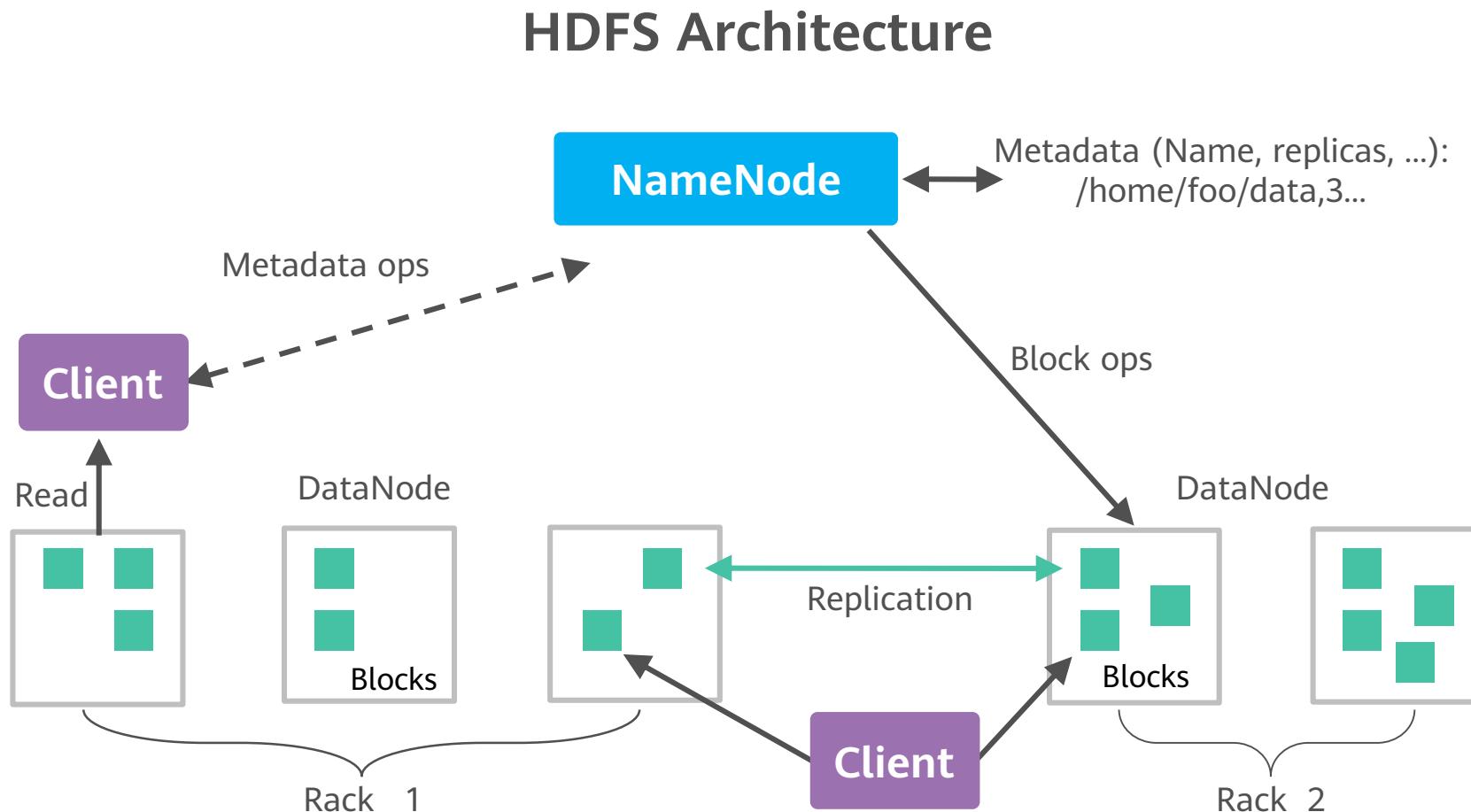


Distributed File System

- The distributed file system stores files on multiple computer nodes. Thousands of computer nodes form a computer cluster.
- Currently, the computer cluster used by the distributed file system consists of common hardware, which greatly reduces the hardware overhead.
- The Hadoop Distributed File System (HDFS) is a distributed file system running on universal hardware and is designed and developed based on the GFS paper.



HDFS Architecture



HDFS Communication Protocol

- HDFS is a distributed file system deployed on a cluster. Therefore, a large amount of data needs to be transmitted over the network.
 - All HDFS communication protocols are based on the TCP/IP protocol.
 - The client initiates a TCP connection to the NameNode through a configurable port and uses the client protocol to interact with the NameNode.
 - The NameNode and the DataNode interact with each other by using the DataNode protocol.
 - The interaction between the client and the DataNode is implemented through the Remote Procedure Call (RPC). In design, the name node does not initiate an RPC request, but responds to RPC requests from the client and DataNode.

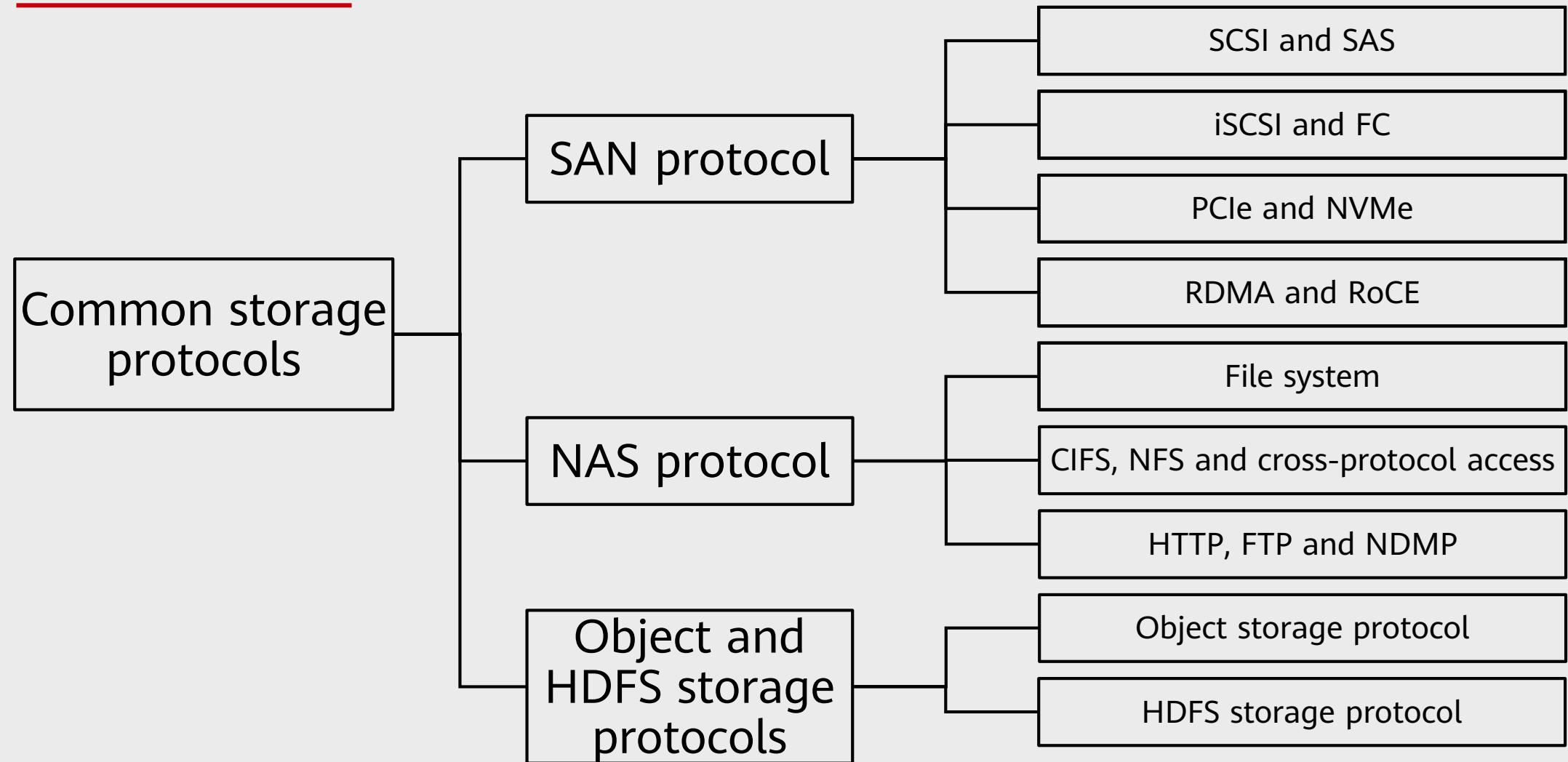
Quiz

1. Which networks are included in FC topologies?
 - A. Fibre Channel Arbitrated Loop (FC-AL)
 - B. Fibre Channel point-to-point (FC-P2P)
 - C. Switched network
 - D. Fibre Channel dual-switch
2. Which NFS versions are available currently?
 - A. NFSv1
 - B. NFSv2
 - C. NFSv3
 - D. NFSv4

Quiz

3. Which of the following are file sharing protocols?
 - A. HTTP
 - B. iSCSI
 - C. NFS
 - D. CIFS
4. Which of the following operations are involved in the CIFS protocol?
 - A. Protocol handshake
 - B. Security authentication
 - C. Share connection
 - D. File operations
 - E. Disconnection

Summary



Recommendations

- Huawei official websites:
 - Enterprise service: <https://e.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/en/index.html>
 - Online learning: <https://learning.huawei.com/en/>
- Popular tools
 - HedEx Lite
 - Network documentation tool center
 - Information query assistant

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.



Storage Network Architecture



Foreword

- With the development of host, disk, and network technologies, the storage system architecture evolves, and the storage network architecture also develops to meet service requirements. This course introduces the storage network architecture.

Objectives

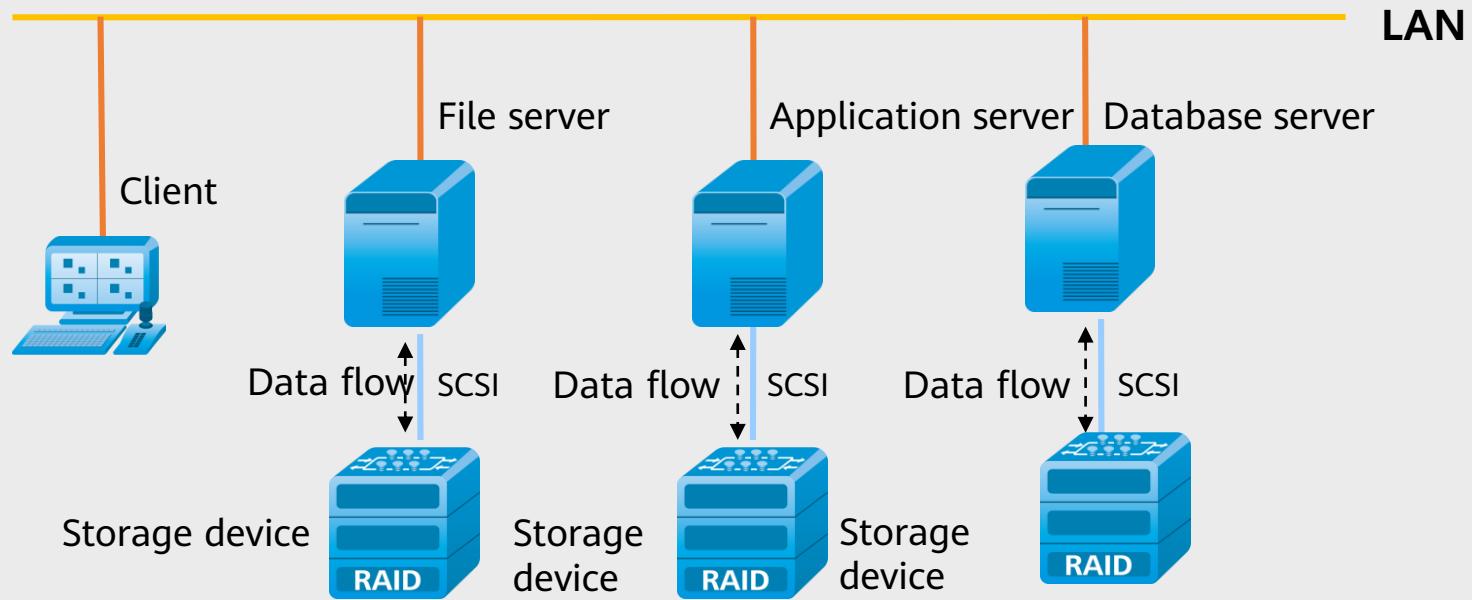
Upon completion of this lesson, you will be able to understand:

- Storage network architecture evolution
- Storage network technology evolution

Contents

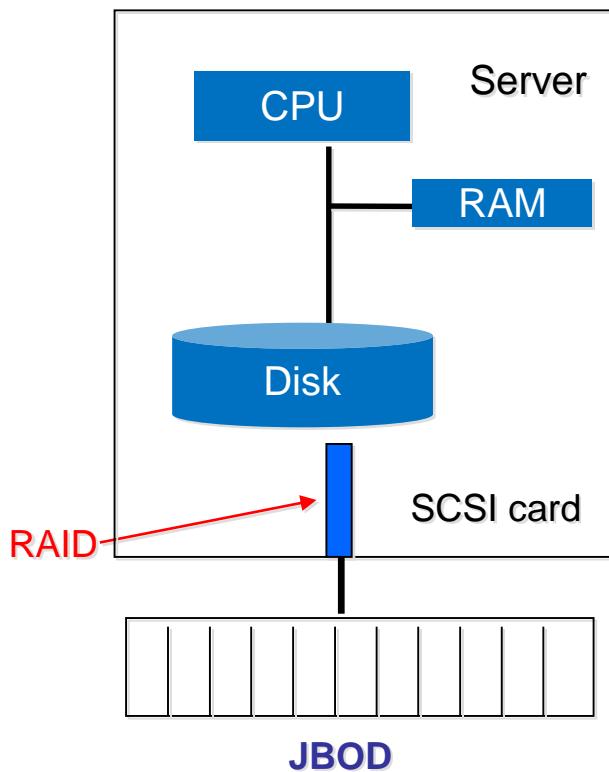
- 1. DAS**
2. NAS
3. SAN
4. Distributed Architecture

DAS

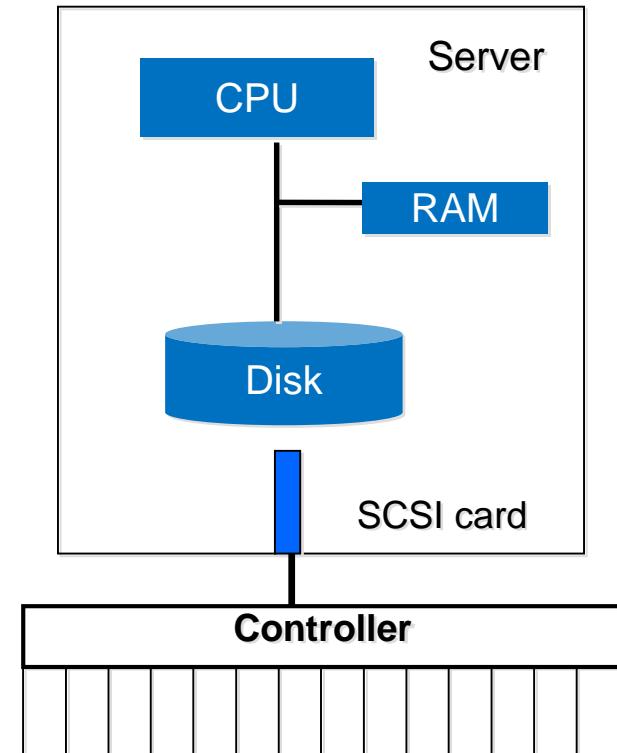


DAS

External disk array (DAS)



Smart disk array (DAS)



Challenges for DAS

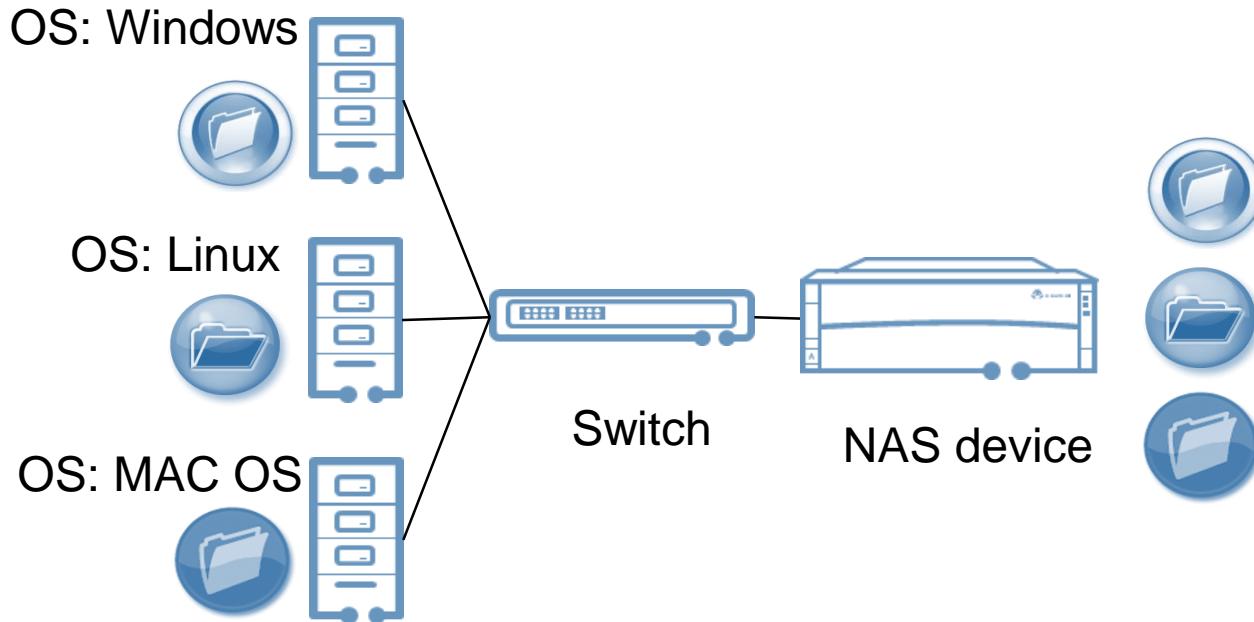
Challenges	Description
Low Scalability	Limited number of ports that can be connected to a host.
	Limited number of addressable disks.
	Limited distance.
Inconvenient Maintenance	The system needs to be powered off during maintenance.
Insufficient Resource Sharing	Front-end ports and storage space are difficult to share.
	Resource silos: For example, the DAS with insufficient storage space cannot share the remaining space of the DAS with excessive storage resources.

Contents

1. DAS
- 2. NAS**
3. SAN
4. Distributed Architecture

NAS

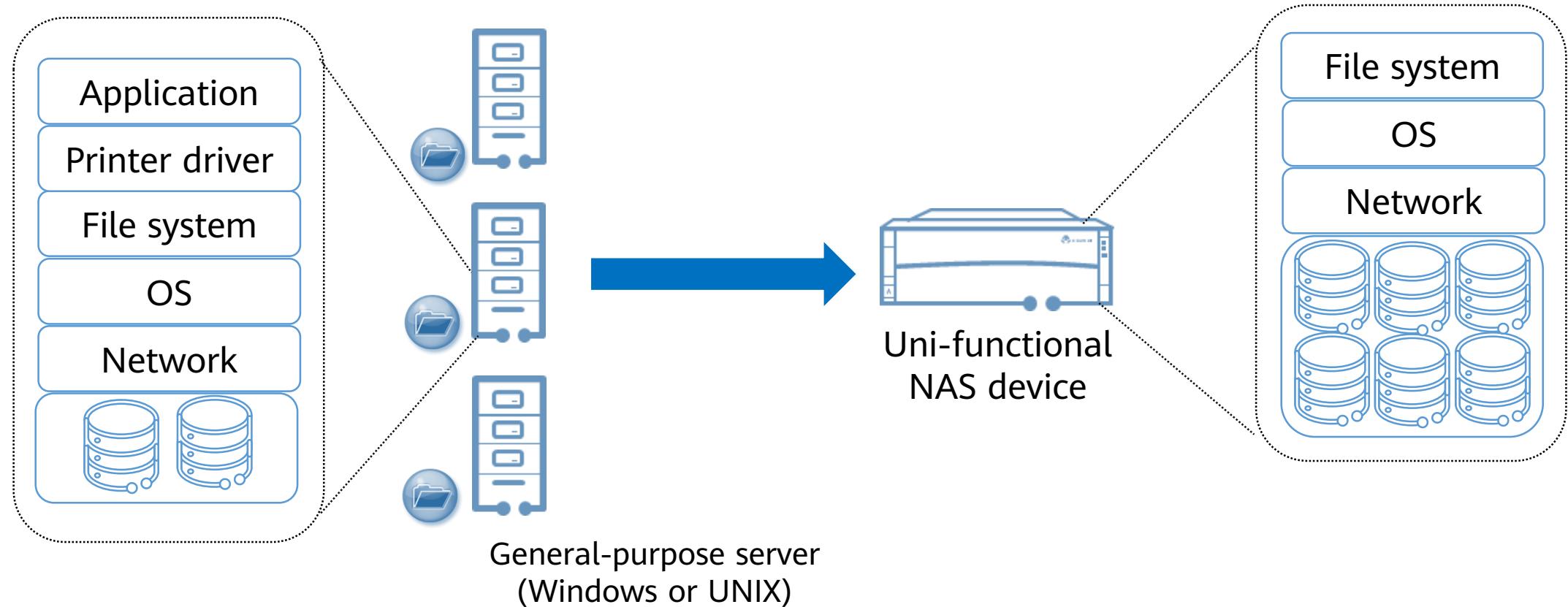
- Network-attached storage (NAS) connects storage devices to the live network and provides data and file services.
- The most commonly used network sharing protocols for NAS are Common Internet File System (CIFS) and Network File System (NFS).



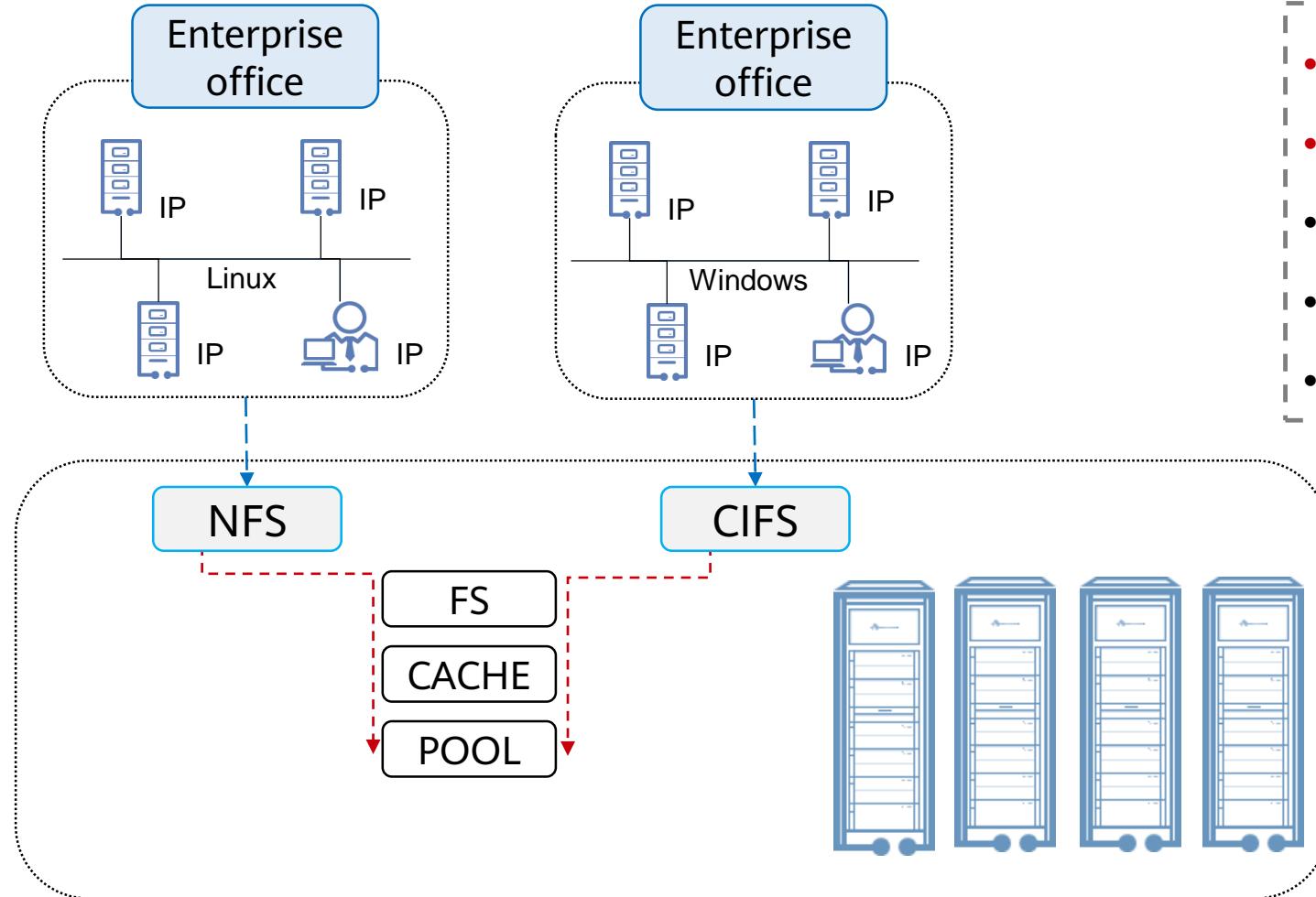
- **Benefits:**

- Improved efficiency
- Improved flexibility
- Centralized storage
- Simplified management
- High scalability
- High availability
- Security (user authentication and authorization)

General-Purpose Server and NAS Devices

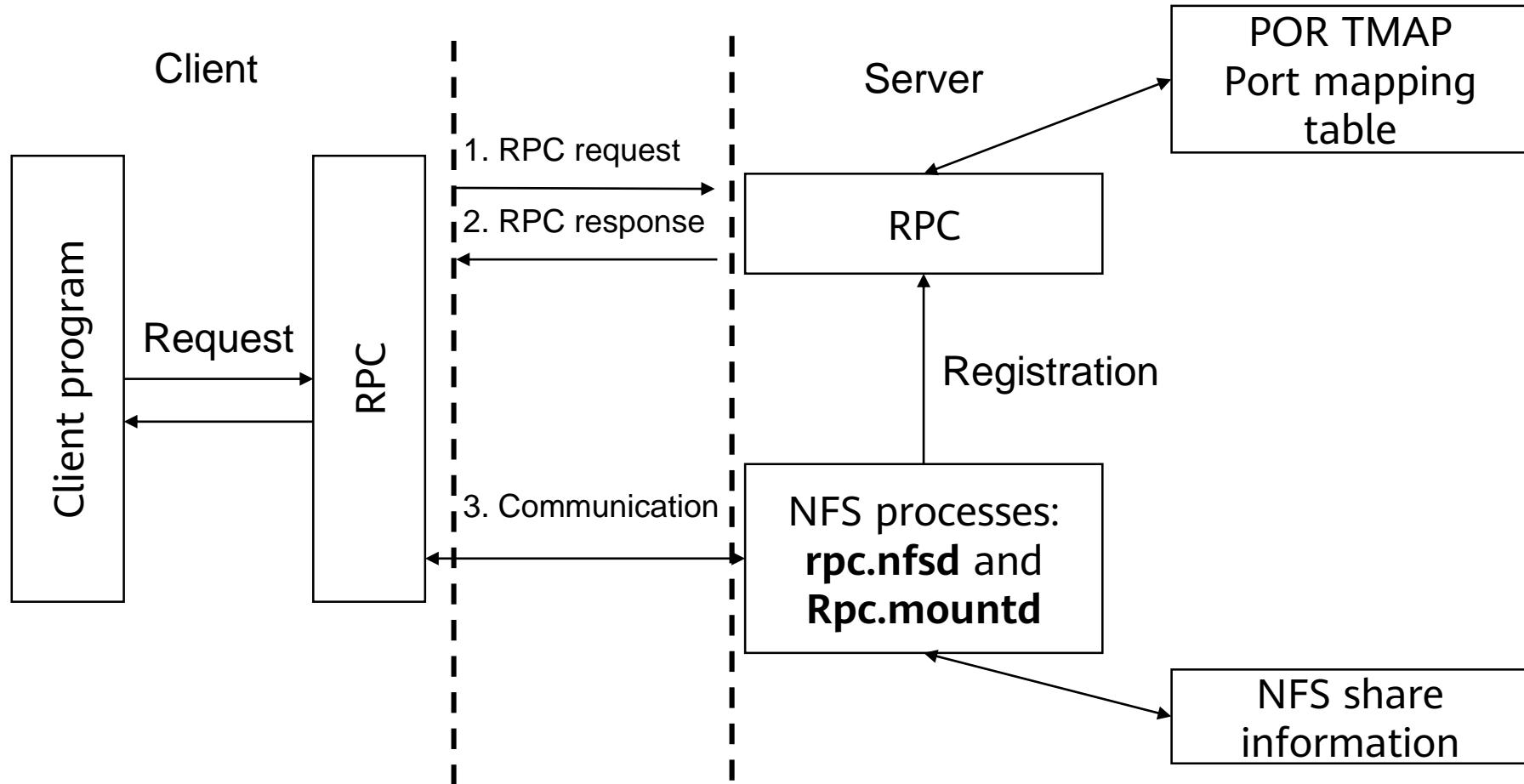


NAS Protocols



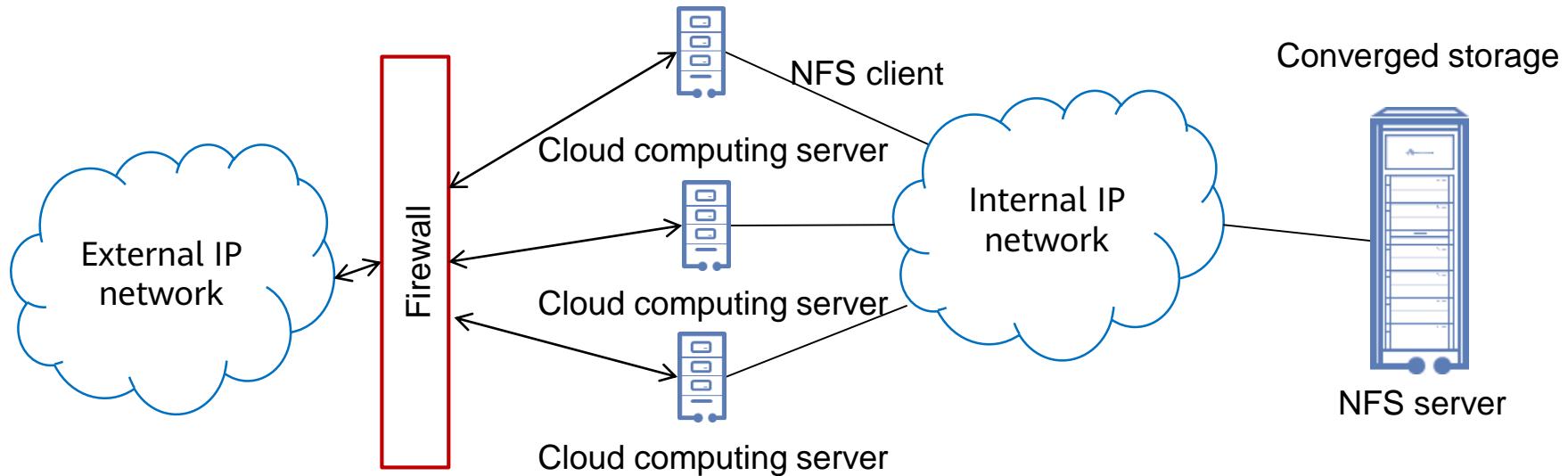
- NFS
- CIFS
- FTP
- HTTP
- NDMP

Working Principles of NFS

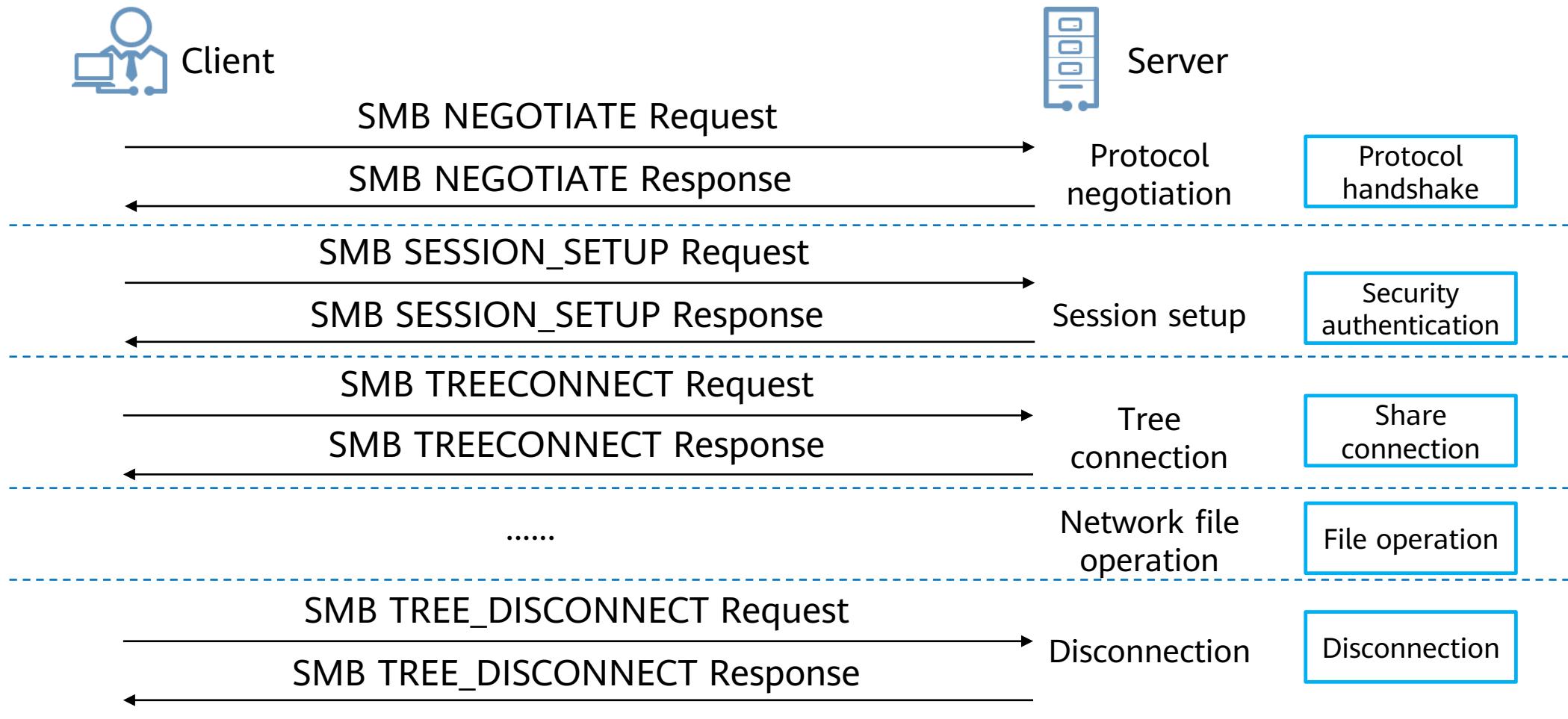


Typical Application of NFS: Shared Storage for Cloud Computing

- Cloud computing uses the NFS server as the internal shared storage.

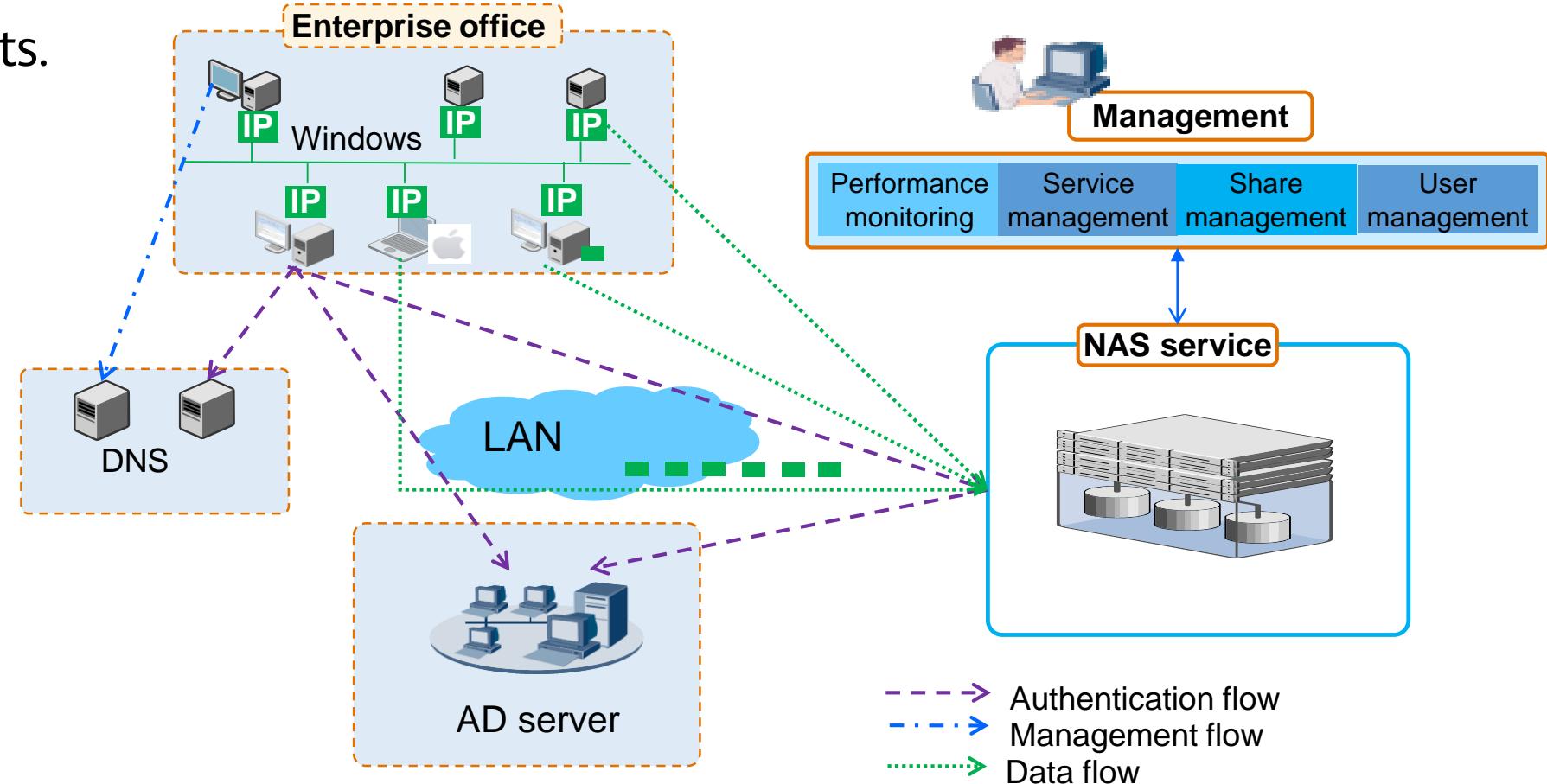


Working Principles of CIFS



Typical Application of CIFS: File Sharing Service

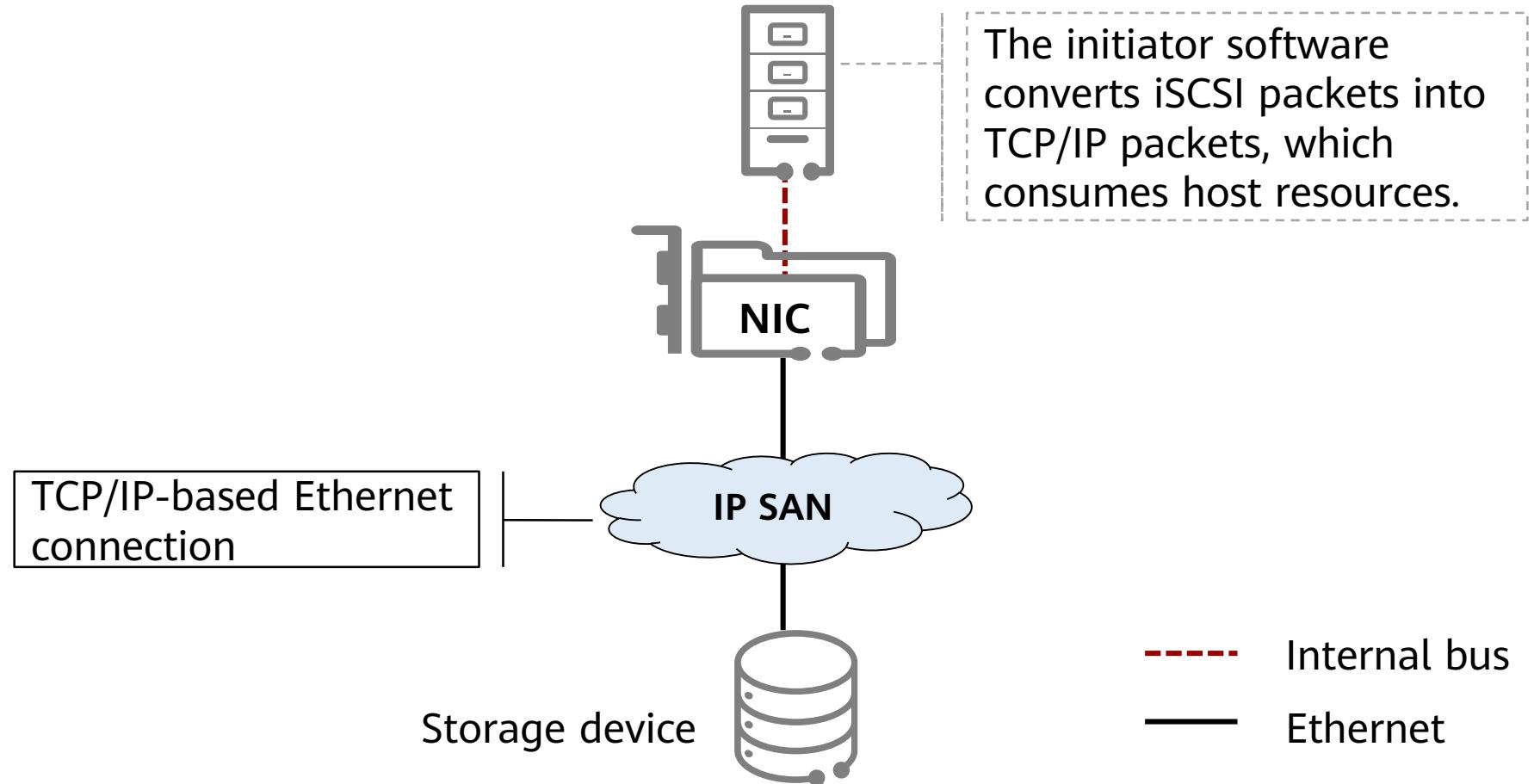
- The file sharing service applies to scenarios such as enterprise file servers and media assets.



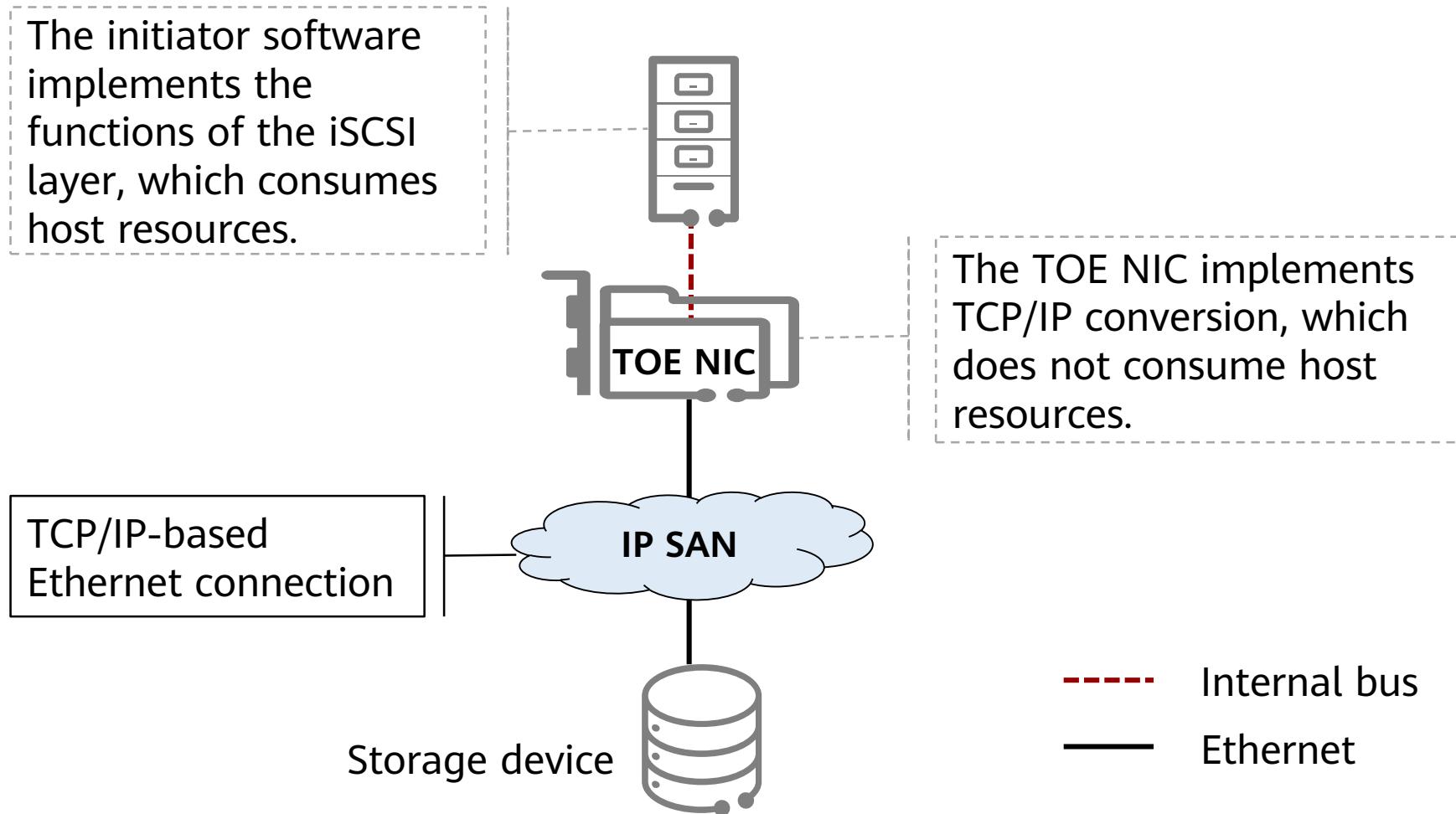
Contents

1. DAS
2. NAS
- 3. SAN**
 - IP SAN Technologies
 - FC SAN Technologies
 - Comparison Between IP SAN and FC SAN
4. Distributed Architecture

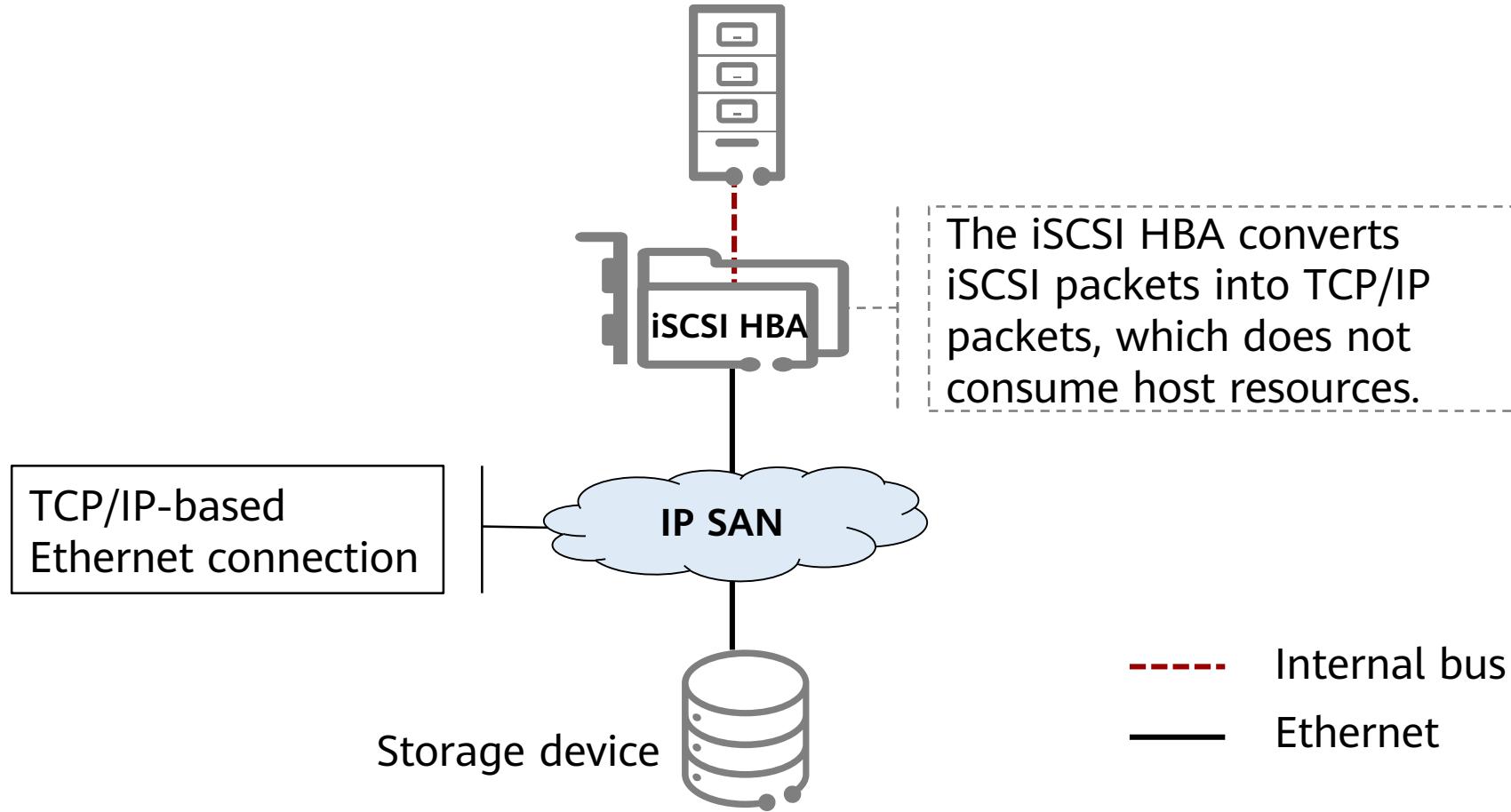
NIC + Initiator Software



TOE NIC + Initiator Software



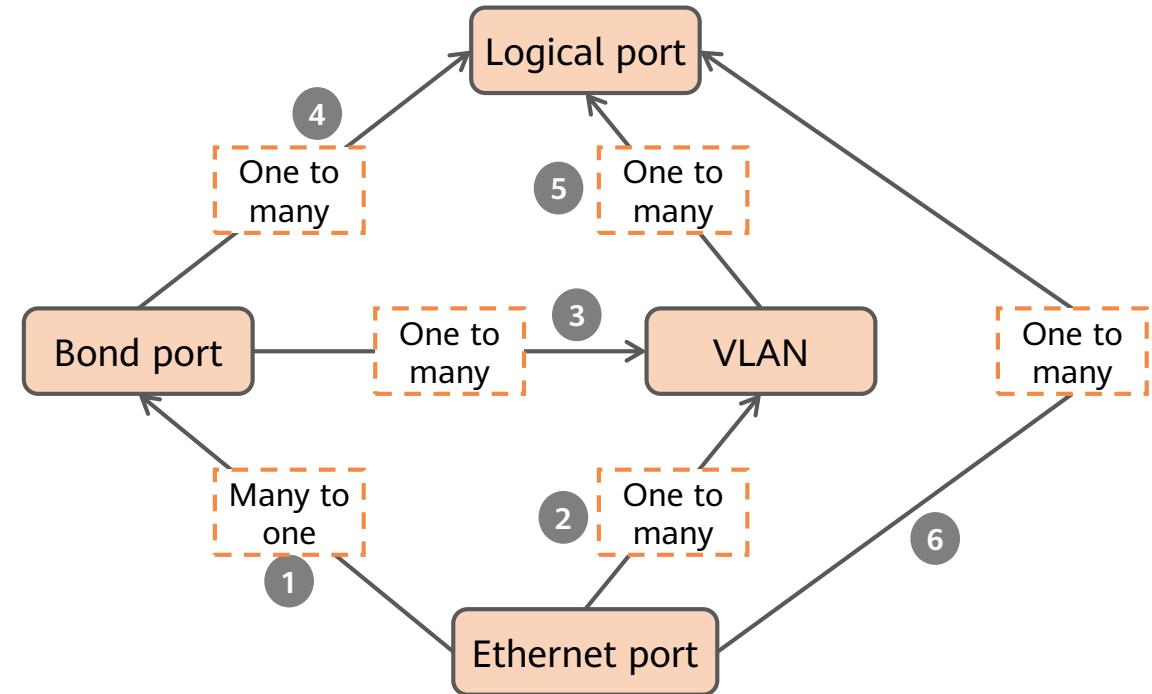
iSCSI HBA



Logical Port

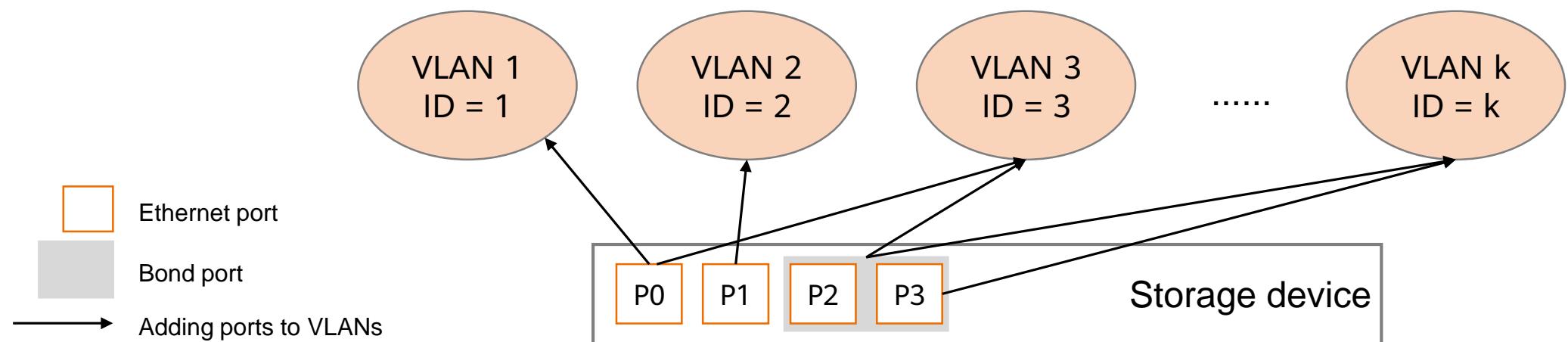
- Logical ports are created based on bond ports, VLAN ports, or Ethernet ports. The logical ports are virtual ports that carry host services.
- A unique IP address is allocated to each logical port for carrying its services.

No.	Description
1	Indicates that multiple Ethernet ports are bonded to form a bond port.
2	Indicates that an Ethernet port is added to multiple VLANs.
3	Indicates that a bond port is added to multiple VLANs.
4	Indicates that a bond port is used to create multiple logical ports.
5	Indicates that a VLAN port is used to create multiple logical ports.
6	Indicates that an Ethernet port is used to create multiple logical ports.



VLAN Configuration

- VLAN is a technology that logically divides a physical LAN into multiple broadcast domains.
- Ethernet ports or bond ports in a storage system can be added to multiple independent VLANs. You can configure different services in different VLANs to ensure the security and reliability of service data.



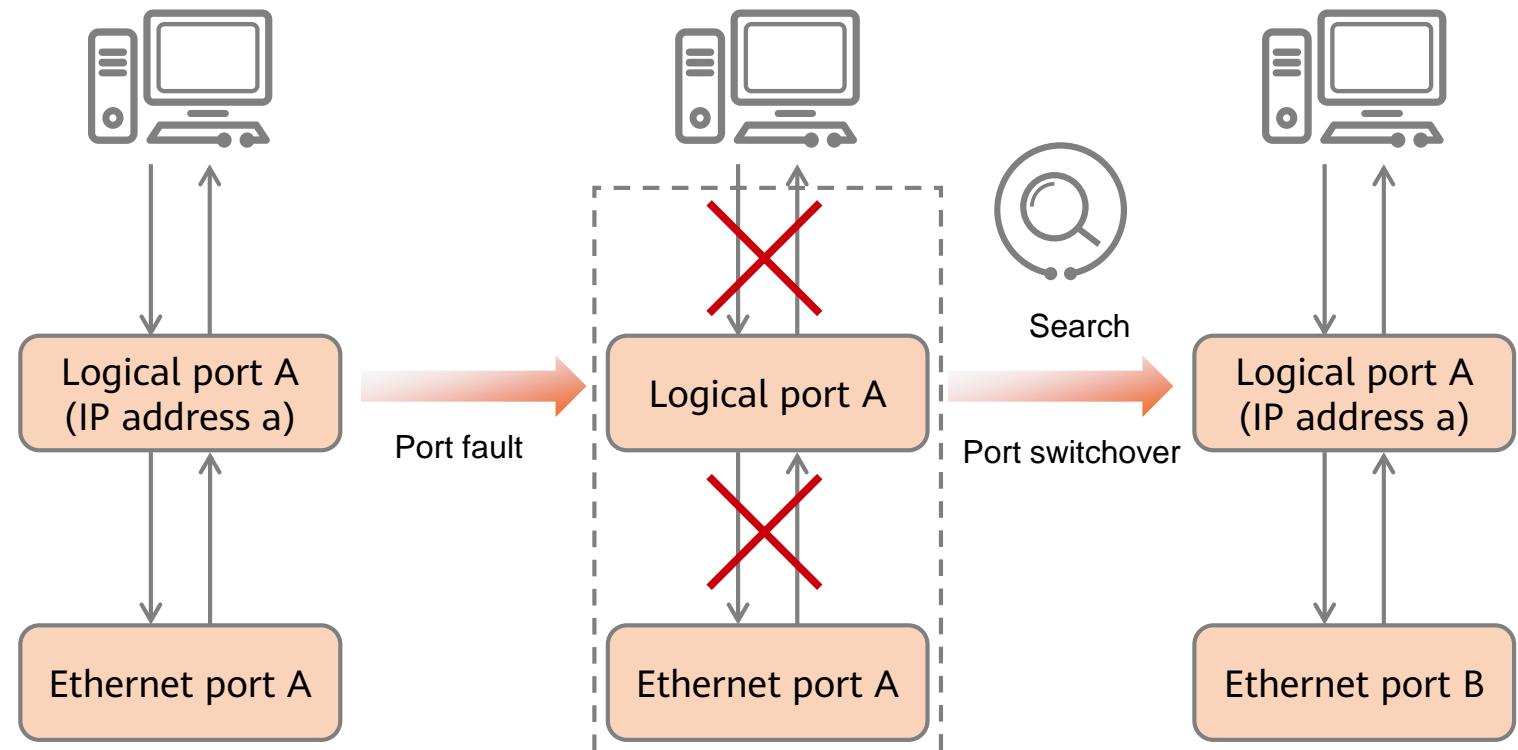
IP Address Failover

- IP address failover indicates that a logical IP address fails over from a faulty port to an available port. In this way, services are switched from the faulty port to the available port without interruption. The faulty port can take over services back after being recovered.
- During the IP address failover, services are switched from the faulty port to the available port, ensuring service continuity and improving reliability of paths for accessing file systems. This process is transparent to users.
- The essence is a service switchover between ports. The ports can be Ethernet ports, bond ports, or VLAN ports.

Ethernet Port-based IP Address Failover

- To improve reliability of paths for accessing file systems, you can create logical ports based on Ethernet ports.
- When the Ethernet port that corresponds to a logical port fails, the system will:

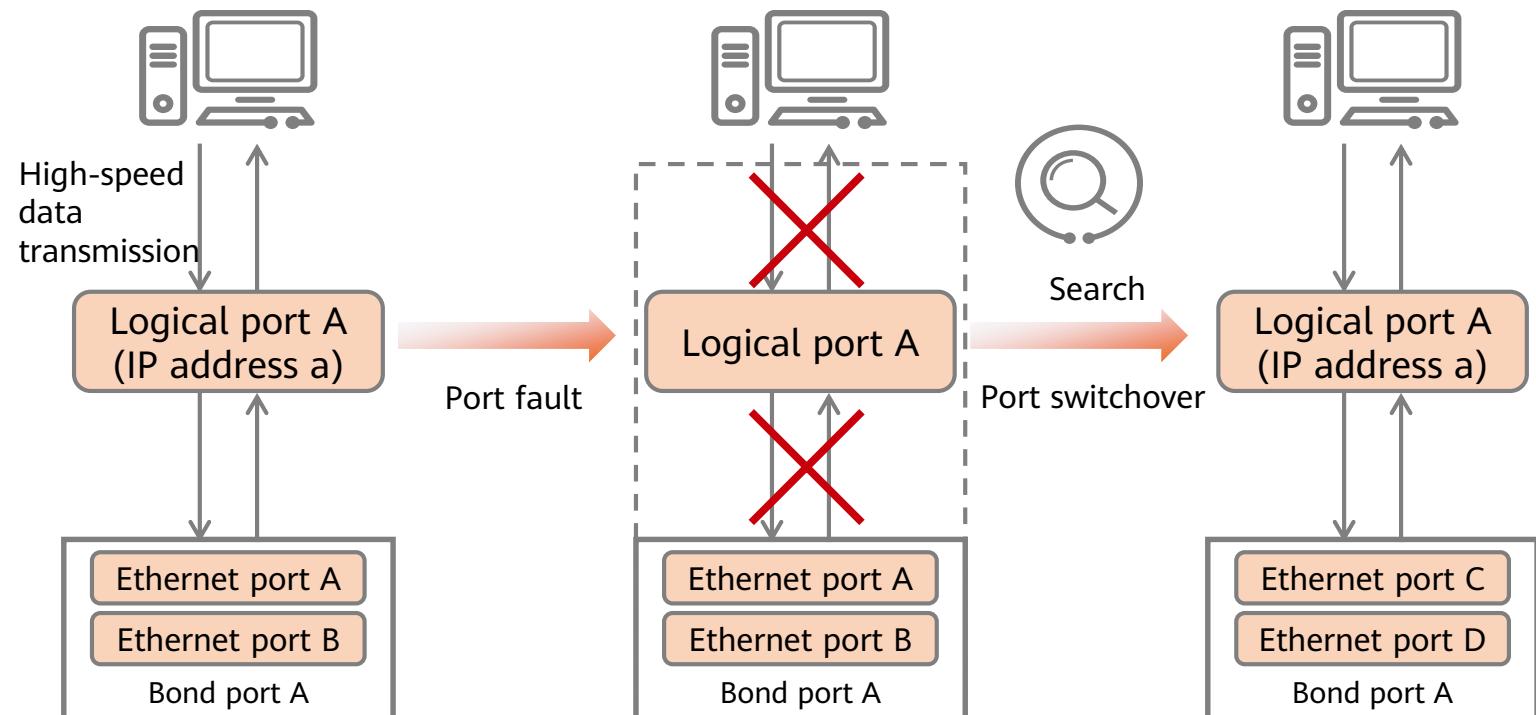
- Locate an available Ethernet port of the same type.
- Delete the logical port from the faulty Ethernet port.
- Create the same logical port on the available Ethernet port to carry services.
- Ensure service continuity.



Bond Port-based IP Address Failover

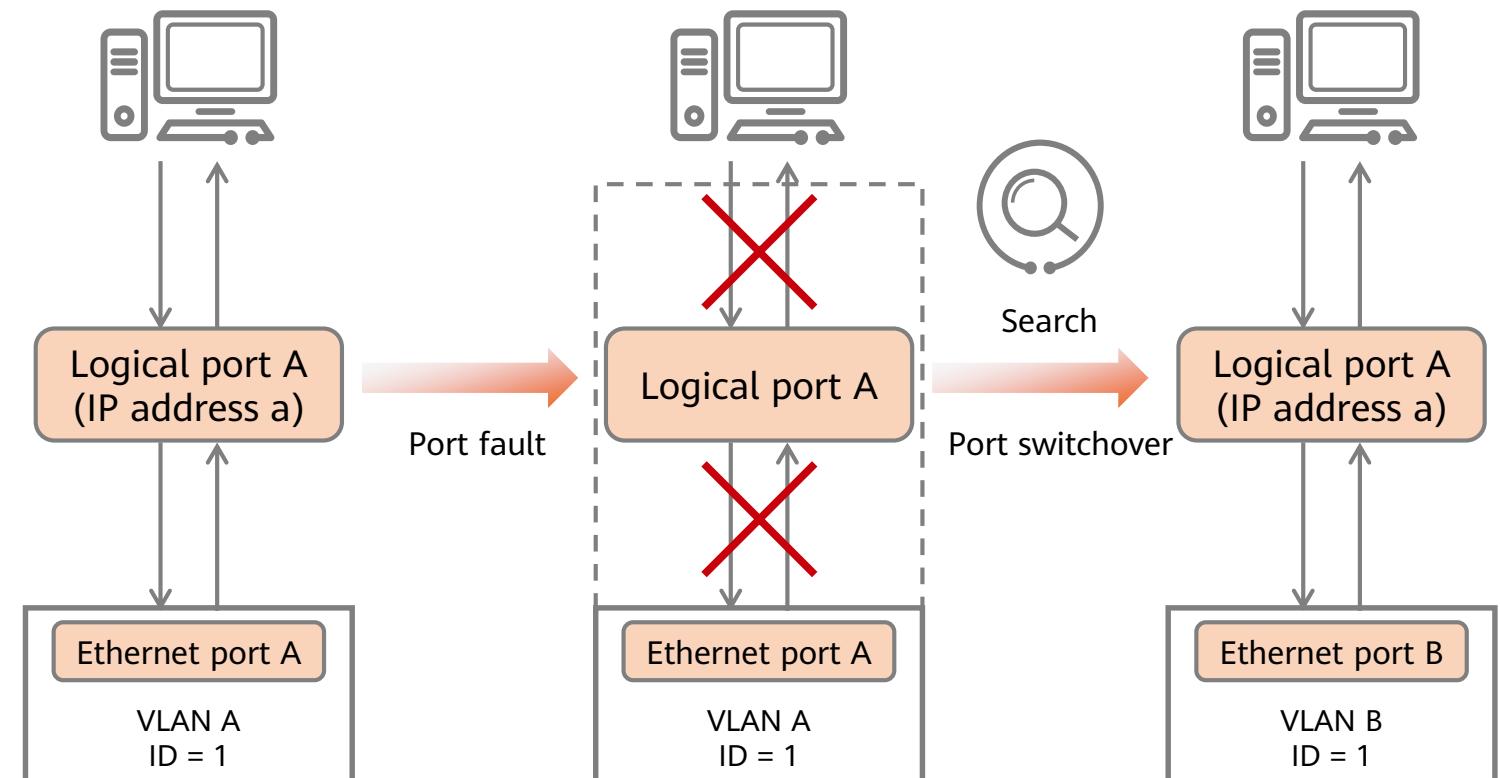
- To improve reliability of paths for accessing file systems, you can bond multiple Ethernet ports to form a bond port.
- When the Ethernet ports that are used to create the bond port fails, the system will:
 - Locate an available port.
 - Delete the logical port created on the faulty port.
 - Create a logical port with the same IP address on the available port.
 - Switch services to the available port.

After the faulty port recovers, it can take over services again.



VLAN-based IP Address Failover

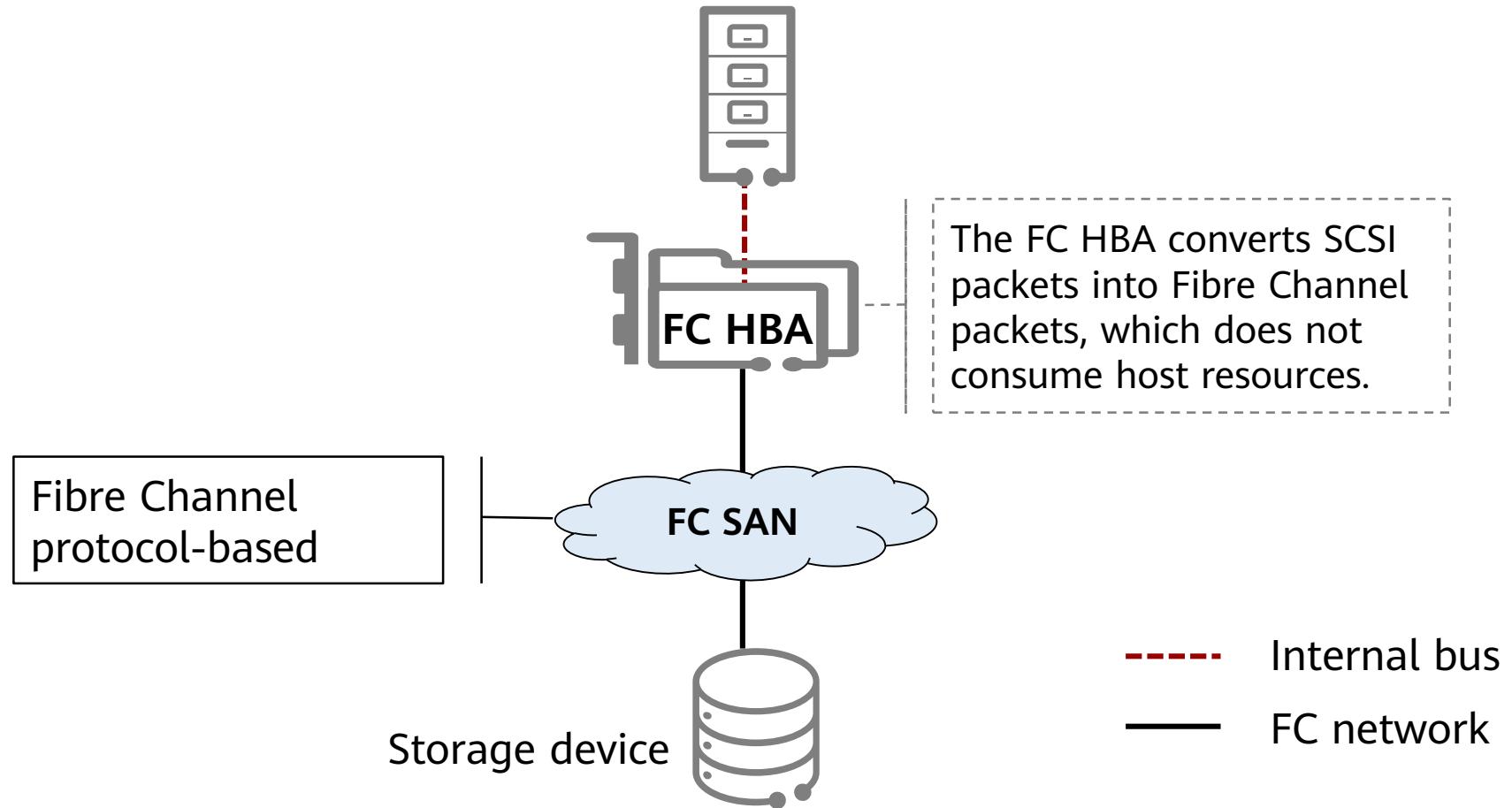
- You can create VLANs to isolate different services.
- When an Ethernet port on a VLAN fails, the system will:
 - Locate an available port of the same type.
 - Delete the logical port from the faulty port.
 - Create the same logical port on the available port.
 - Switch services to the available port.



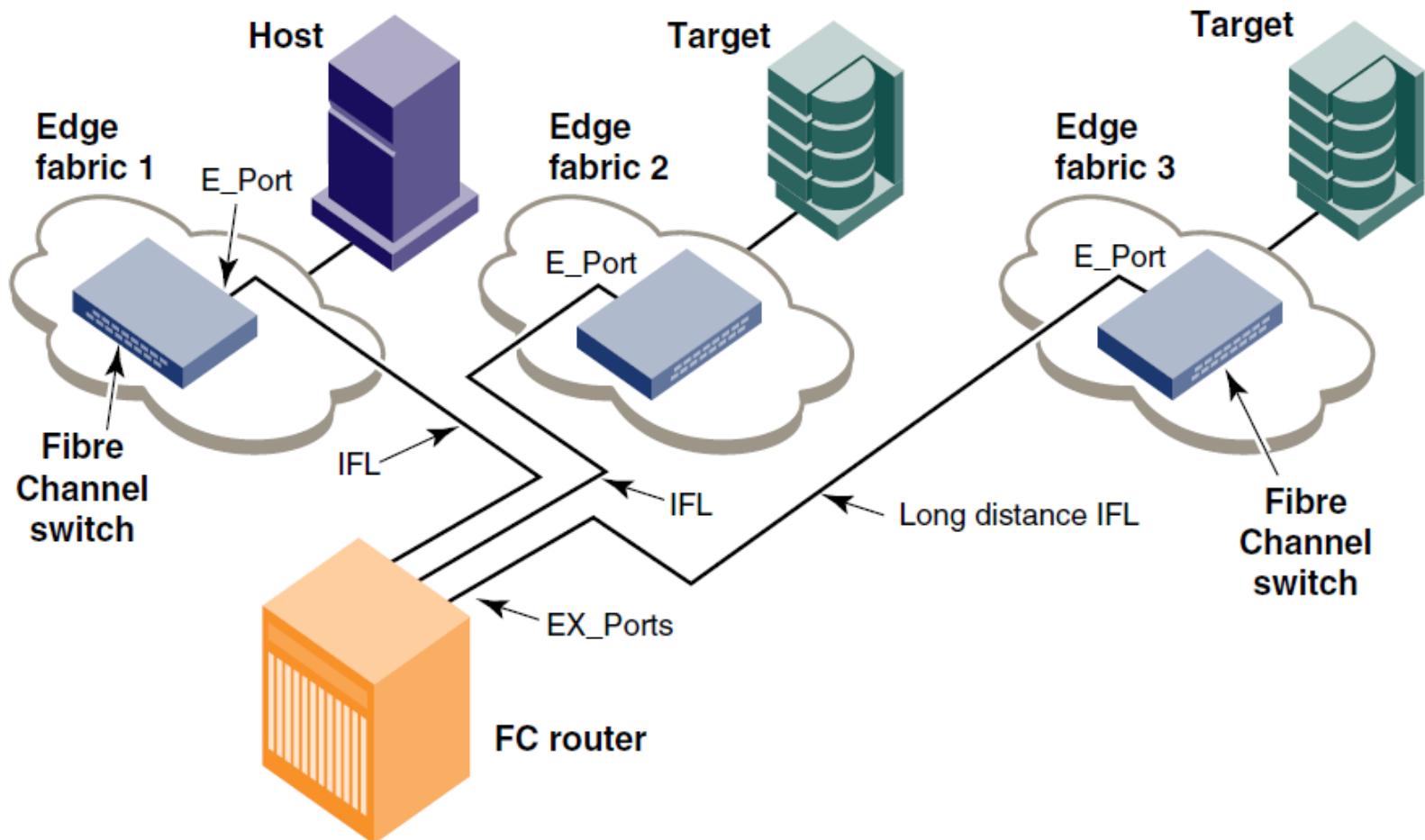
Contents

1. DAS
2. NAS
- 3. SAN**
 - IP SAN Technologies
 - FC SAN Technologies
 - Comparison Between IP SAN and FC SAN
4. Distributed Architecture

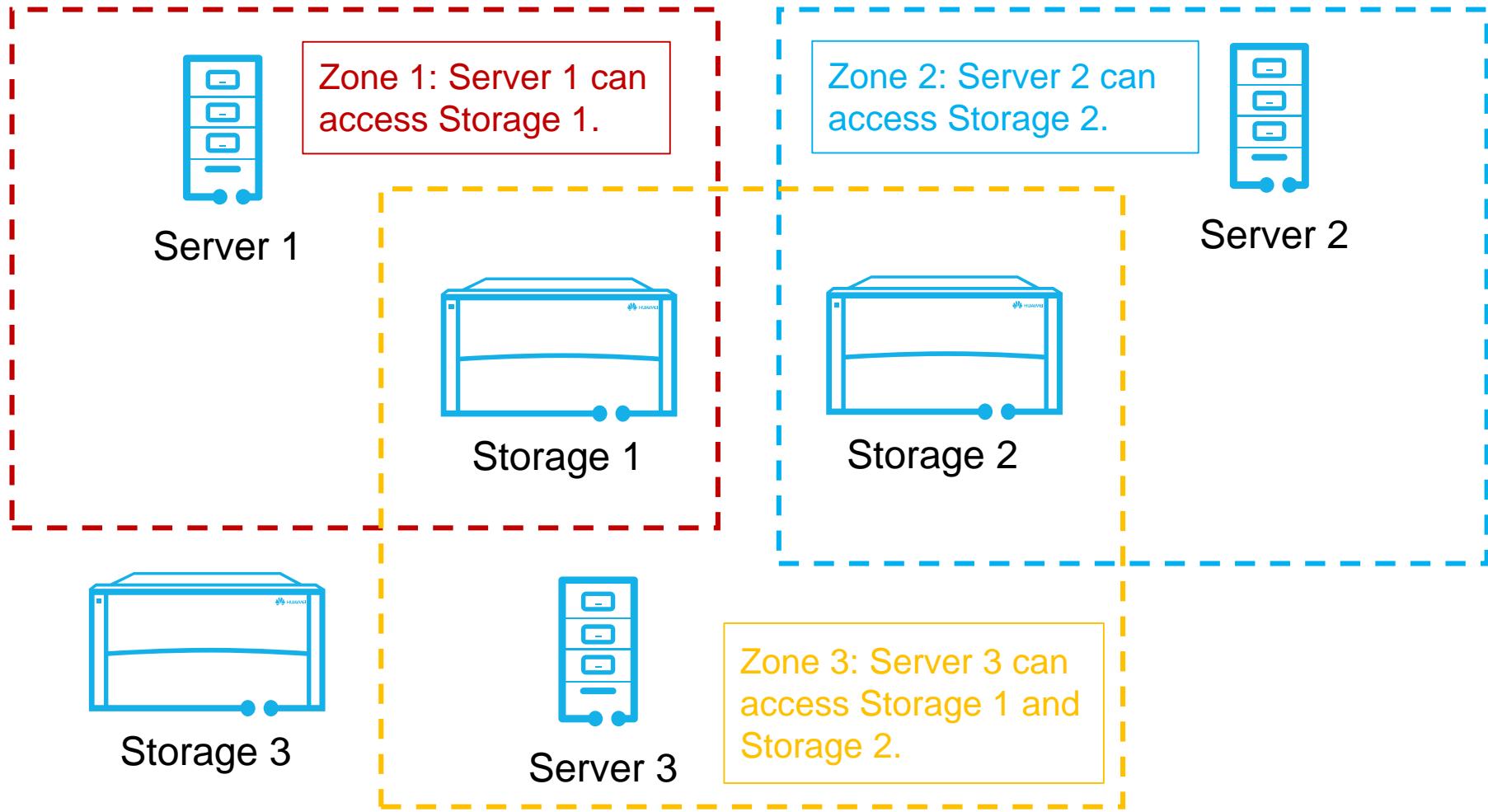
FC HBA



FC Network



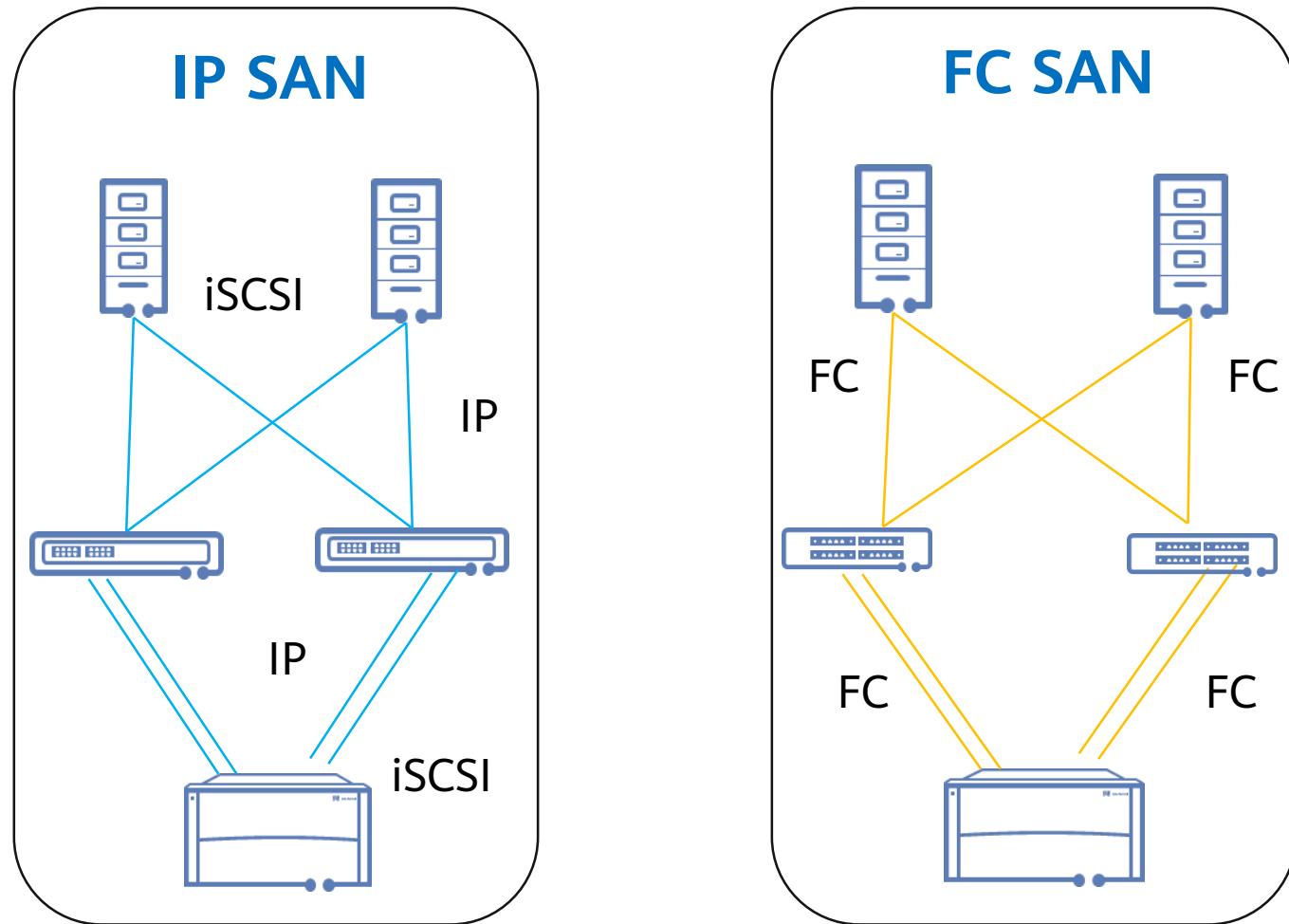
Zoning



Contents

1. DAS
2. NAS
- 3. SAN**
 - IP SAN Technologies
 - FC SAN Technologies
 - Comparison Between IP SAN and FC SAN
4. Distributed Architecture

IP SAN and FC SAN



Comparison Between IP SAN and FC SAN

Item	IP SAN	FC SAN
Network architecture	Existing IP networks	Dedicated Fibre Channel networks and HBAs
Transmission distance	Not limited theoretically	Limited by the maximum transmission distance of optical fibers
Management and maintenance	As simple as operating IP devices	Complicated technologies and management
Compatibility	Compatible with all IP network devices	Poor compatibility
Cost	Lower purchase and maintenance costs than FC SAN, higher return on investment (ROI)	High purchase (Fibre Channel switches, HBAs, Fibre Channel disk arrays, and so on) and maintenance (staff training, system configuration and supervision, and so on) costs
Disaster recovery (DR)	Local and remote DR available based on existing networks at a low cost	High hardware and software costs for DR
Security	Relatively low	Relatively high

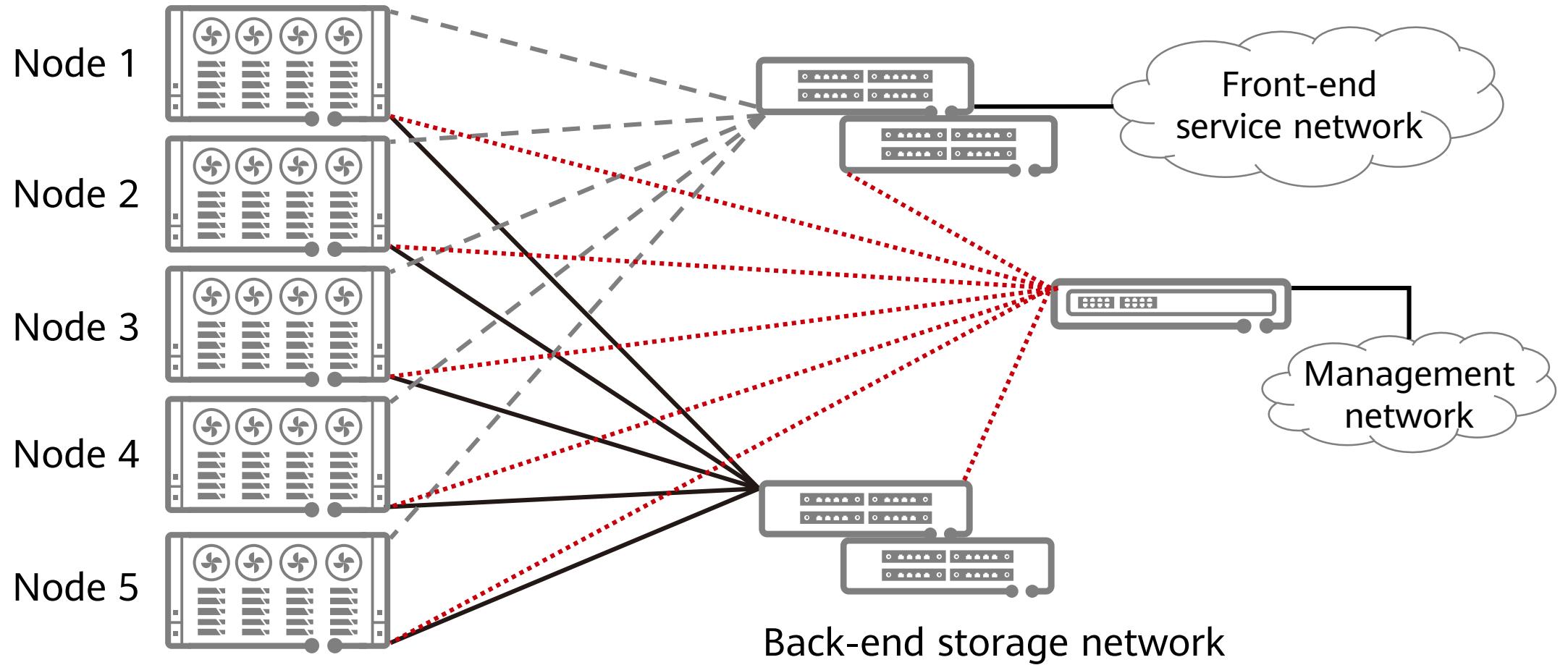
Comparison Between DAS, NAS, and SAN

Storage System Architecture	DAS	NAS	SAN
Data transmission protocol	SCSI/FC/ATA	TCP/IP	FC
Transport object	Data block	File	Data block
Using standard file sharing protocols	No	Yes (NFS/CIFS.....)	No
Centralized management	Not sure.	Yes	Management tools required
Improving server efficiency	No	Yes	Yes
Disaster tolerance	Low	High	High, dedicated solution
Application scope	SME servers and JBOD	SME, monitoring, and broadcasting	Large enterprises and data centers
Application environment	LAN Documents are seldom shared, the operation platform is independent, and the number of servers is small.	LAN Documents are highly shared, and different media formats have high storage requirements.	Fibre channel storage domain network Complex network environment, high degree of document sharing, heterogeneous operating system platform, and a large number of servers.
Capacity expansion capability	Low	Medium	High

Contents

1. DAS
2. NAS
3. SAN
- 4. Distributed Architecture**

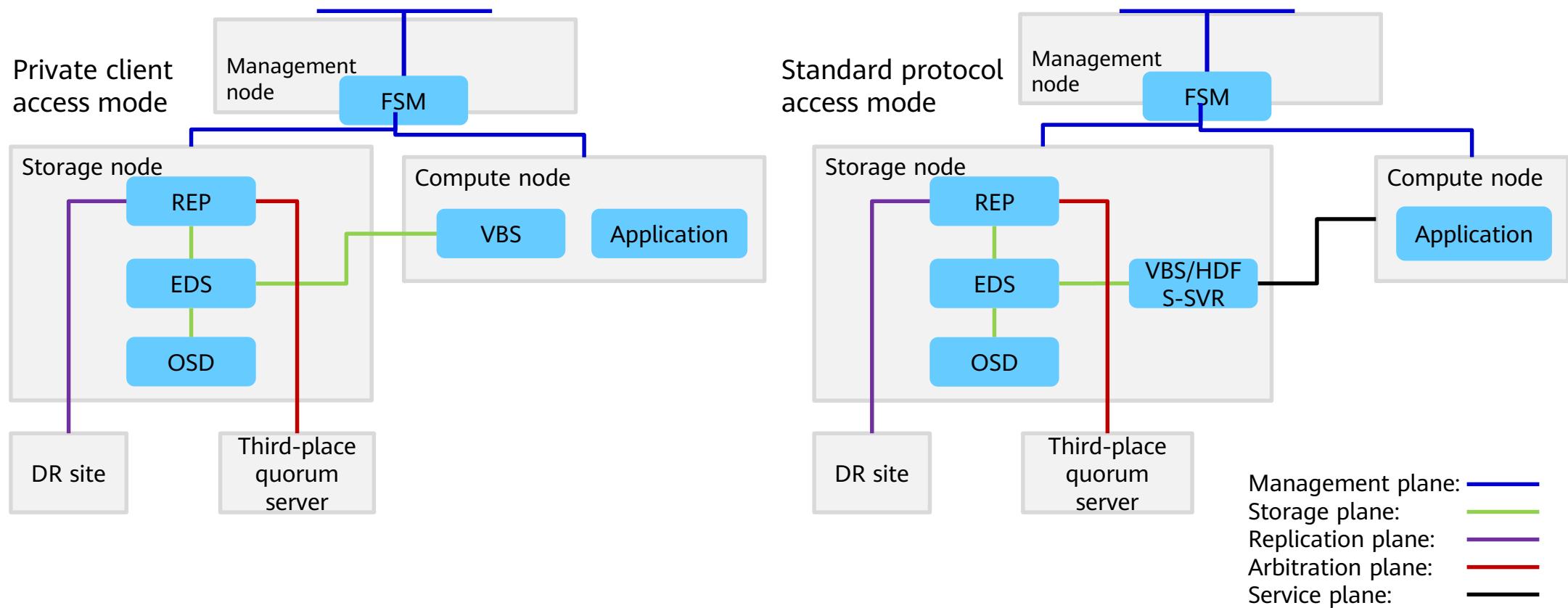
Scale-out Storage Networking



Networking Overview

- Front-end service/Tenant network
 - The front-end service/tenant network is used to interconnect the scale-out storage with the customer network. It provides the tenant UI for tenant users to complete operations such as resource application and usage query, and processes service requests sent by tenant clients or APIs.
- Back-end storage network
 - The back-end storage/internal management network is used for internal interconnection between nodes. It provides heartbeat communication between high availability (HA) components such as the data service subsystem (DSS), and internal communication and data interaction between components.
- Management network
 - The management network is used to interconnect with the customer's maintenance network. It provides a management UI for the system administrator to perform service operations such as system configuration, tenant management, resource management, and service provisioning, as well as maintenance operations such as alarm, performance, and topology management. In addition, the Mgmt ports of all physical nodes can be aggregated to provide remote device maintenance capabilities, such as remotely logging in to the virtual KVM of a device and viewing hardware running data such as temperature and voltage.

Network Planes



Networking Rules

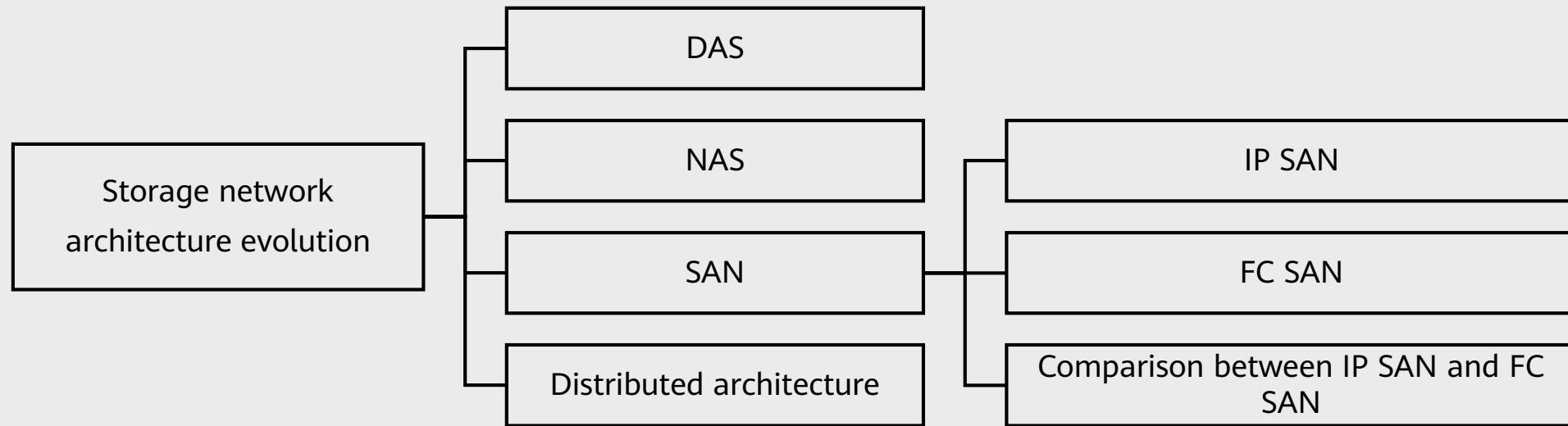
- Nodes must be placed in a cabinet from bottom to top.
- A deployment solution is usually chosen based on project requirements. The total power consumption and weight of the storage nodes, switches, and KVM in a cabinet must be calculated and the number of nodes that can be housed by a cabinet must be determined based on the equipment room conditions.
- In typical configuration, nodes and switches in the base cabinet are connected through network cables and SFP+ cables, and nodes in an expansion cabinet connect to switches in the base cabinet through network cables and optical fibers.

Quiz

1. Which of the following are included in scale-out storage networking?
 - A. Management network
 - B. Front-end service network
 - C. Front-end storage network
 - D. Back-end storage network

2. Which of the following protocols are commonly used in SAN networking?
 - A. FC
 - B. iSCSI
 - C. CIFS
 - D. NFS

Summary



Recommendations

- Huawei official websites
 - Enterprise business: <https://e.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/en/index.html>
 - Online learning: <https://www.huawei.com/en/learning>
- Popular tools
 - HedEx Lite
 - Network Document Tool Center
 - Information Query Assistant

Acronyms and Abbreviations

Direct Attached Storage (DAS): An external storage device is directly connected to a computer through a cable.

Redundant Array of Independent Disks (RAID): It is a technology that provides a disk group (logical disk) consisting of multiple disks (physical disks) combined in different modes. The disk group features higher storage performance over a single disk and supports data redundancy.

Redirect on write (ROW): A core technology used to create file system snapshots. When a source file system receives a write request to modify existing data, the storage system writes the new data to a new location and directs the BP of the modified data block to the new location.

Virtual Local Area Network (VLAN): A VLAN is a group of hosts with a common set of requirements that communicate as if they were attached to the same broadcast domain, regardless of their physical location. VLAN membership can be configured through software instead of physically relocating devices or connections.

HBA: Host Bus Adapter

KVM: keyboard, video, and mouse. You can use the KVM to remotely view the screen of the terminal host or use the local mouse and keyboard to remotely control the terminal host. In this way, the administrator can remotely solve the problems that occur on the terminal host.

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.



Huawei Intelligent Storage Products



Foreword

- This course describes features, positioning, and typical application scenarios of Huawei intelligent storage products, including Huawei all-flash storage, hybrid flash storage, scale-out storage, hyper-converged storage, and backup storage.

Objectives

Upon completion of this course, you will understand the following key information about Huawei intelligent storage products:

- Features
- Positioning
- Typical Application Scenarios

Contents

- 1. Panorama**
2. All-Flash Storage
3. Hybrid Flash Storage
4. Scale-Out Storage
5. Hyper-Converged Storage
6. Backup Storage

Huawei Storage Products

Solutions



Intelligent data lake



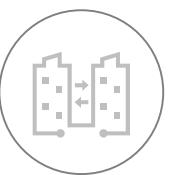
Video cloud



HPDA



Enterprise application acceleration



Active-active



Backup

Data management

Intelligent O&M



eService

Data management engine



DME Storage

Device management



DeviceManager

Storage products



All-flash storage



Hybrid flash storage



Scale-out storage



Hyper-converged storage



Backup storage

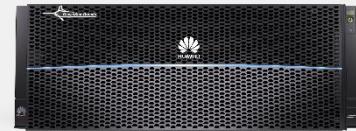
Contents

1. Panorama
- 2. All-Flash Storage**
3. Hybrid Flash Storage
4. Scale-Out Storage
5. Hyper-Converged Storage
6. Backup Storage

All-Flash Products

High-end

OceanStor Dorado 8000



OceanStor Dorado 18000 V6



Mid-range

OceanStor Dorado 3000



OceanStor Dorado 5000

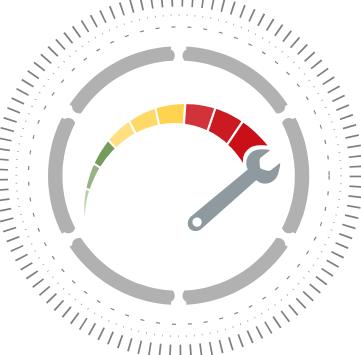


OceanStor Dorado 6000



Product Features

Ever fast



Industry-high performance
Industry-low latency

21 million IOPS
0.05 ms latency
30% higher NAS
performance than the
industry average

Always-on



SmartMatrix full-mesh
architecture
Always-on key services

**Tolerance of failures of seven
out of eight controllers**
Integrated SAN and NAS and
active-active architecture

AI-powered



Ultimate convergence
of SAN and NAS
Full-lifecycle intelligent
management

Intelligent O&M
Device-cloud synergy



Huawei OceanStor
Dorado all-flash storage

Device Model Examples

High-end controller enclosure

Back panel



4 U, 28 interconnect I/O modules

Mid-range controller enclosure



2 U, 2 controllers per enclosure, 12 interface modules

Entry-level controller enclosure



2 U, 2 controllers per enclosure, 6 interface modules

Smart disk enclosure



Front panel



4 U, 4 controllers per enclosure

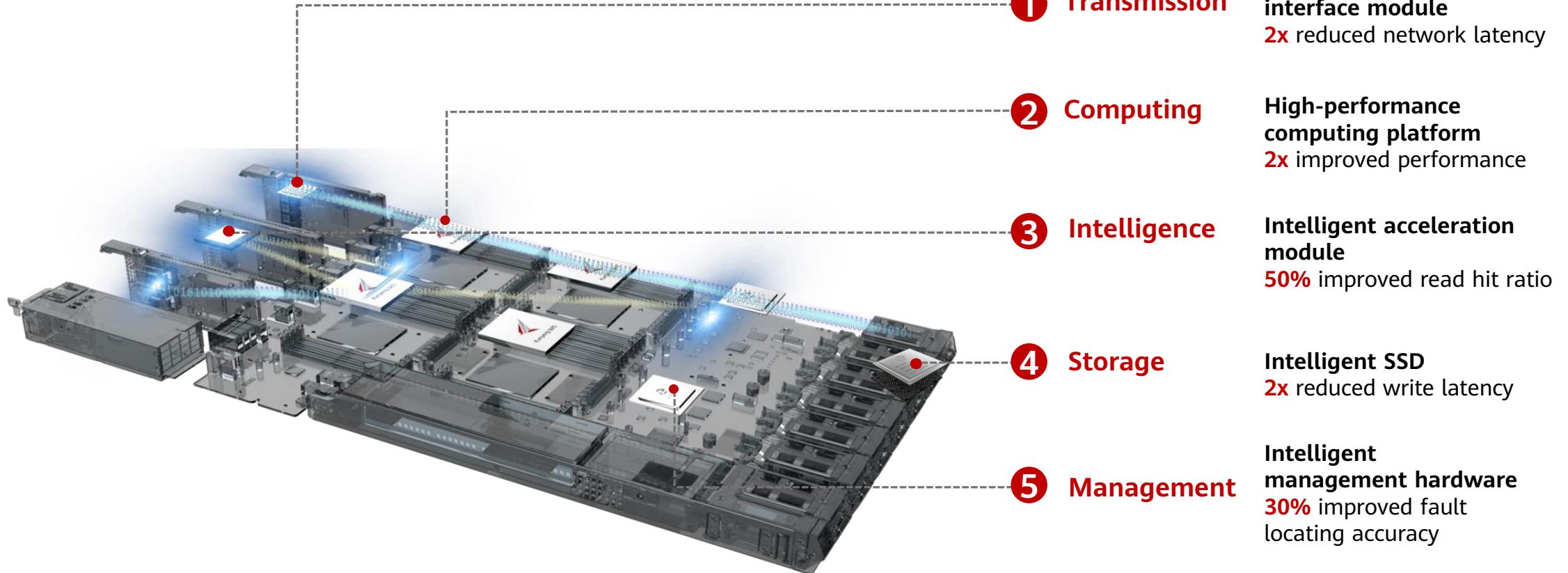


2 U, 36 NVMe SSDs (high-density)

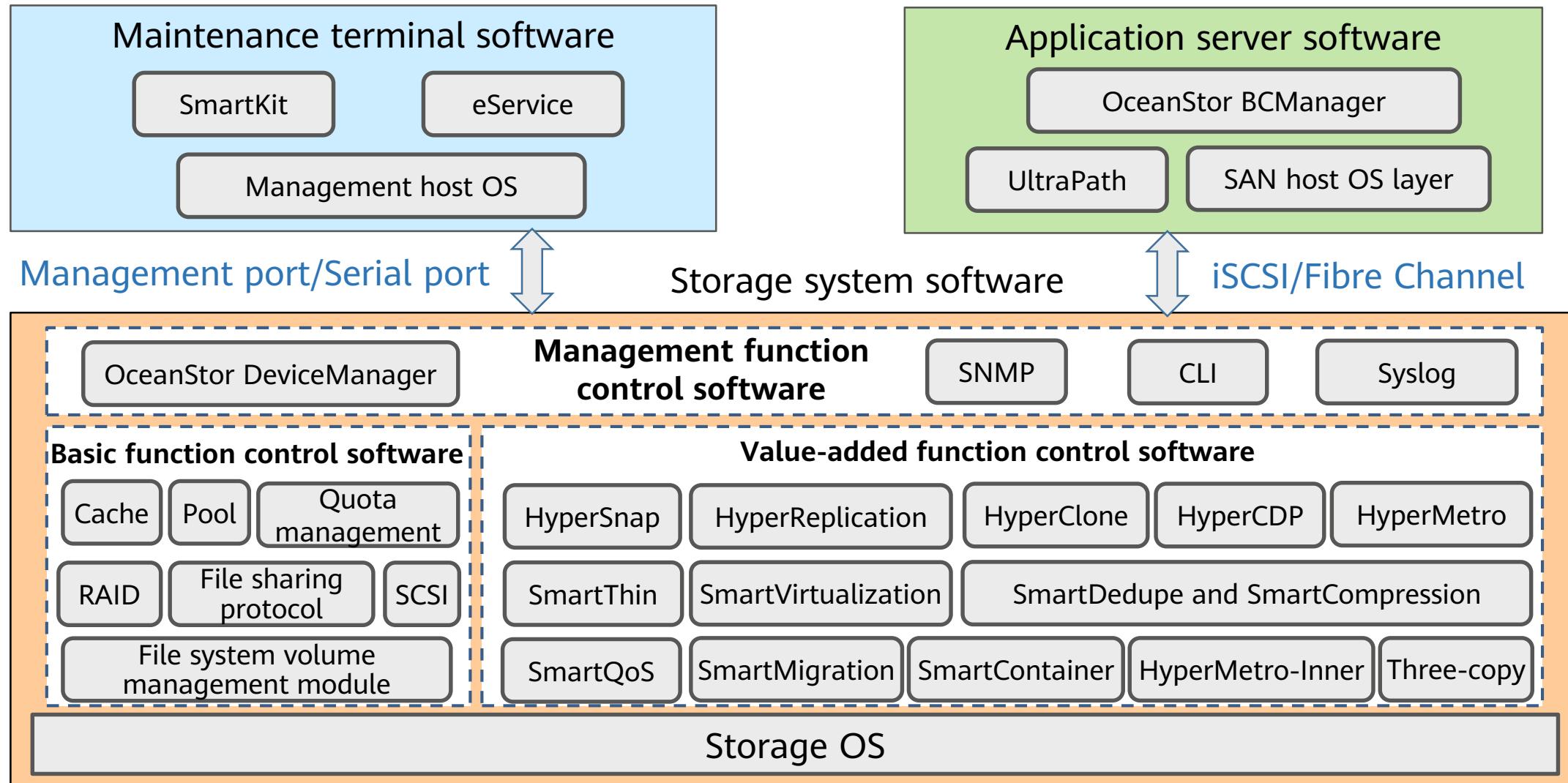


2 U, 25 SAS SSDs

Innovative and Intelligent Hardware Accelerates Critical Paths (Ever Fast)



Software Architecture



Intelligent Chips

Controller with five chips



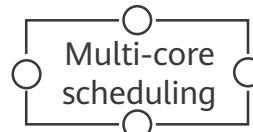
Smart disk enclosure



SSD



FlashLink® intelligent algorithm

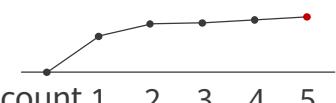


Kunpeng chip +
multi-core algorithm

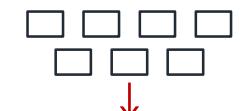


X00 minutes/TB 1X minutes/TB

Kunpeng chip + service offloading
Faster reconstruction



AI chip + cache algorithm
Improved read hit ratio



Full-stripe write
Reduced write amplification

Metadata

New data

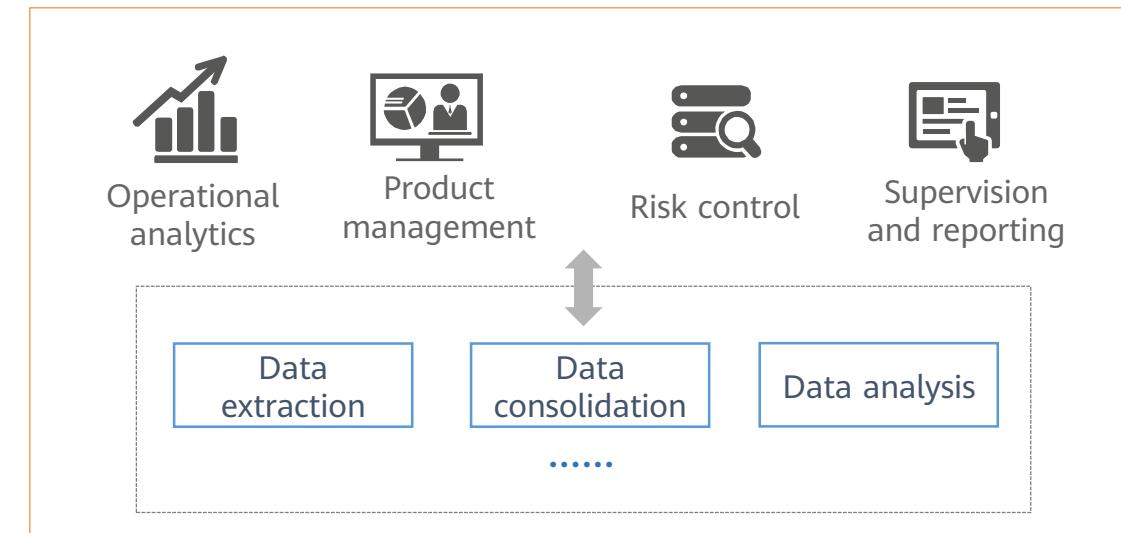
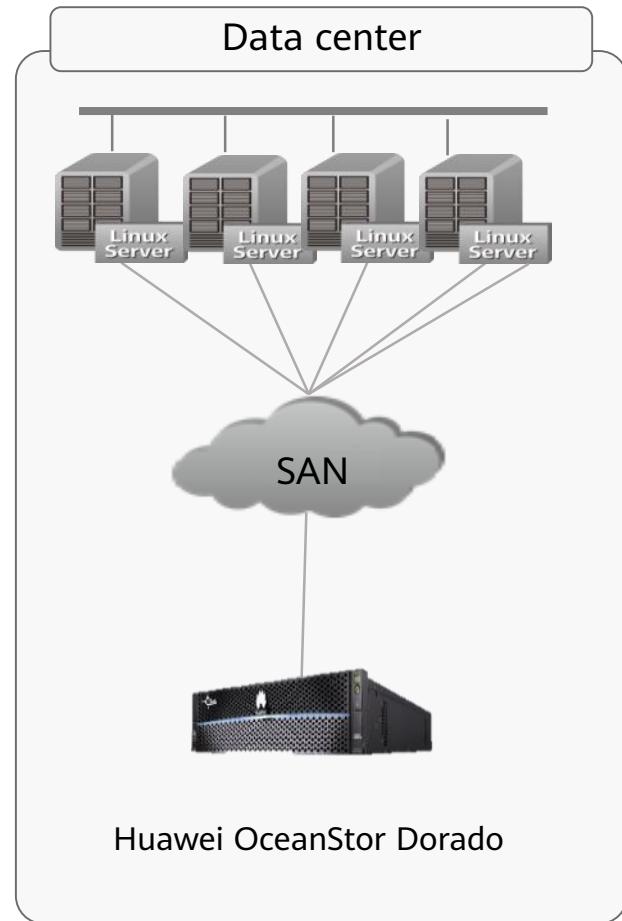
GC migration

Multi-stream data partitioning
Reduced garbage collection



Global I/O priority adjustment
Constant low latency

Typical Application Scenario – Mission-Critical Service Acceleration



Contents

1. Panorama
2. All-Flash Storage
- 3. Hybrid Flash Storage**
4. Scale-Out Storage
5. Hyper-Converged Storage
6. Backup Storage

Hybrid Flash Storage

High-end

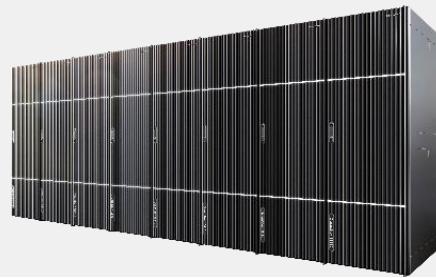
OceanStor 6810



OceanStor 18510



OceanStor 18810



Mid-range

OceanStor 5310



OceanStor 5510



OceanStor 5610



New-Gen OceanStor Hybrid Flash Storage



Fully upgraded to provide ultimate efficiency

- SmartMatrix active-active architecture with load balancing, maximizing performance
- Adaptive layout of hot data, improving performance by 100%



Diversified future-oriented functions

- Compatible with different application ecosystems and supporting various cloudification routes
- Multiple security protection methods, protecting valuable enterprise data

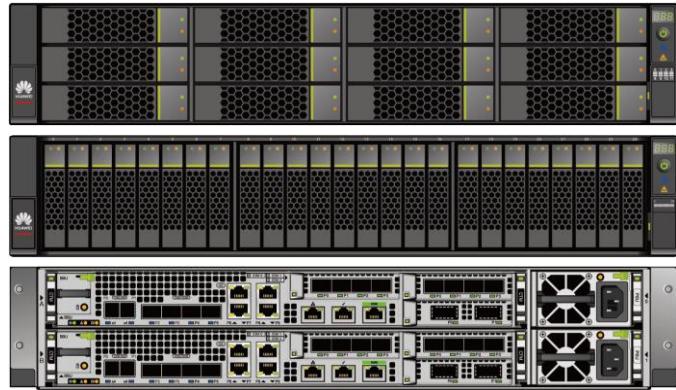


Simpler, smarter, and more economical

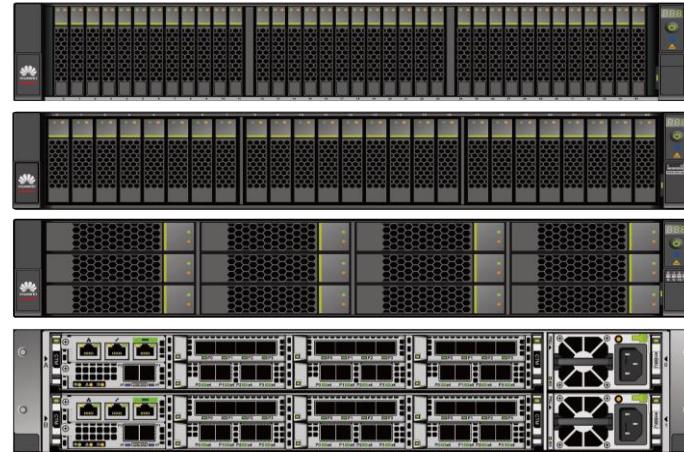
- Cross-generation device interconnection and reuse, SAS HDDs not required, reducing construction costs
- Intelligent and automatic O&M, greatly reducing costs

Device Model Examples

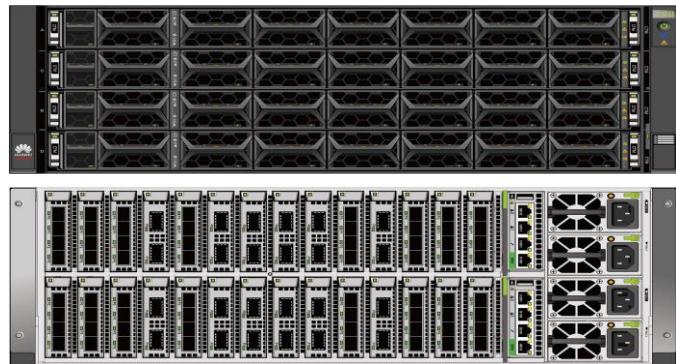
2 U: Huawei new-gen OceanStor hybrid flash storage 5310



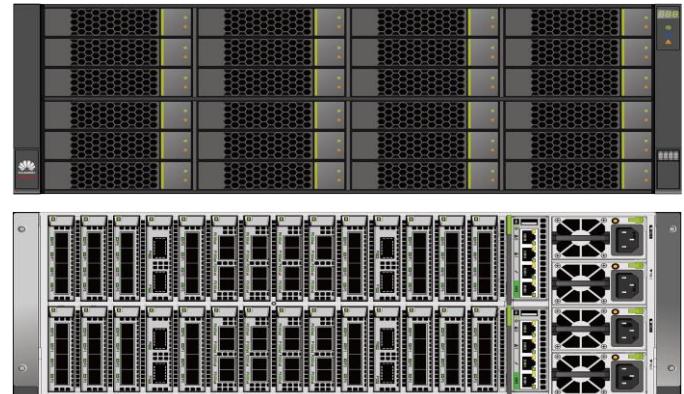
2 U: Huawei new-gen OceanStor hybrid flash storage 5510/5610



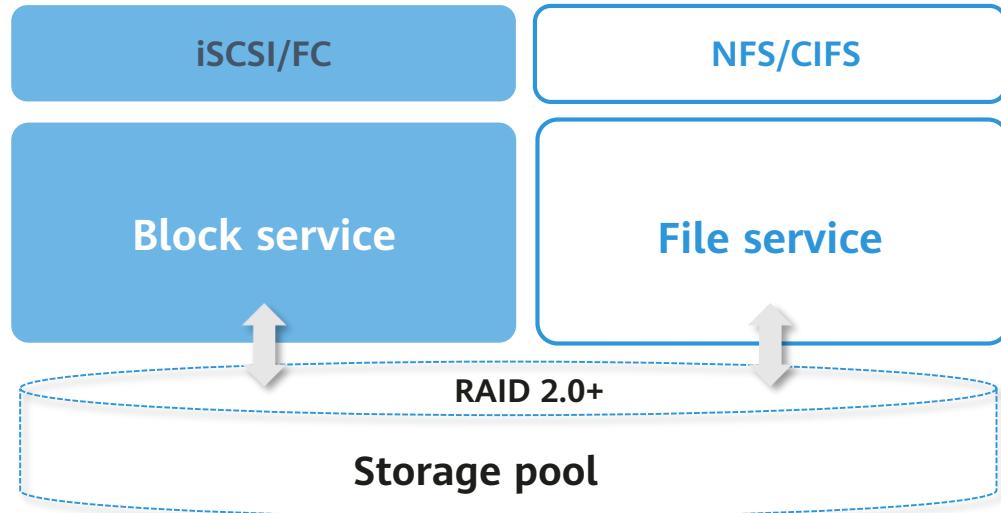
4 U: Huawei new-gen OceanStor hybrid flash storage 6810



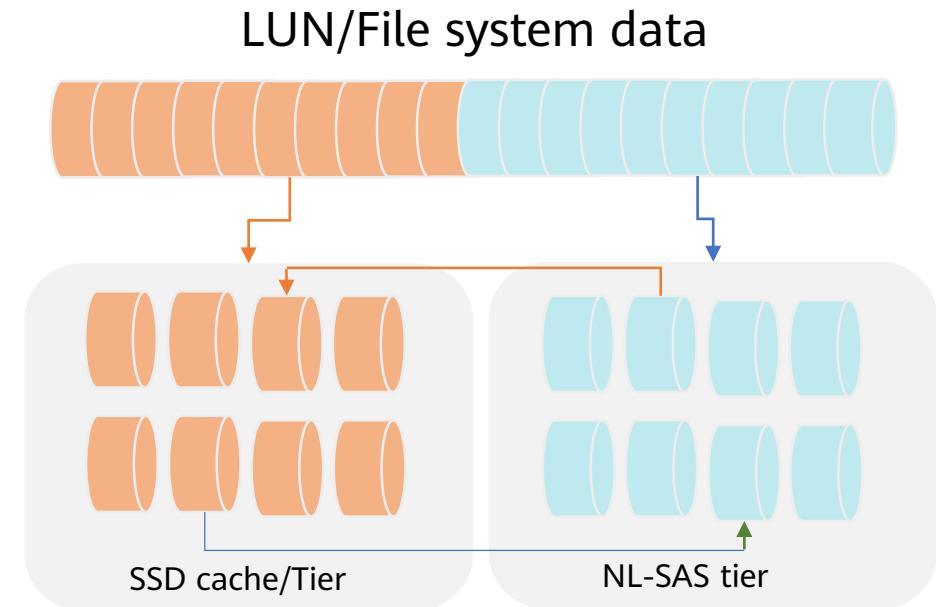
4 U: Huawei new-gen OceanStor hybrid flash storage 18510/18810



Converged SAN and NAS

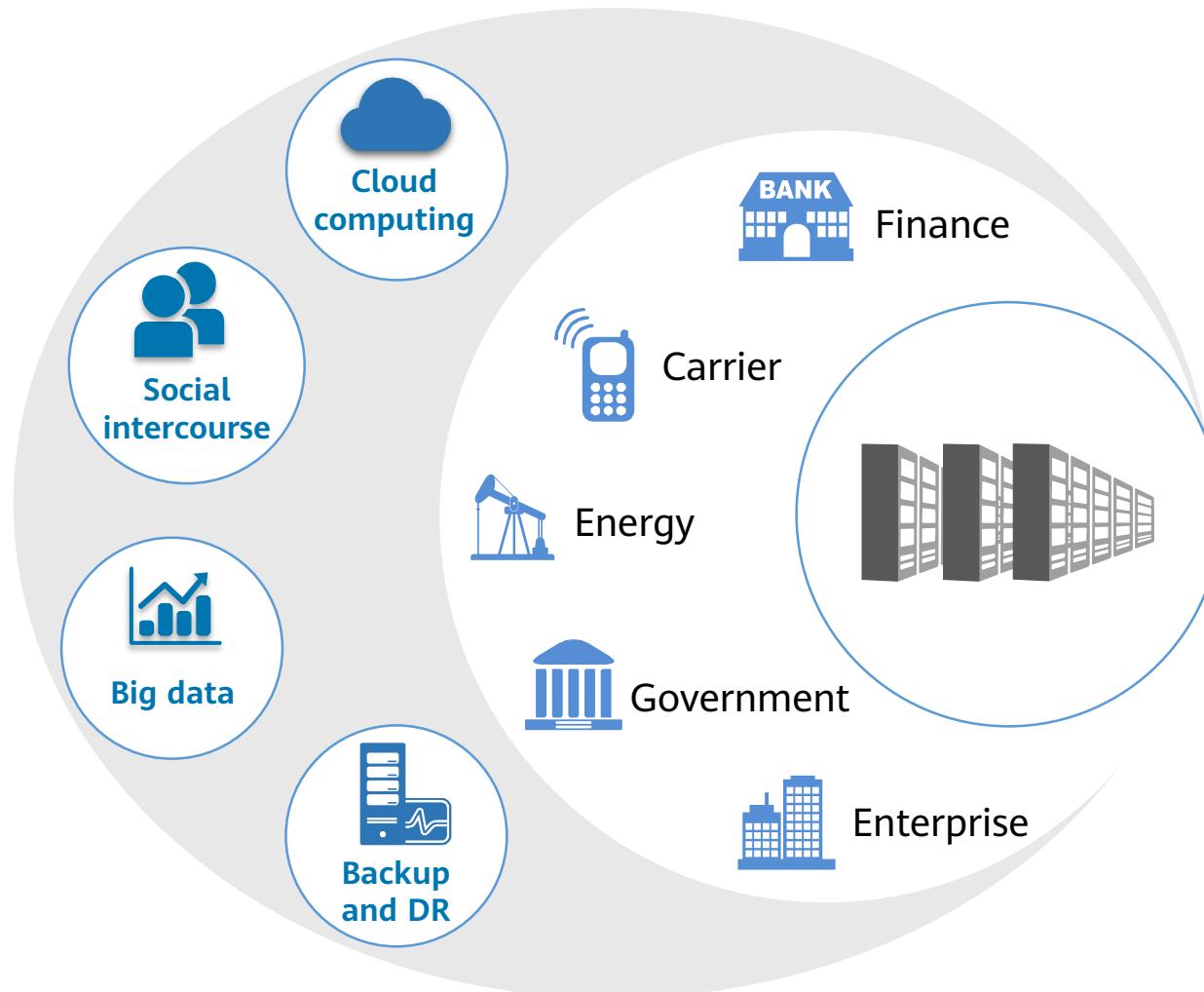


Convergence of SAN and NAS resources of the Huawei new-gen OceanStor hybrid flash series

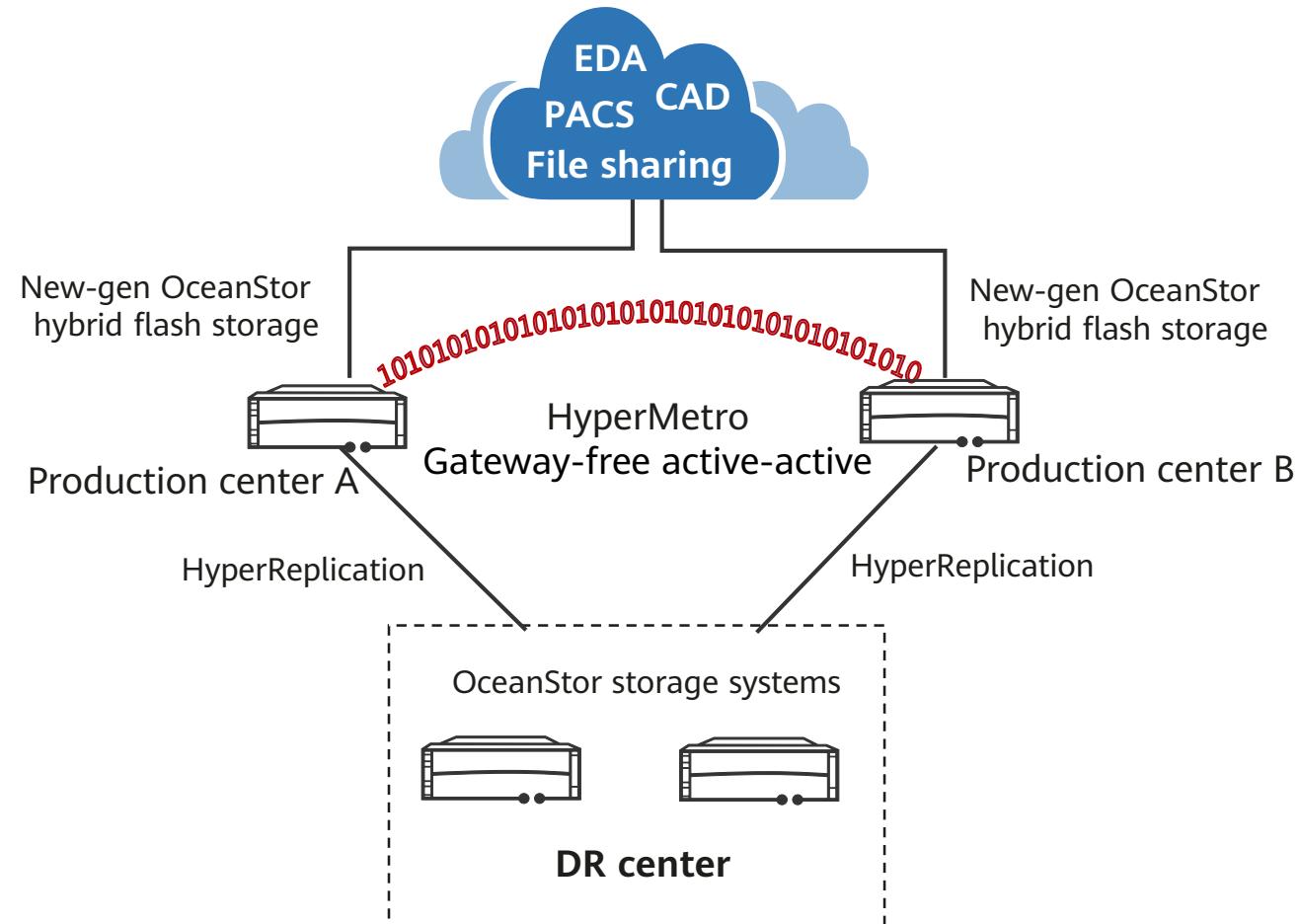


Intelligent tiering cold and hot data of SAN and NAS

Support for Multiple Service Scenarios



Application Scenario – Active-Active Data Centers



Contents

1. Panorama
2. All-Flash Storage
3. Hybrid Flash Storage
- 4. Scale-out Storage**
5. Hyper-Converged Storage
6. Backup Storage

Scale-Out Storage

Performance-oriented

OceanStor Pacific 9950

OceanStor Pacific 9920



Capacity-oriented

OceanStor Pacific 9550

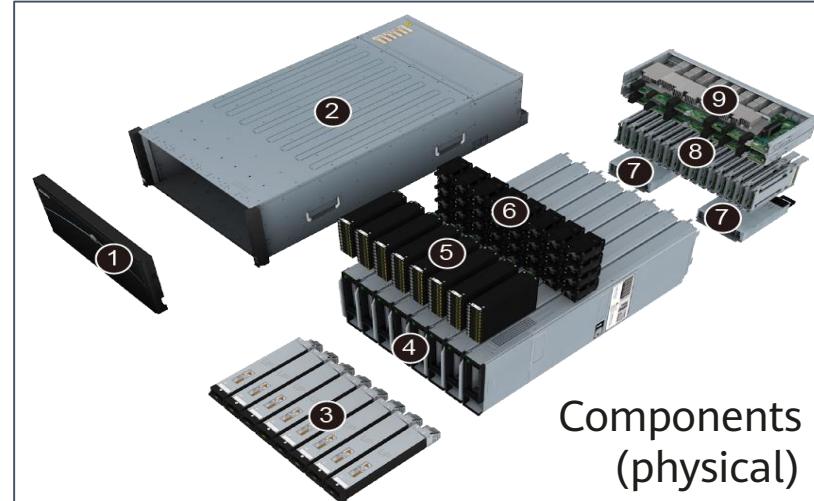
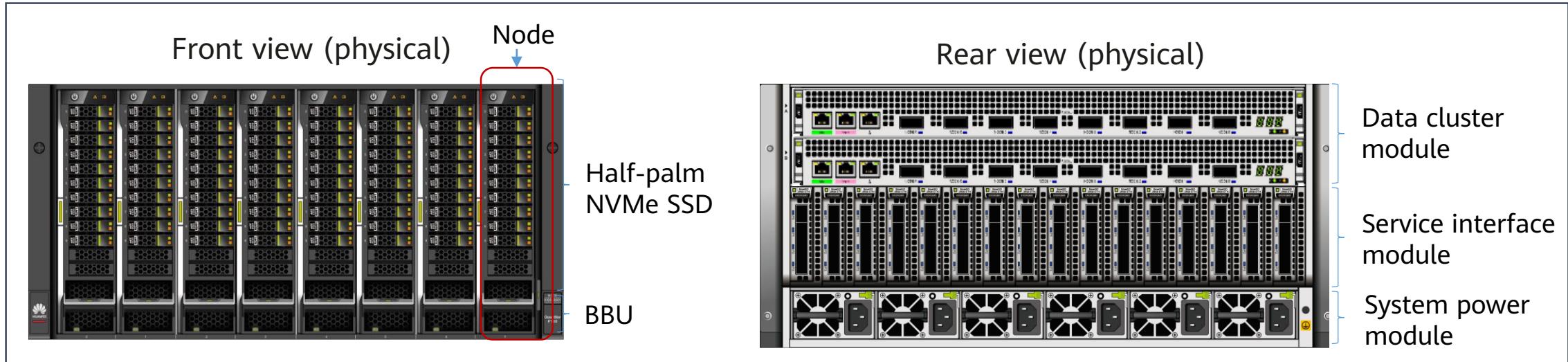
OceanStor Pacific 9520



OceanStor Pacific 9540



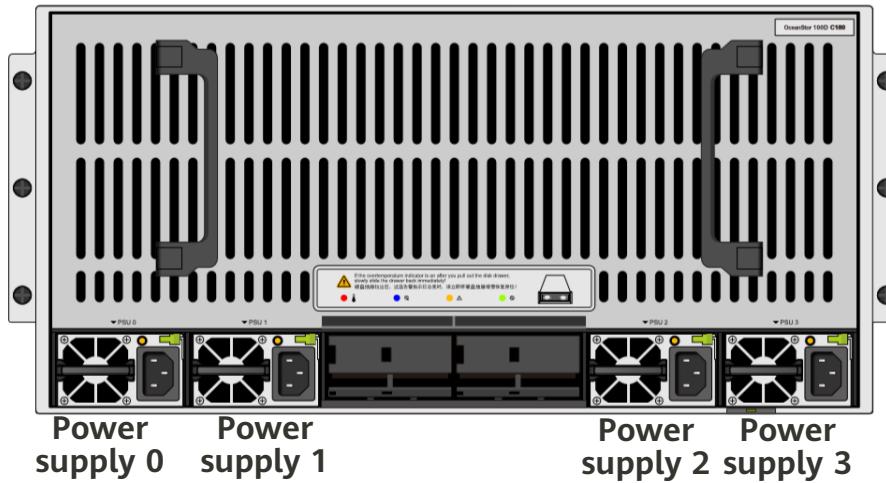
Appearance of OceanStor Pacific 9950



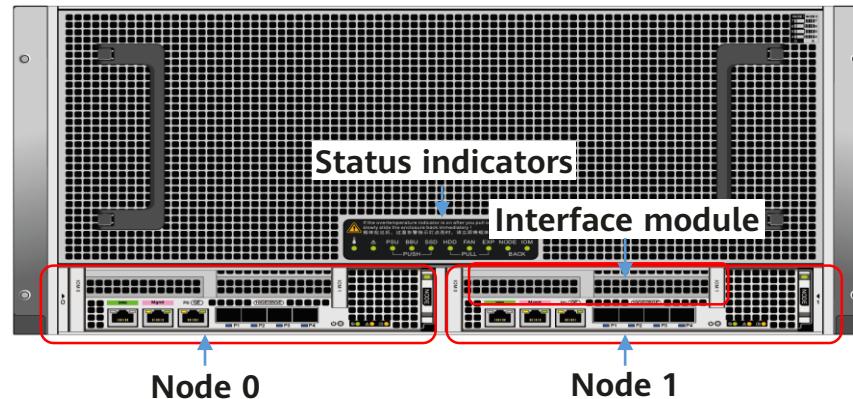
1	Front panel cover	2	System subrack
3	BBU	4	Node
5	Half-palm NVMe SSD	6	Fan module
7	Power module	8	Interface module
9	Data cluster module		

Appearance of OceanStor Pacific 9550

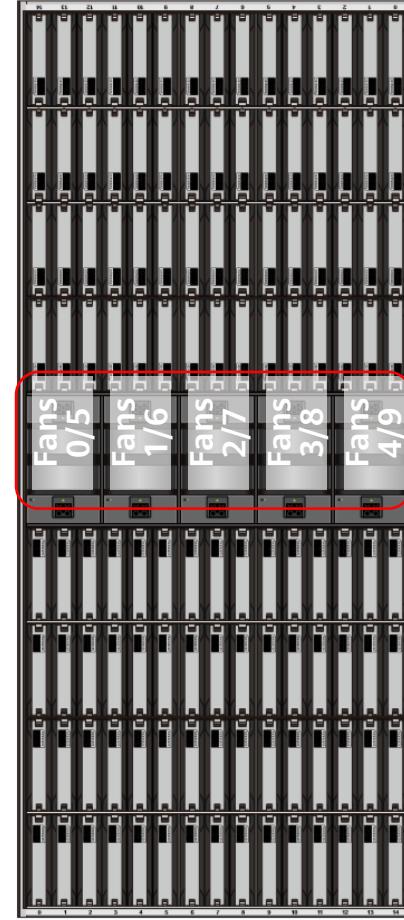
Front view (without the cover)



Rear View

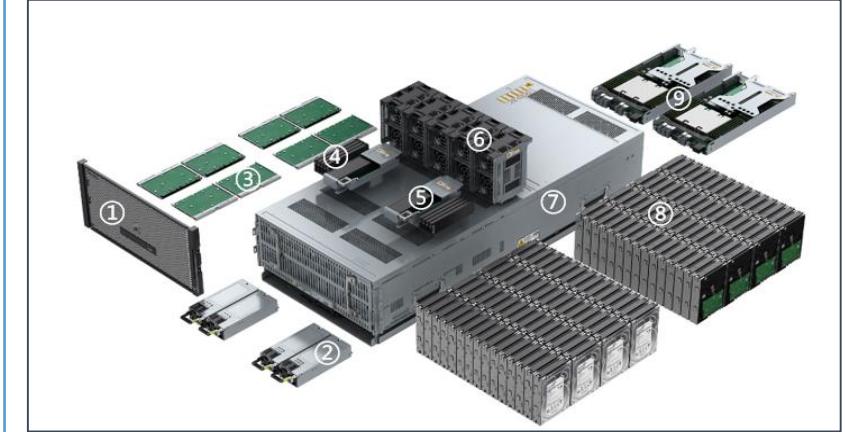


Top view



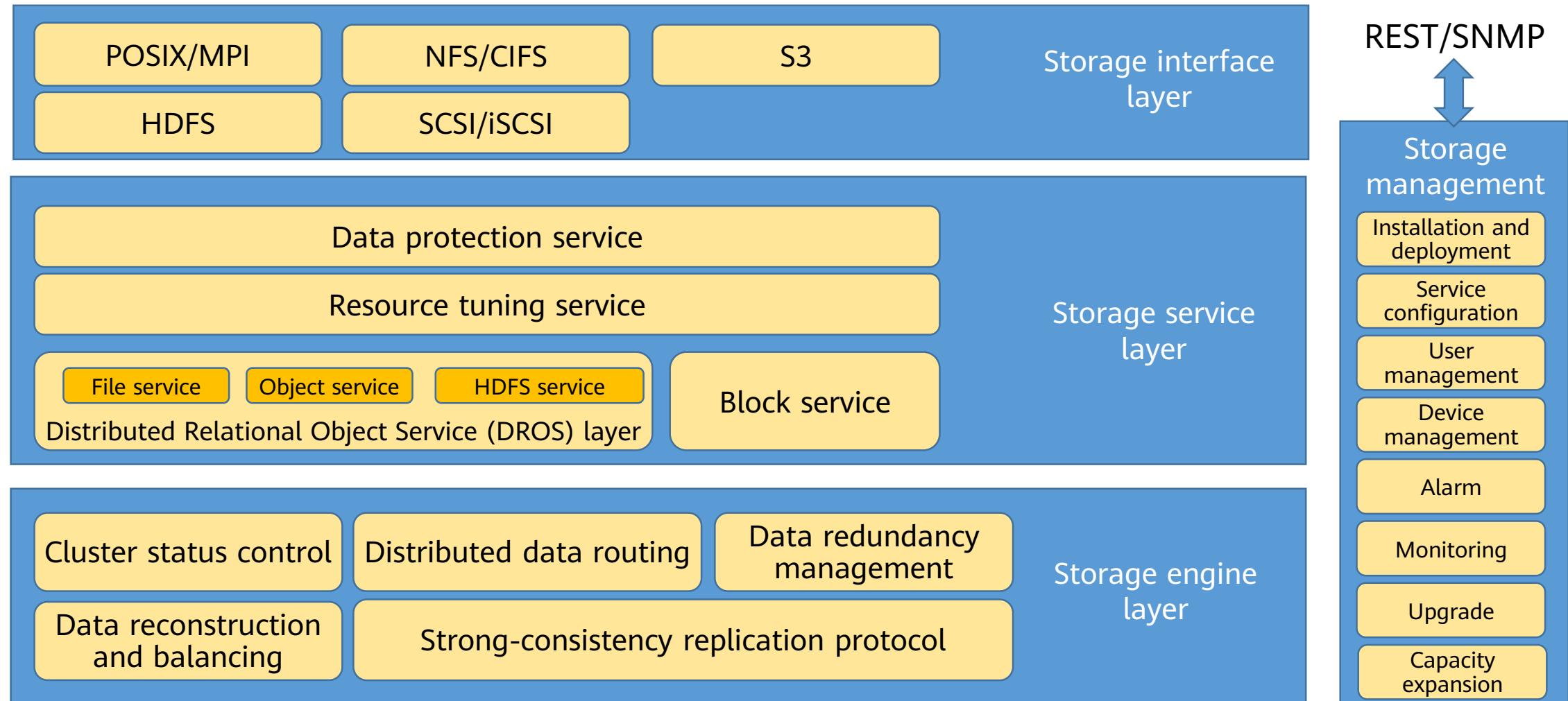
Rear

Components (physical)

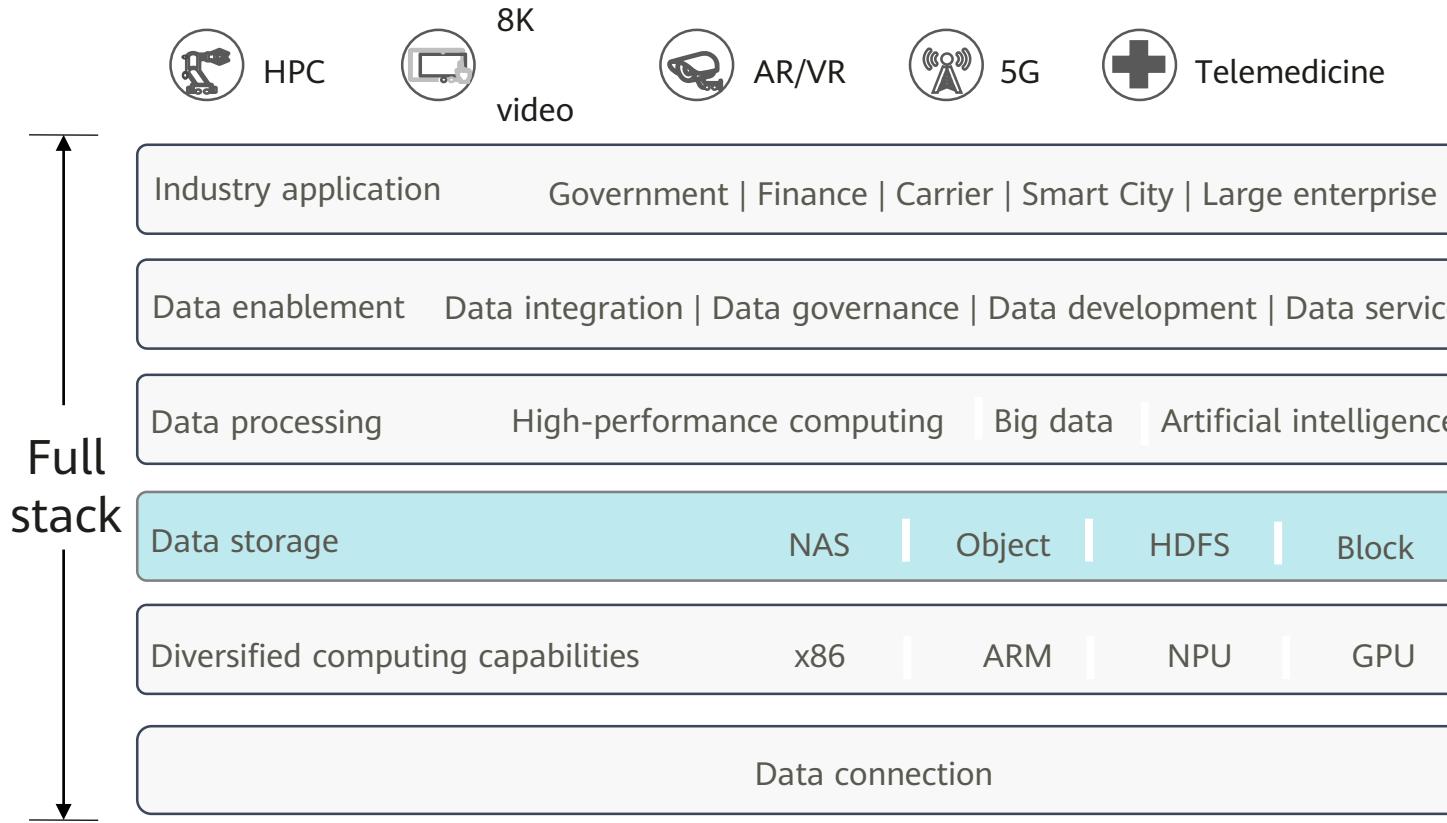


1	Front panel cover	2	Power module
3	Expansion module	4	Half-palm NVMe SSD
5	BBU	6	Fan module
7	Chassis	8	3.5-inch HDD
9	Node		

Overall Software Architecture



Product Features

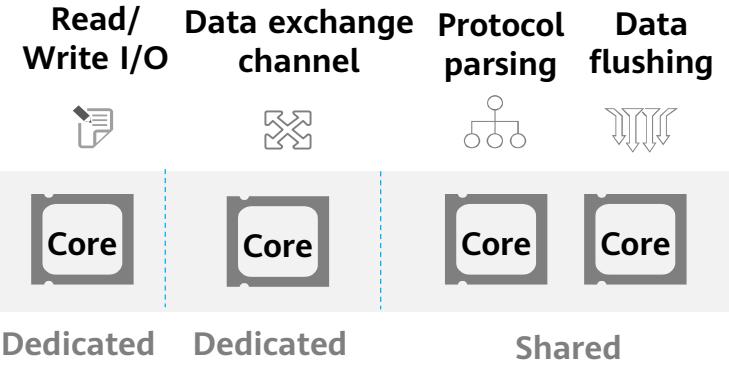


OceanStor Pacific series intelligent scale-out storage

- **High efficiency**
Multi-protocol interworking and zero data migration
- **Secure storage**
Solution-, system-, device-, and I/O-level HA designs, ensuring 24/7 service continuity
- **Cost reduction**
Full-stack optimization from hardware to solutions and from algorithms to the architecture, reducing TCO by more than 30%

FlashLink - Multi-core Technology

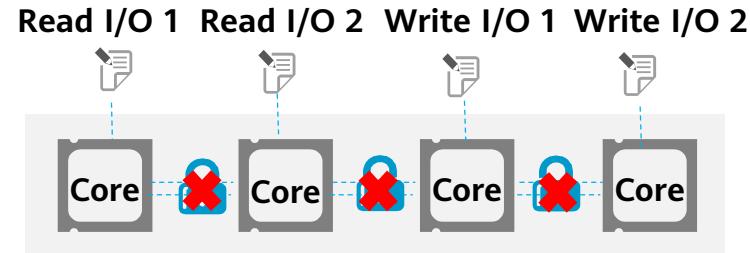
CPU grouping algorithm



I/O grouping to avoid mutual interference

Dedicated cores are used to ensure the key service resource investment and reduce the latency. Shared cores are used to balance the load of multiple services.

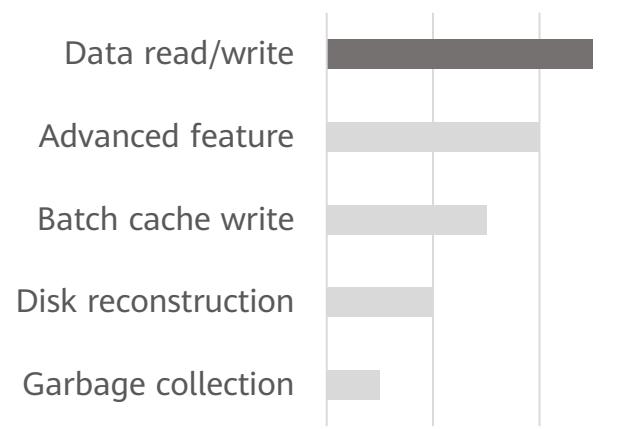
CPU core-based algorithm



Continuous I/O execution to avoid consumption caused by switchover

A request is continuously processed on the same core to prevent thread switchovers for atomic and lock-free operations. This avoids frequent multi-core switchovers and improves the CPU cache hit ratio.

Intelligent I/O scheduling

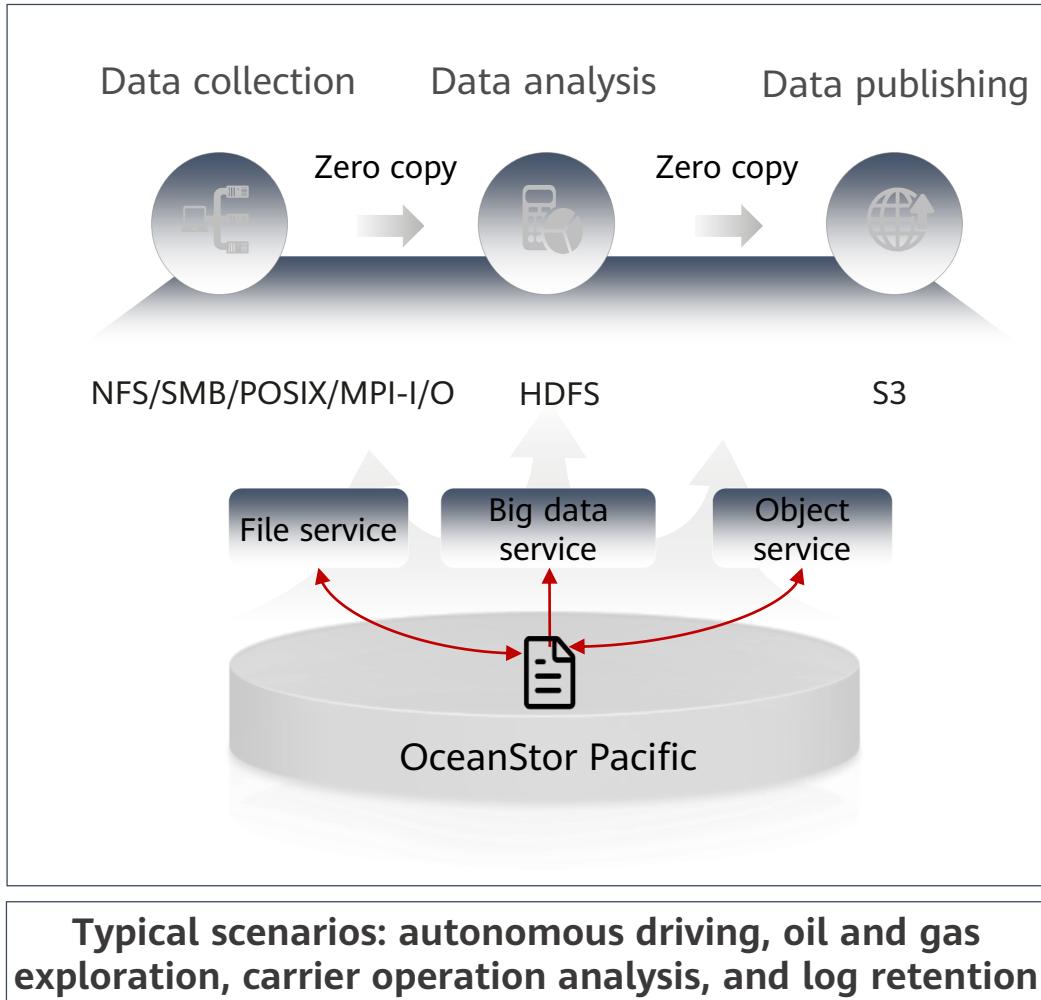


I/O priority guarantee, reducing impact on services

Data read and write I/Os always have the highest priority to ensure the lowest latency.

Continuous low latency ensures fast response of mission-critical services.

SmartInterworking



Improved analysis efficiency

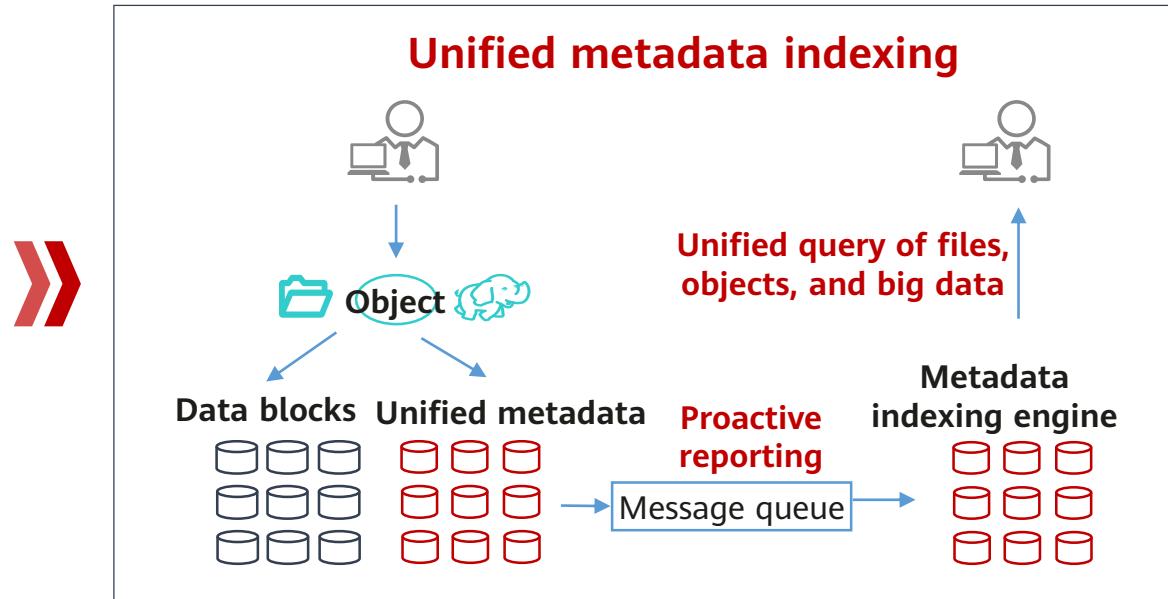
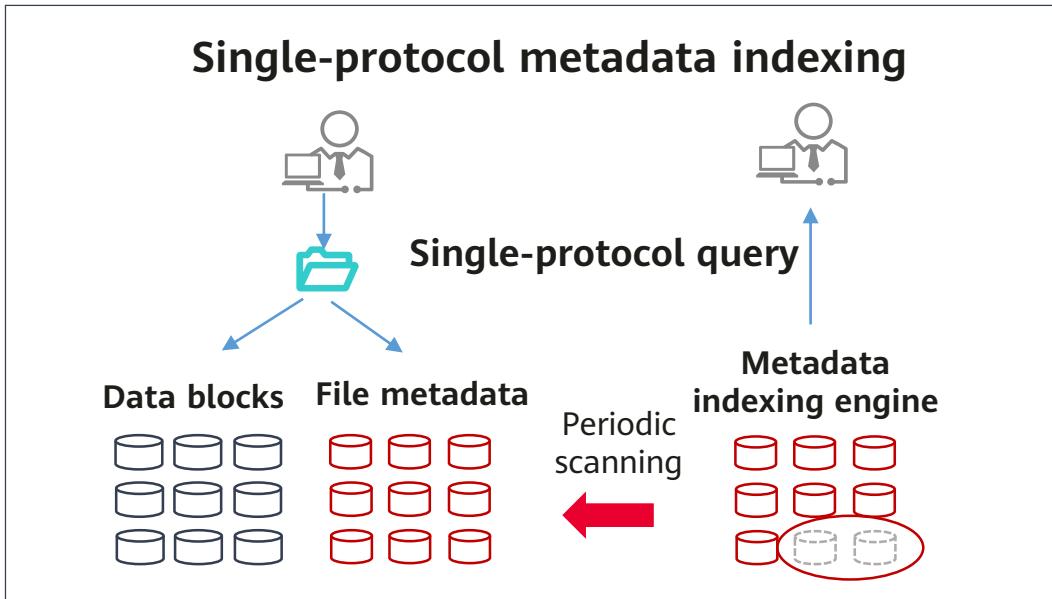
Multiple protocols share the same data without migration required.

Saved storage space

Repeated data storage is avoided, saving space overhead.

Specifications	Product P of Vendor D	Product E of Vendor D	Product H of Vendor H	OceanStor Pacific
Native semantics	✓	✗	✗	✓
Gateway-free	✓	✗	✓	✓
Semantic integrity	Medium	Low	Low	High
Impact on performance	Medium	High	Medium	Low
Cross-protocol real-time access	✓	✓	✗	✓
Multi-protocol sharing	✗	✗	✗	✓

SmartIndexing



Multi-protocol unified search

Unified metadata for object, file, and big data storage



Proactive reporting of updates

Proactive reporting of metadata updates, improving intelligence

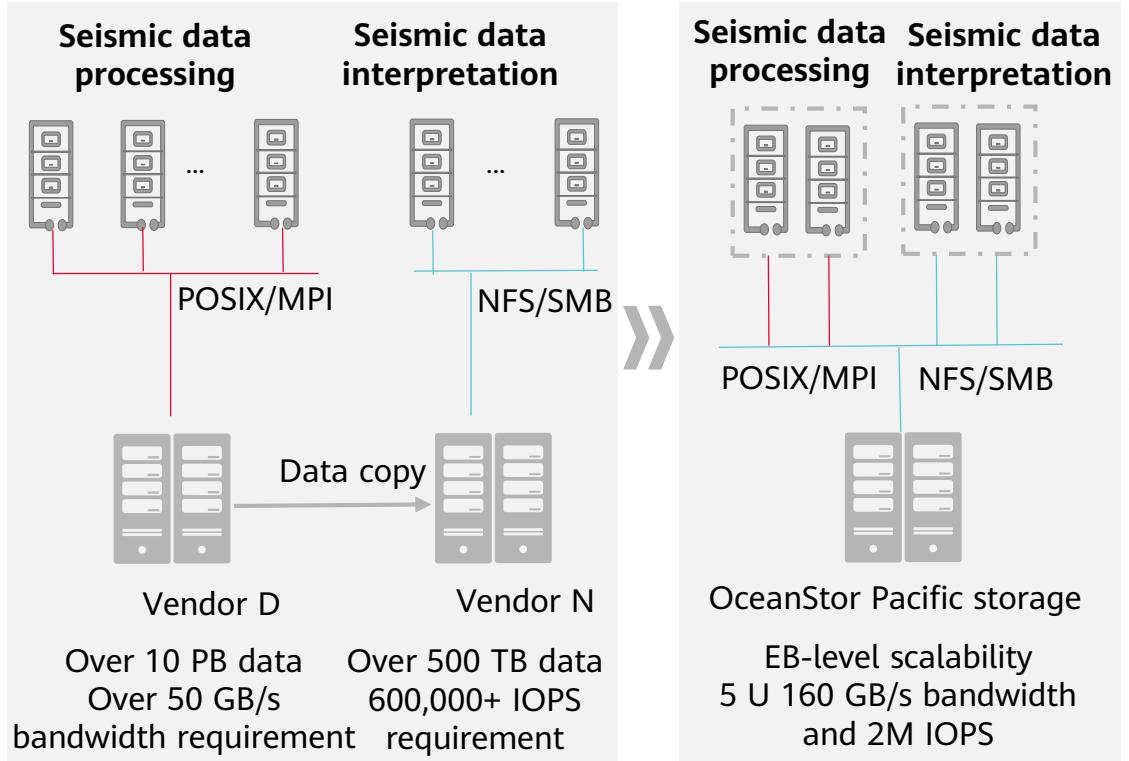


Results returned in seconds

Customized search for hundreds of billions of metadata records

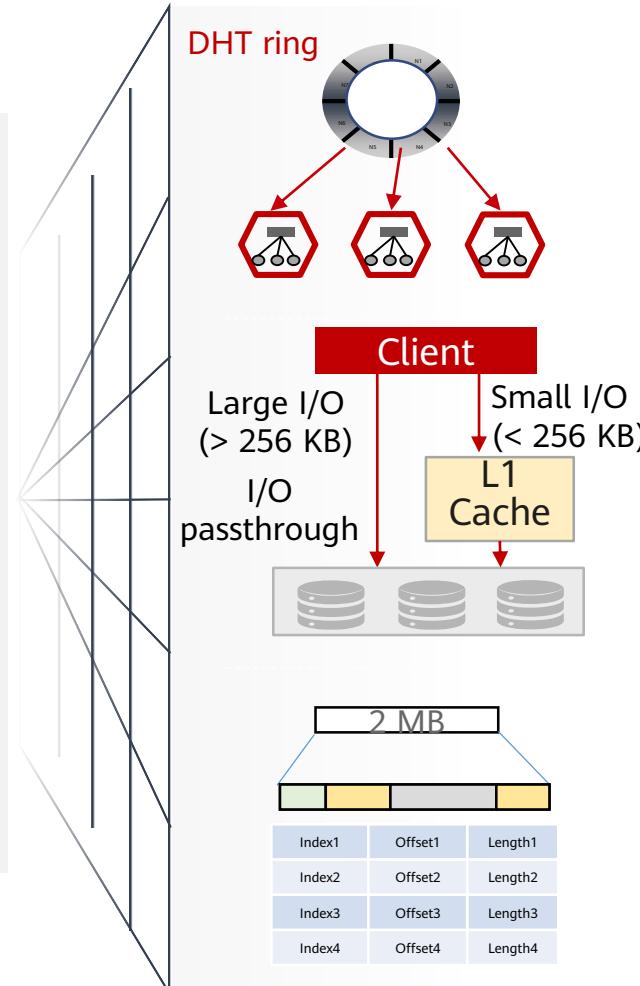
Parallel File System

- Hybrid load-based design



One storage system meets the requirements of large files with high bandwidth, small files with high OPS, and MPI-IO.

No data copy is required, improving resource utilization.



Architecture: distributed lock-free design

Metadata is scattered and owned based on directories and small I/O forwarding is processed by the owning nodes, eliminating distributed lock overheads and greatly reducing the read/write latency of small I/Os.

I/O: large I/O passthrough and small I/O aggregation

Large I/Os are written to the persistence layer to reduce forwarding. Small I/Os are written after aggregation at the cache layer, reducing the number of I/O interactions and latency.

Disk: multi-granularity disk space management

Two-level indexes. The primary index uses the fixed-length granularity index to ensure large I/O performance. Sub-indexes are automatically adapted based on I/O sizes to flexibly cope with small I/O read and write.

Distributed Parallel Client (DPC)



MPI-IO accelerates meteorological data analysis and application.

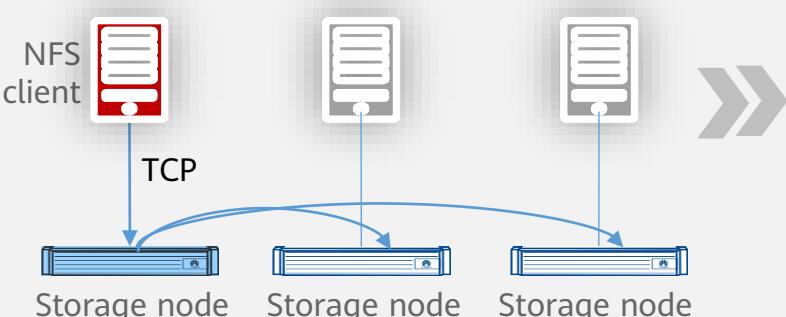


High single-stream performance ensures the receipt of satellite data.



A high-performance client unleashes the potentials of a fat client.

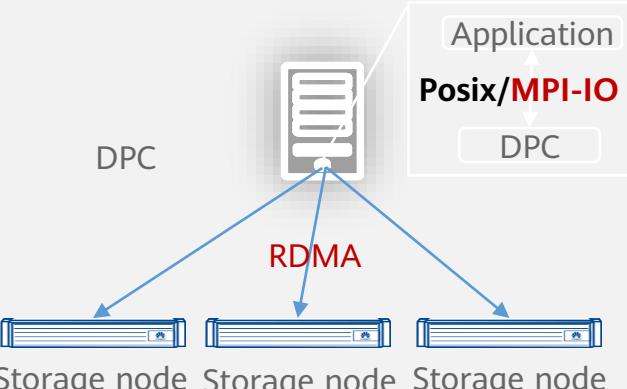
NFS client access mode



The diagram shows a single 'NFS client' icon connected by a blue line labeled 'TCP' to three separate 'Storage node' icons. A large grey arrow points to the right, leading to the 'DPC client access mode' diagram.

- A single client connects to one storage node.
- MPI-IO is not supported.
- TCP network access is supported.

DPC client access mode

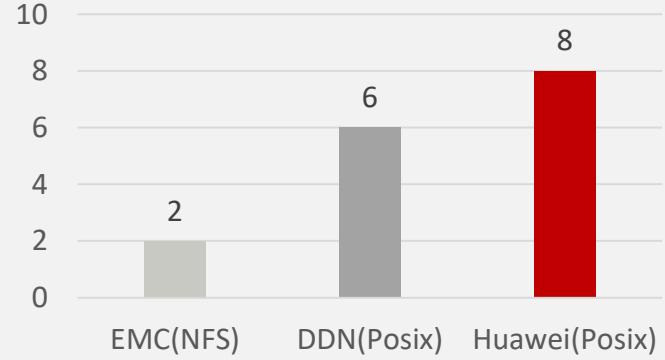


The diagram shows a single 'DPC' icon connected by a blue line labeled 'RDMA' to three separate 'Storage node' icons. The 'DPC' icon also contains a smaller box labeled 'Application Posix/MPI-IO' with a downward arrow pointing to a 'DPC' box.

- A single client connects to **multiple** storage nodes.
- **MPI-IO** is supported.
- **RDMA** network access is supported.

VS.

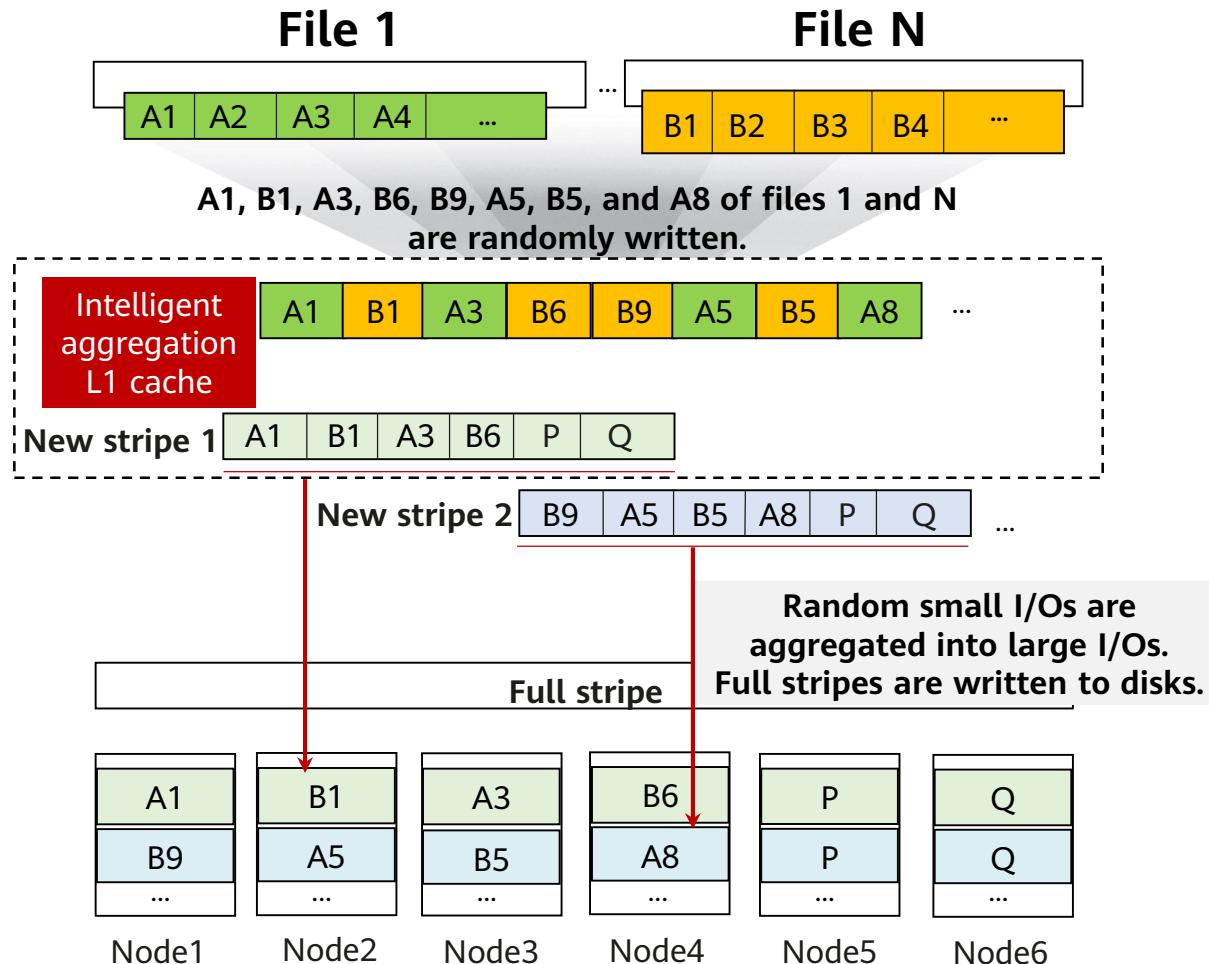
Maximum single-stream bandwidth (GB/s)



Protocol	Maximum Single-Stream Bandwidth (GB/s)
EMC(NFS)	2
DDN(Posix)	6
Huawei(Posix)	8

Up to **8 GB/s** single-stream bandwidth, which is leading in the industry and more than four times that of the NFS protocol.

Intelligent Stripe Aggregation



Large I/O passthrough

Large I/Os form EC stripes and are directly written to disks without being cached, saving cache resources and prolonging the cache service life.

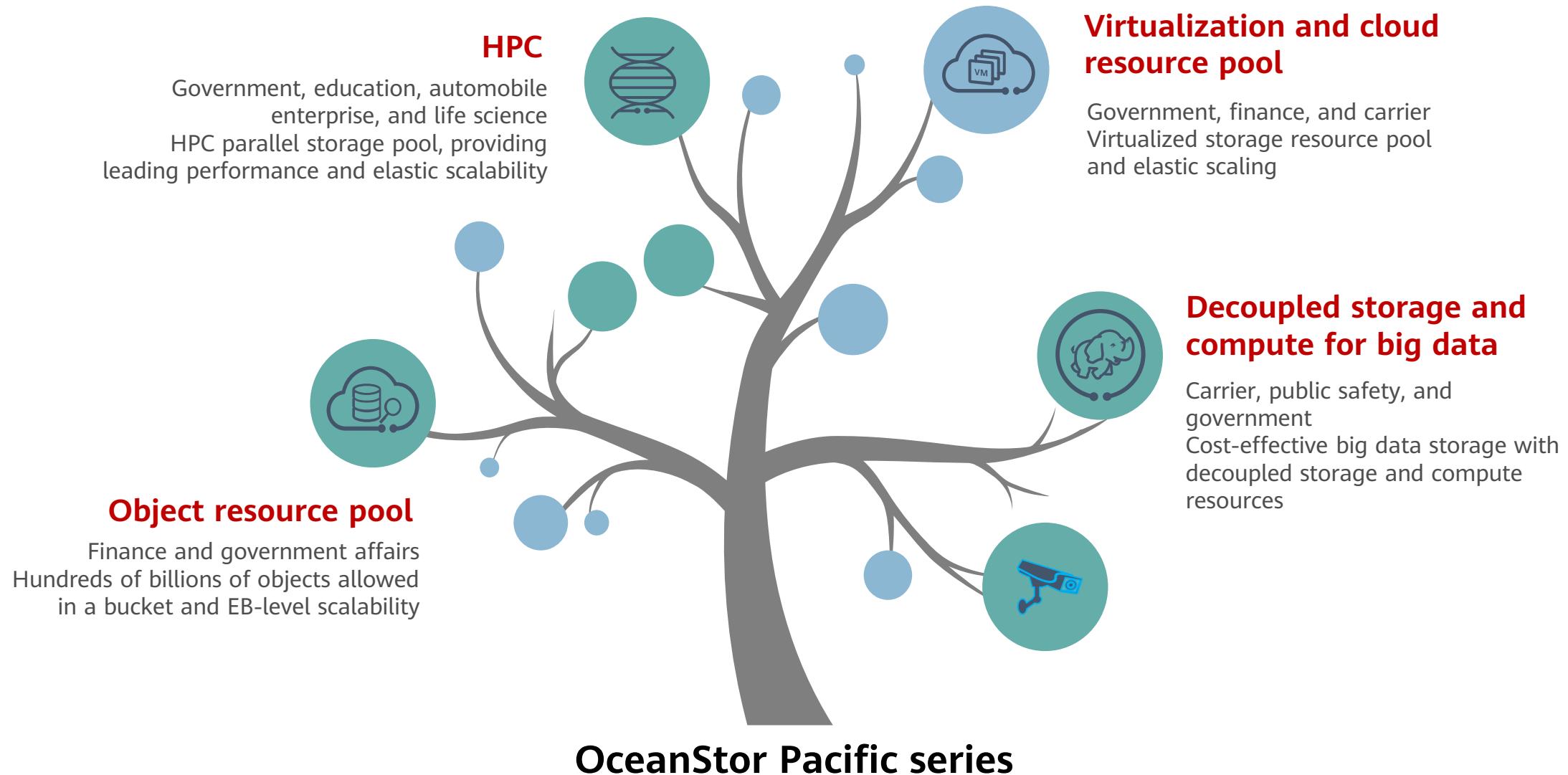
Intelligent small I/O aggregation

I/Os are aggregated into full EC stripes and then written to disks to reduce the write latency.

Intelligent append write

Append writes replace traditional overwrite operations. Random writes change to 100% sequential writes. Small block write is avoided, and the write performance of HDDs is maximized.

Solutions for Four Typical Scenarios



Contents

1. Panorama
2. All-Flash Storage
3. Hybrid Flash Storage
4. Scale-Out Storage
- 5. Hyper-Converged Storage**
6. Backup Storage

Hyper-Converged Series

Virtualization/VDI



FusionCube 1000H

Database



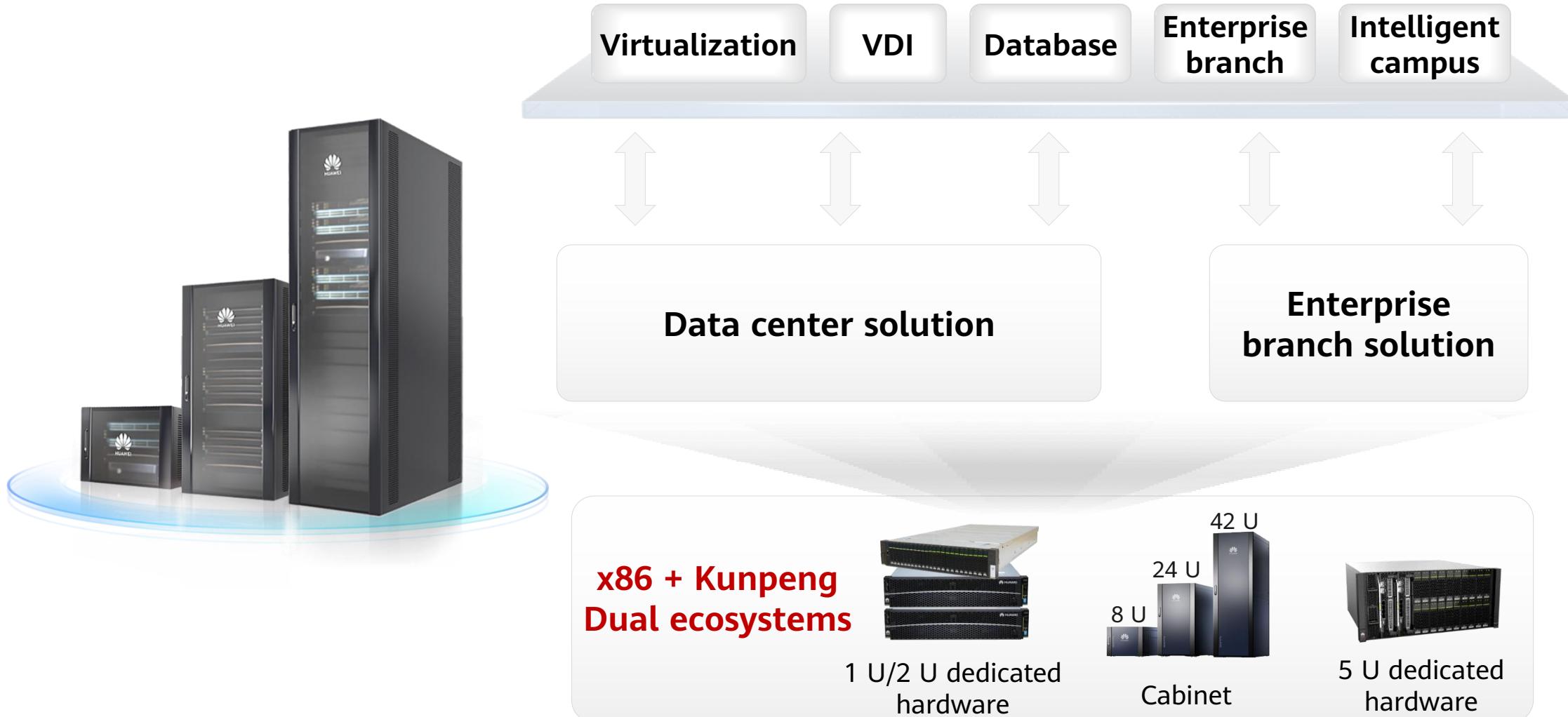
FusionCube 1000D

Enterprise branch

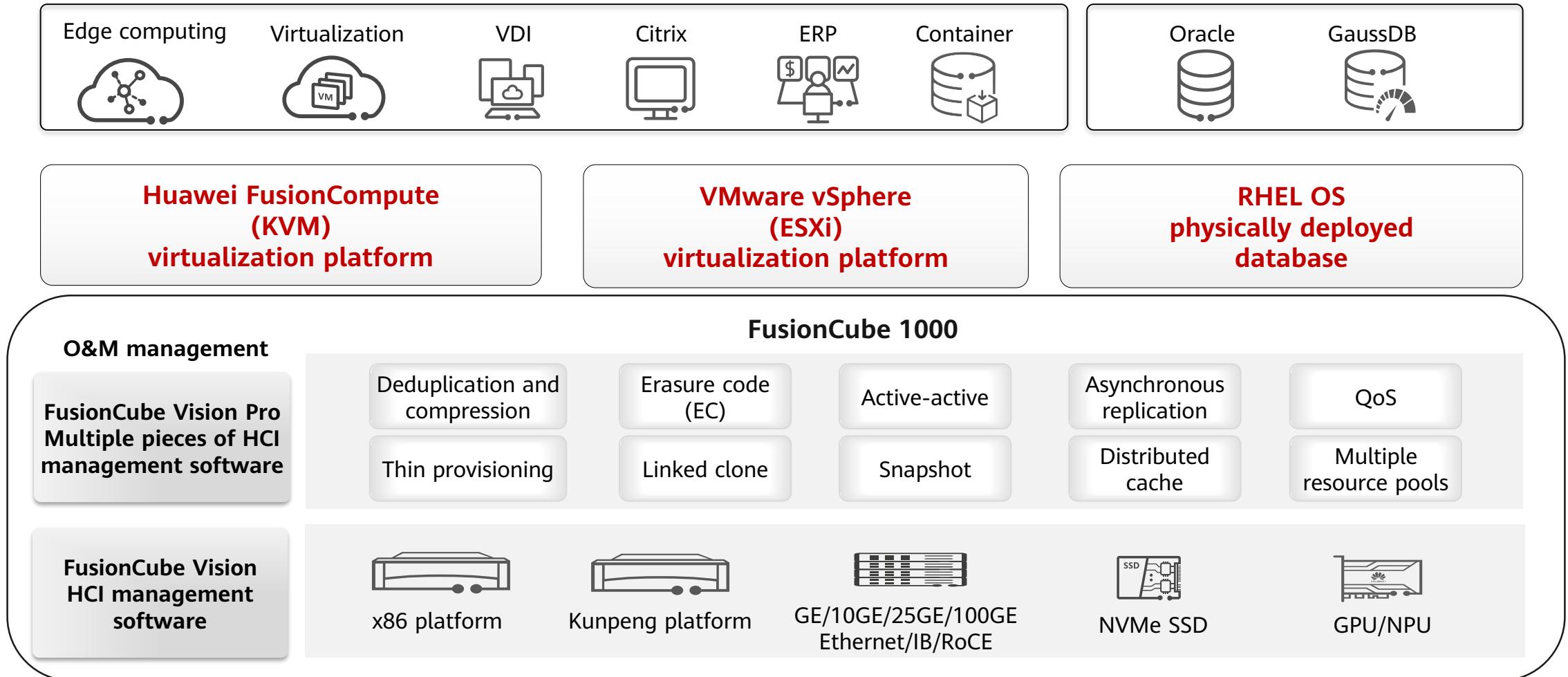


FusionCube 1000C

FusionCube Scenarios and Solutions

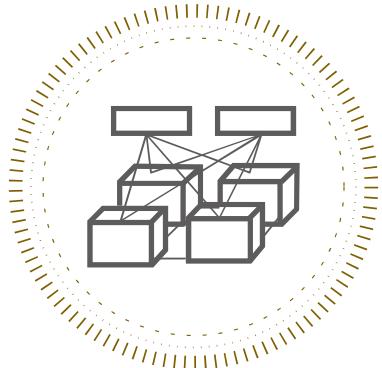


Logical Architecture of the Hyper-converged Data Center Solution



Highlights

Convergence and efficiency



Convergence of virtualization, container, database, desktop cloud, and storage

The minimum configuration requirement is lowered to two nodes.

EC, deduplication, and compression improve the capacity utilization **by three times**.

Four types of services provide convergency and efficiency.

Rock-solid reliability



Ten-level protection for devices, in data centers, and between data centers

Subhealth detection and end-to-end DIF ensure ultimate data reliability.

The **real active-active** architecture ensures 99.9999% reliability and service continuity.

Ten-level protection ensures ultimate reliability.

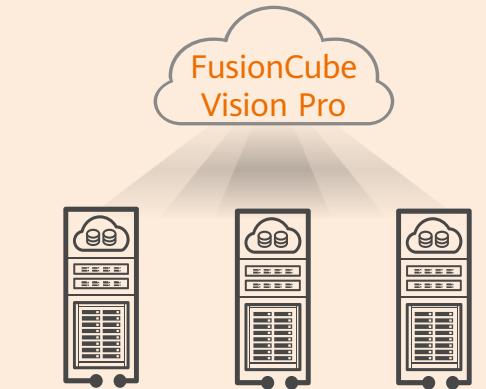
Cloud-based and easy-to-use



Four-level management:
Hyper-convergence, multi-site, resource pool, and service-oriented
Out-of-the-box hyper-convergence and **one-click delivery**
Unified management platform and **simplified O&M**

Cloud-based and easy-to-use four-level management is supported.

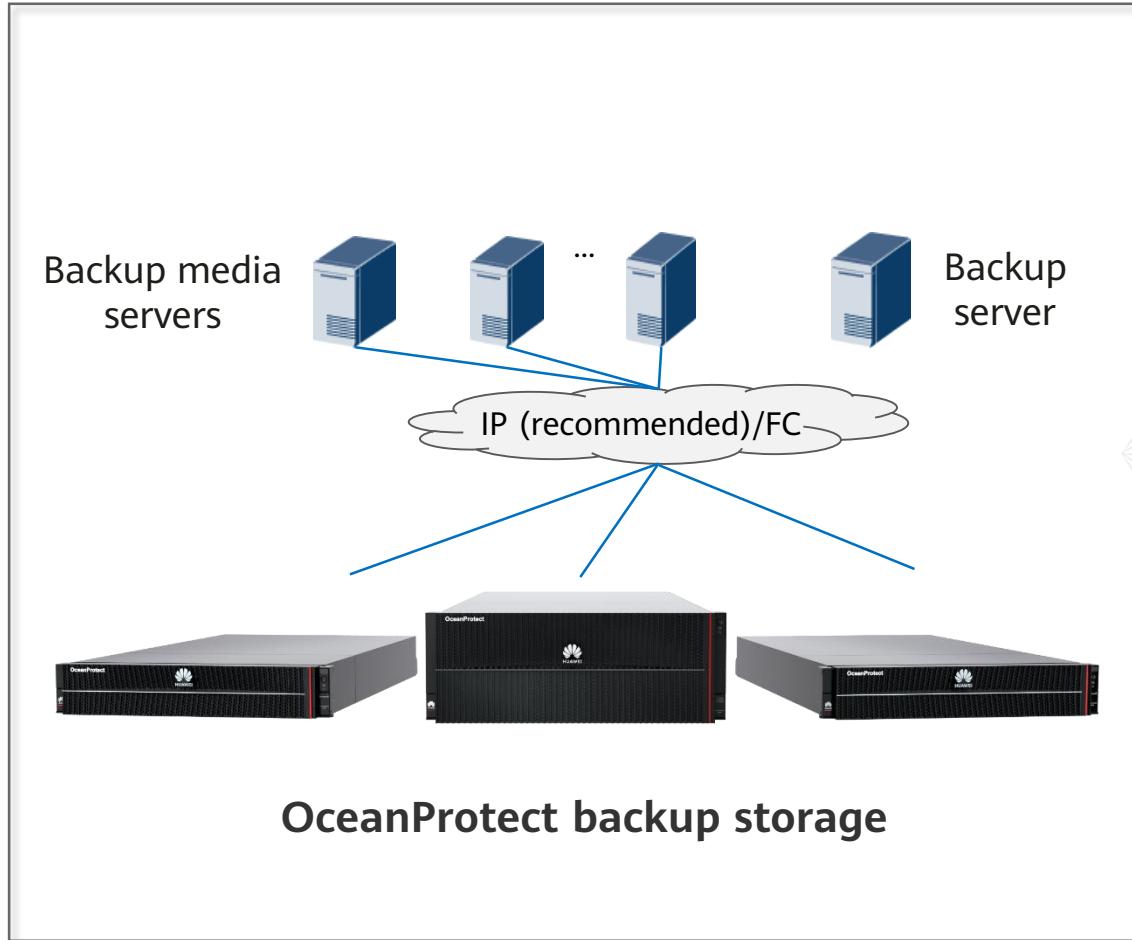
Enterprise Branch: Easy Storage, Unified Management, and Cloud Synergy

Easy storage	Unified management	Cloud synergy
<p>Fast deployment Plug-and-play One-click configuration</p>   <p>Storage Compute AI Network Security</p> <p>Delivery period shortened by 50%</p> <ul style="list-style-type: none">• Full-stack design and one-stop delivery• Plug-and-play and one-click configuration	<p>Unified management Centralized monitoring</p>  <p>Device VM Container Application</p> <p>Unified management of one set of software</p> <ul style="list-style-type: none">• Application, VM, container, server, firewall, switch, gateway, storage device, UPS, and sensor• Visualized remote management of 20,000+ sites	<p>Algorithm collaboration Data collaboration</p>   <p>Inference Training</p> <p>Intelligent autonomy</p> <ul style="list-style-type: none">• Central training and near-end inference• Cloud-based training and near-end inference• Centralized data backup and fast restoration

Contents

1. Panorama
2. All-Flash Storage
3. Hybrid Flash Storage
4. Scale-Out Storage
5. Hyper-Converged Storage
- 6. Backup Storage**

Backup Storage



Rapid backup and restoration

End-to-end backup acceleration

Instant recovery



Efficient reduction

Backup data preprocessing
and variable-length deduplication

Compression and compaction



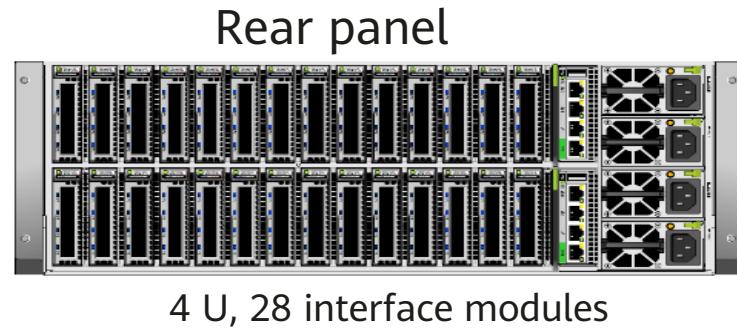
High reliability

Data consistency verification
System-level reliability
Proactive predictive O&M

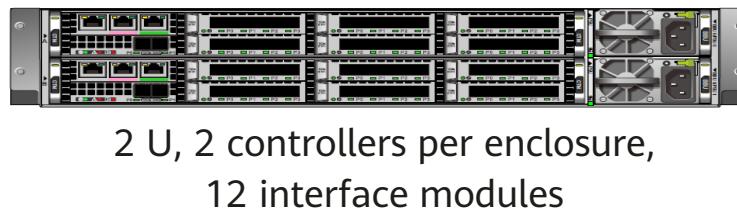


Device Model Examples

OceanProtect X9000 storage controller enclosure



OceanProtect X6000/X8000 storage controller enclosure



SAS SSD disk enclosure



Front panel



4 U, 4 controllers per enclosure



2 U, 25 SAS SSDs



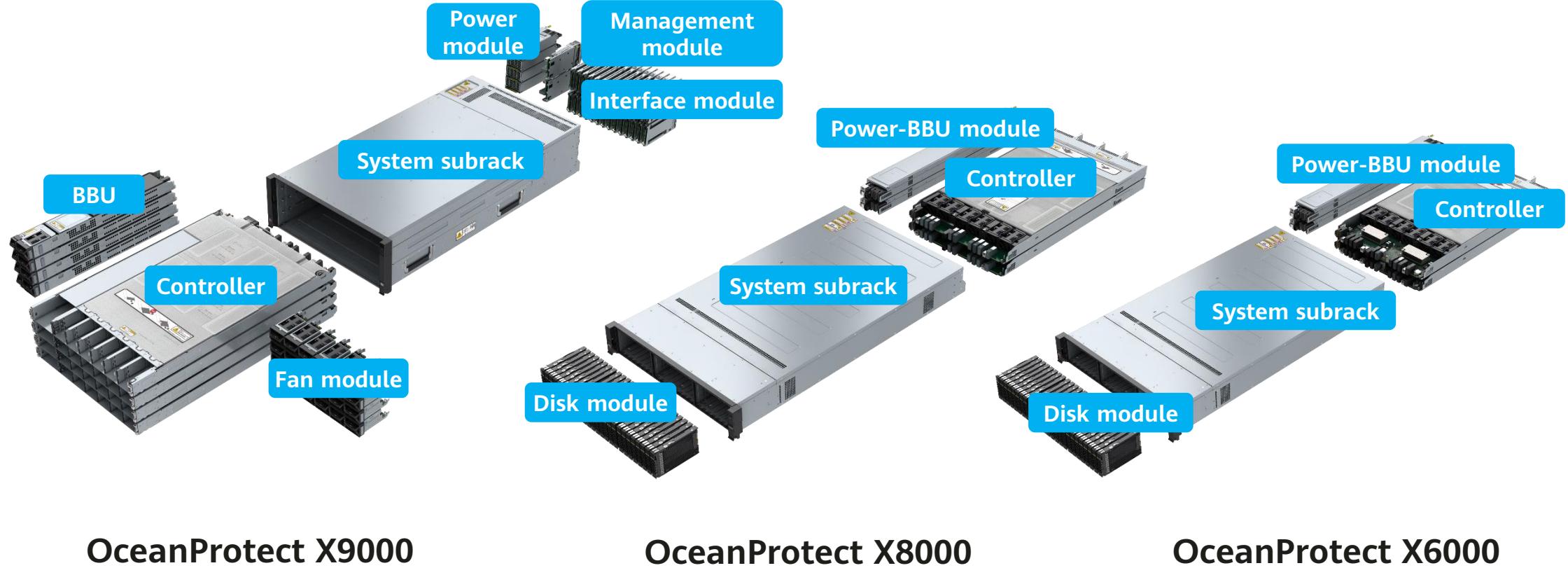
2 U, 25 SAS SSDs

NL-SAS disk enclosure

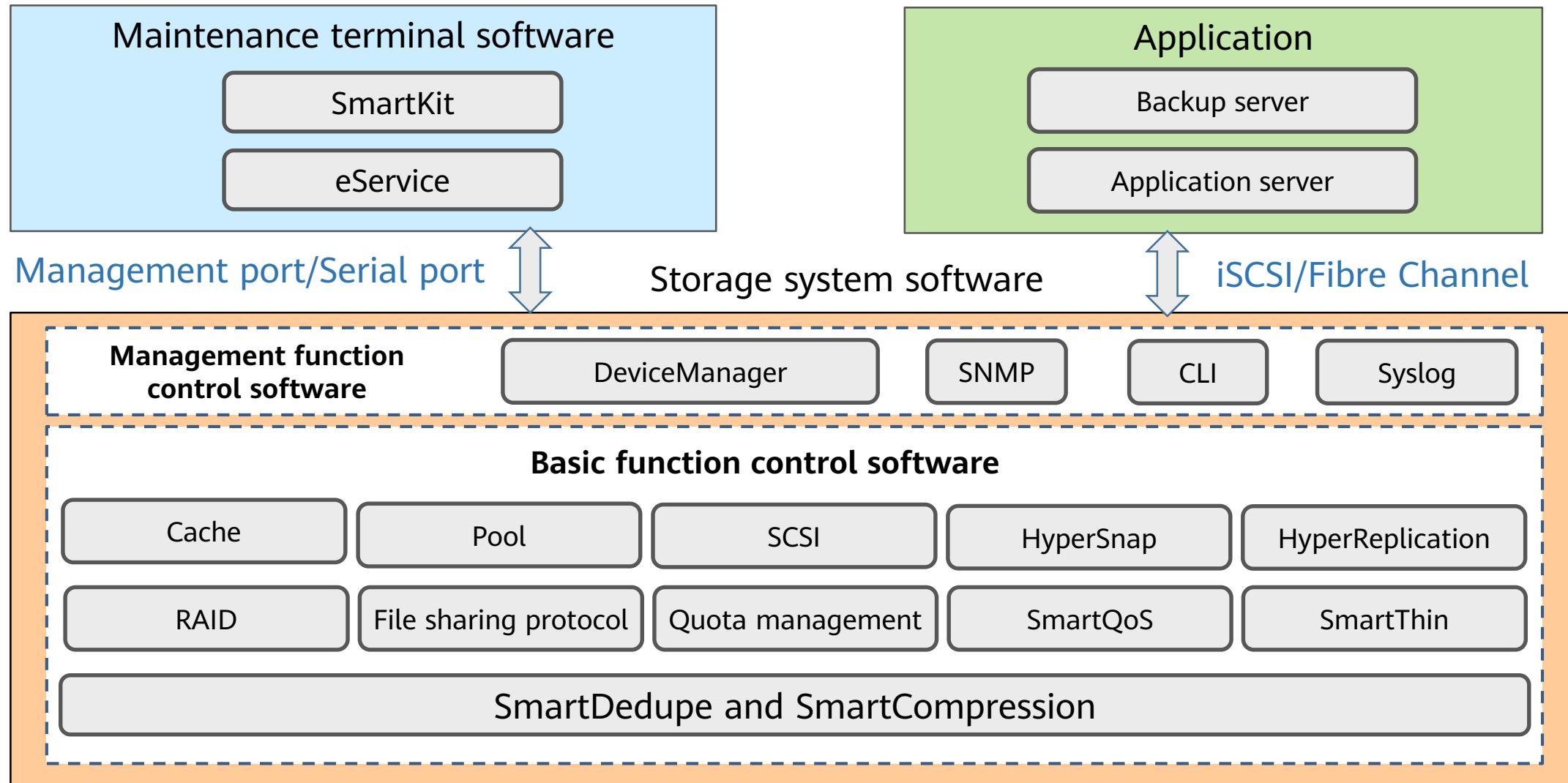


4 U, 24 NL-SAS disks

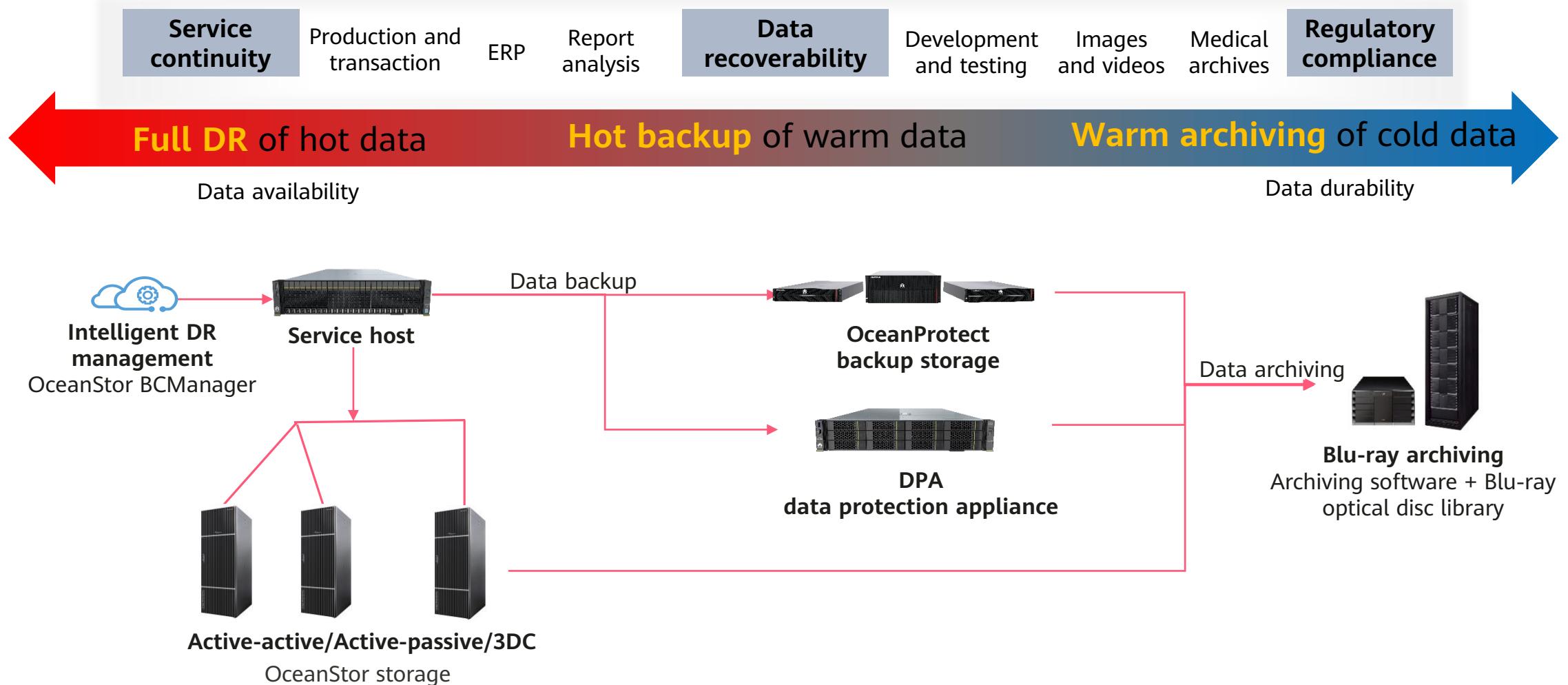
Hardware Architecture



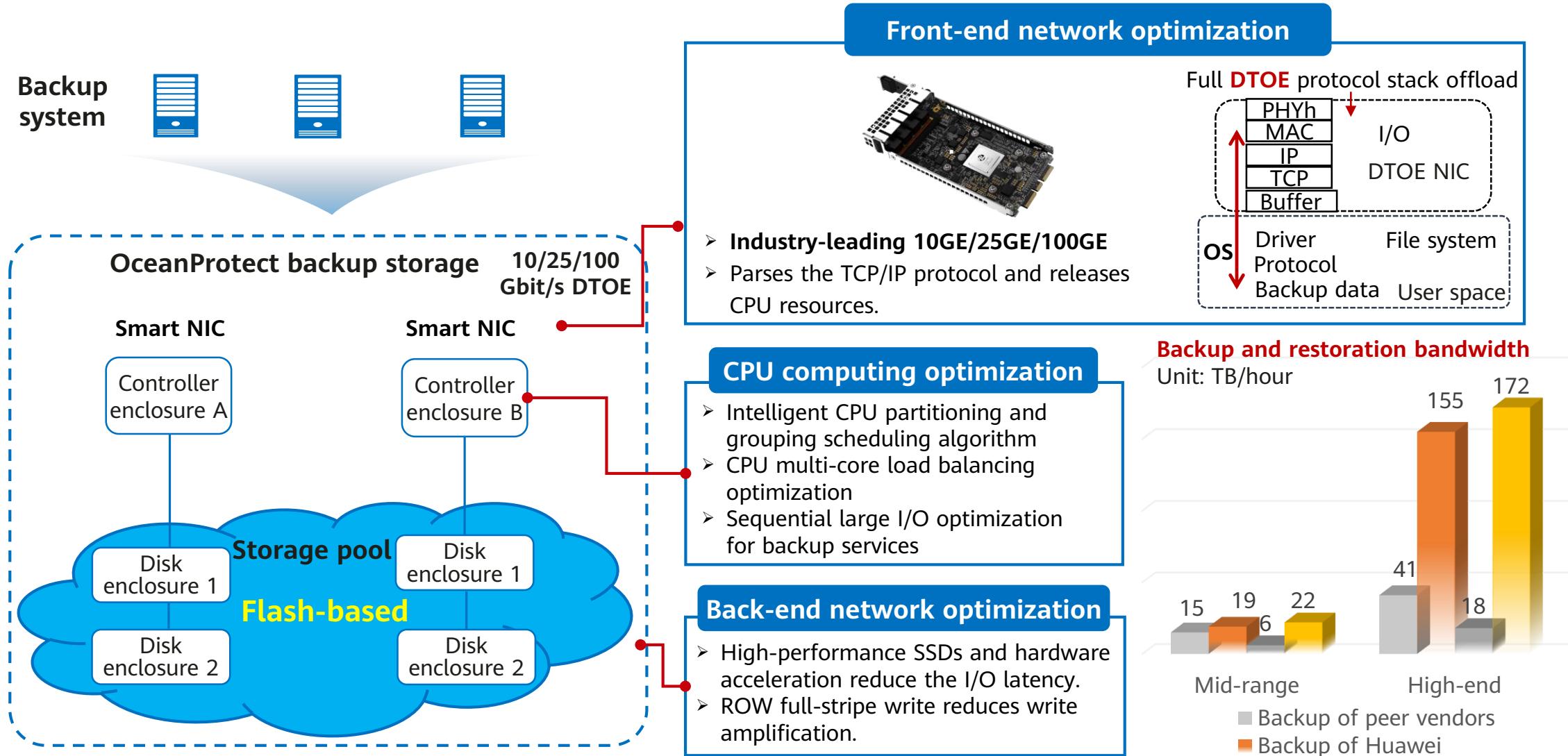
Software Architecture



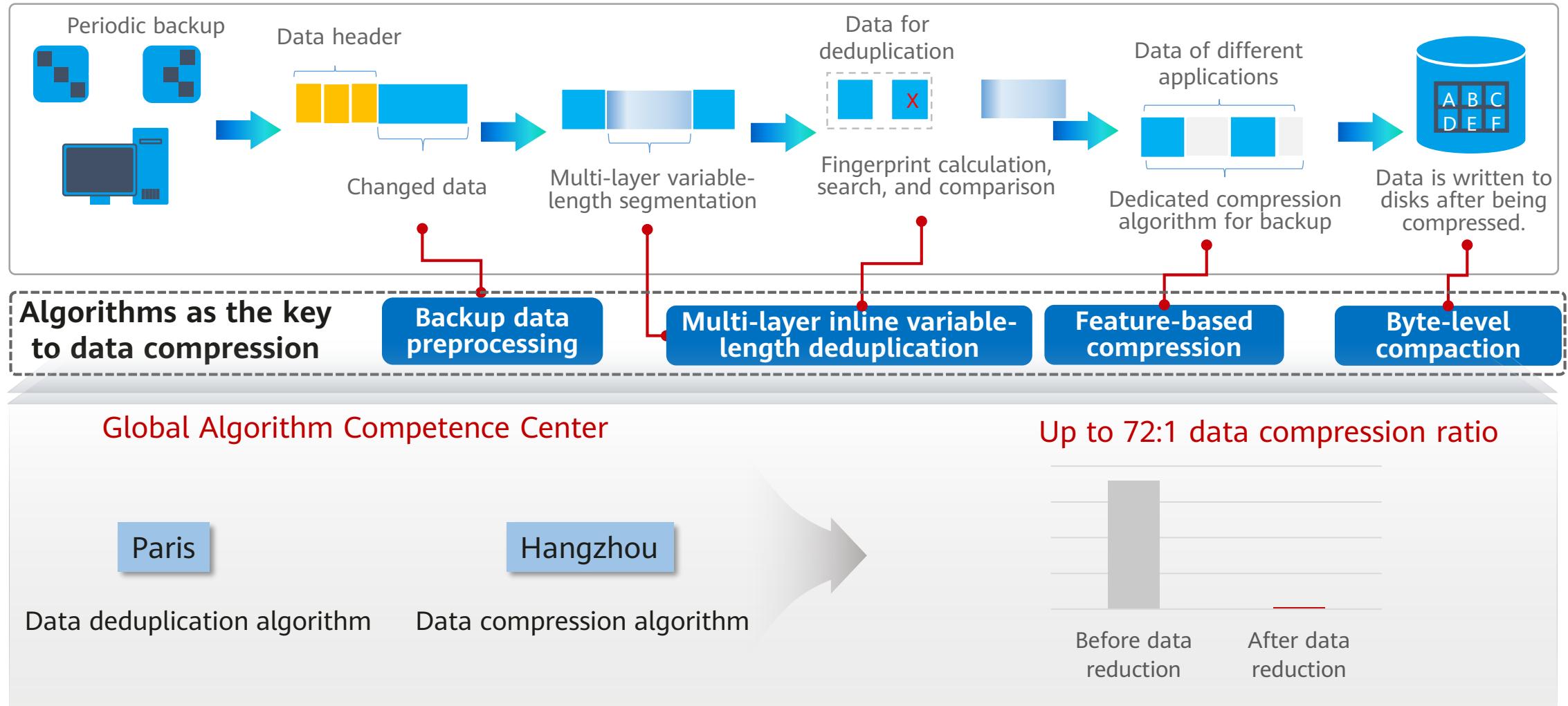
All-Scenario Data Protection for the Intelligent World



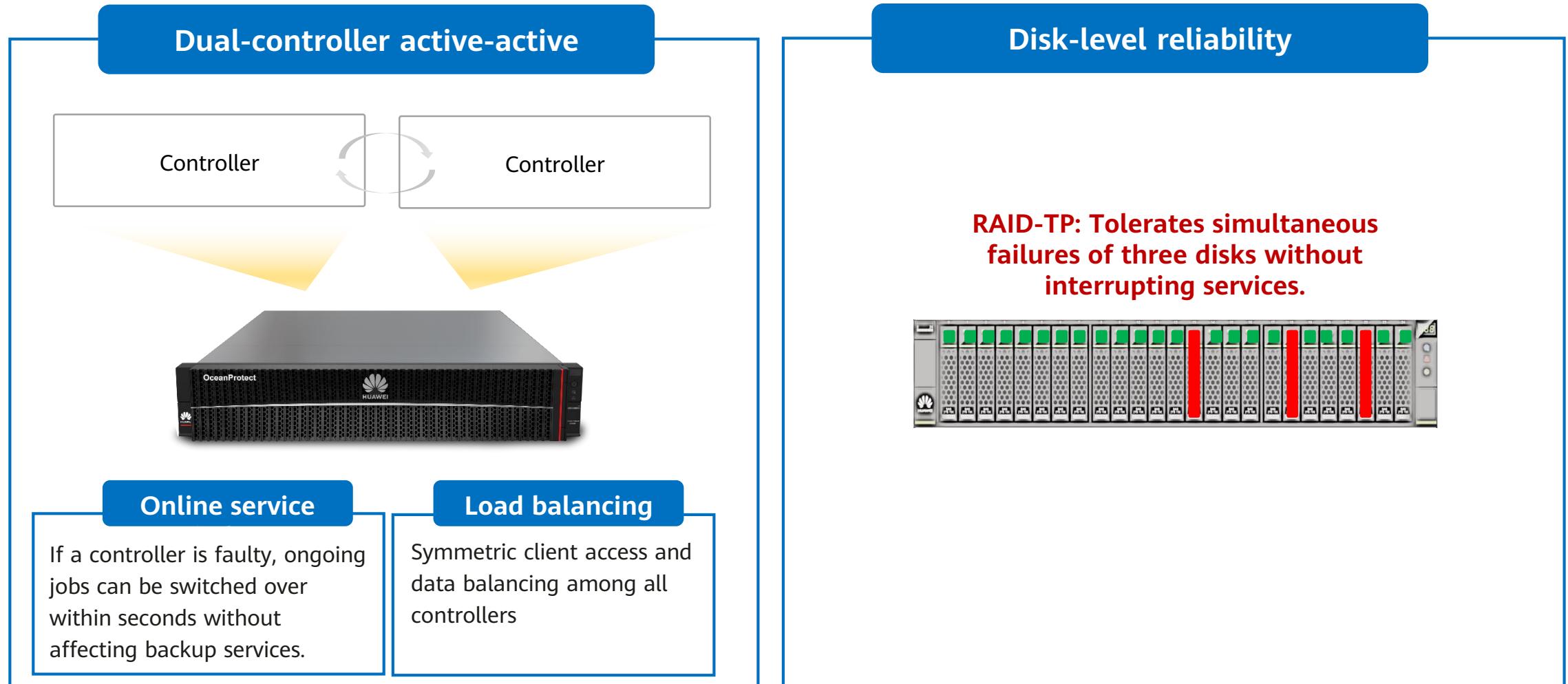
E2E Backup Acceleration Improves System Performance



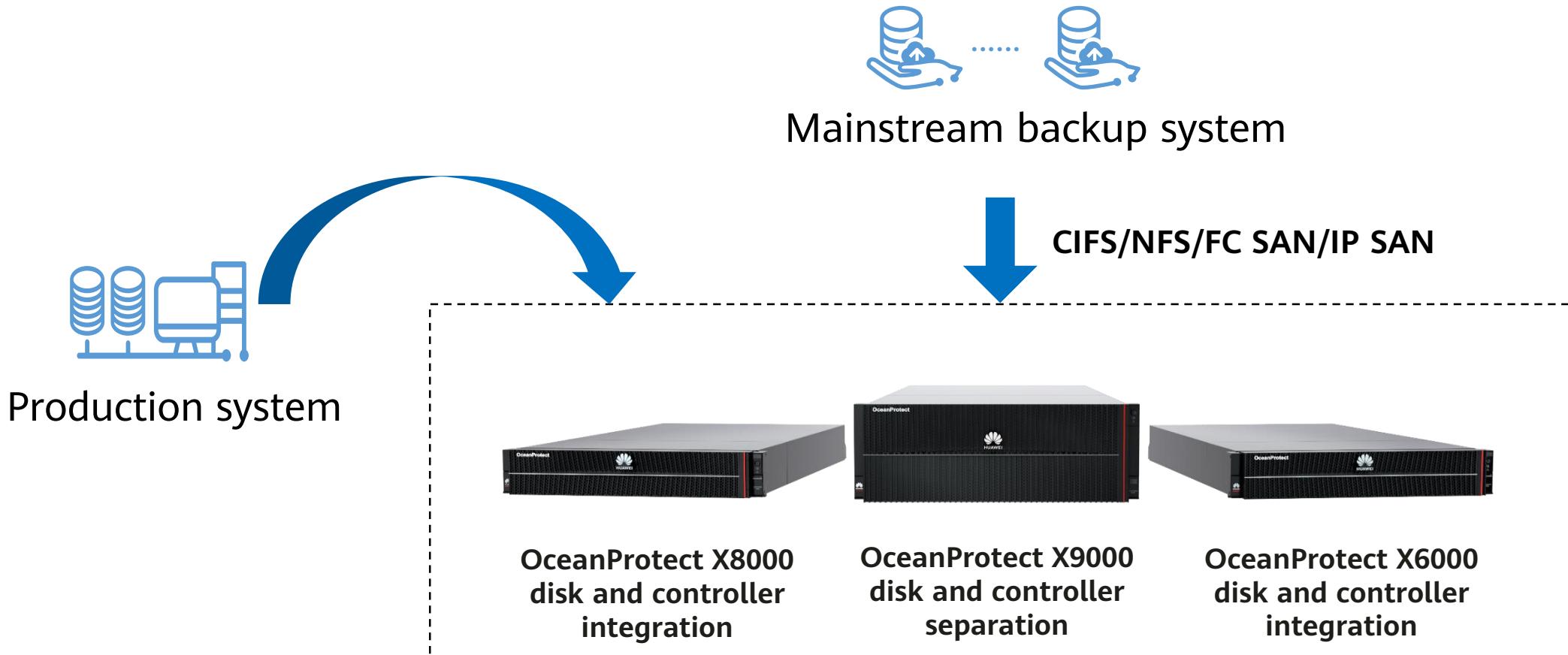
Throughout-Data-Flow Deduplication and Compression Implement Ultimate Data Reduction



The Active-Active Architecture and RAID-TP Ensure System-level Reliability of Services and Data



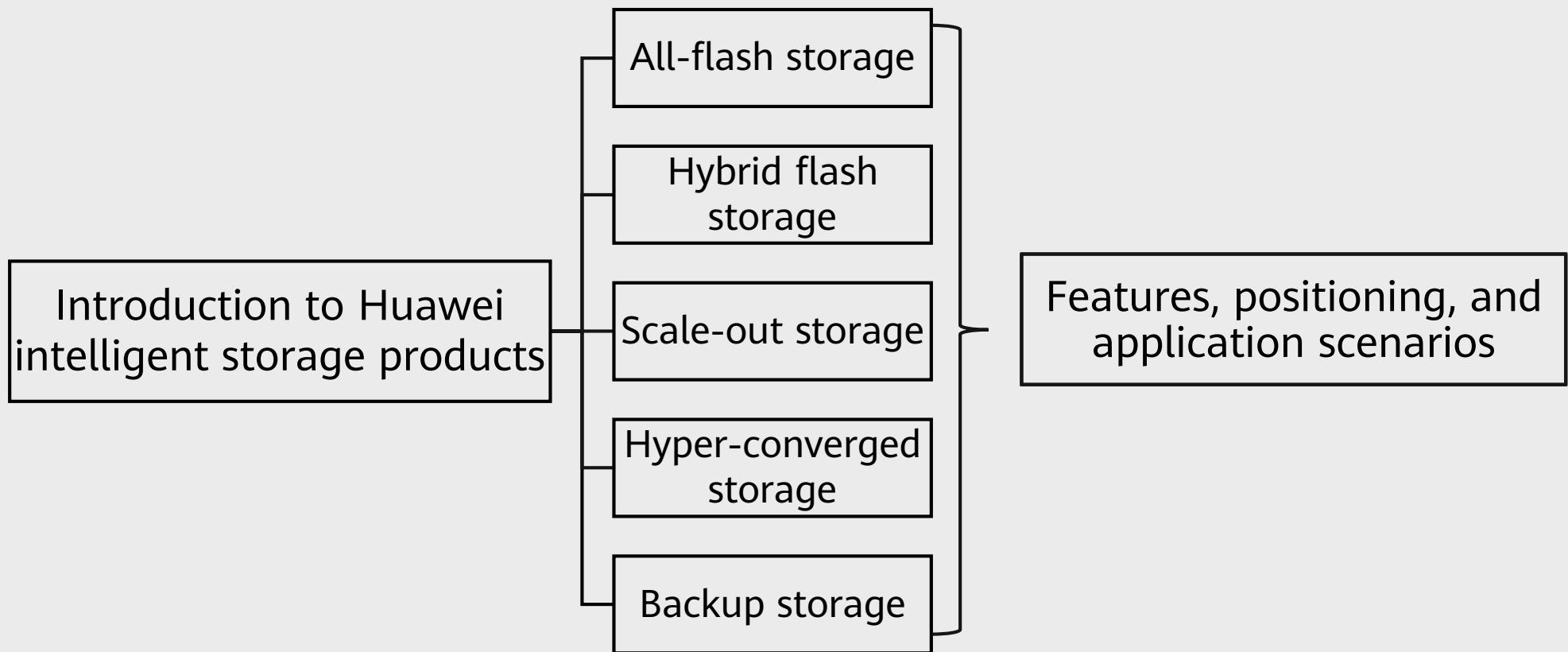
Application Scenario: The Standard NAS/SAN Protocol Is Compatible with the Backup Software Ecosystem



Quiz

1. (Single-answer question) Which of the following statements about Huawei OceanStor all-flash storage is incorrect?
 - A. Supports SSDs and NVMe SSDs.
 - B. Supports SAS disks.
 - C. Supports HDDs.
 - D. Supports palm-sized SSDs.
2. (Multiple-answer question) Which of the following storage services are supported by Huawei scale-out storage?
 - A. Block storage
 - B. File storage
 - C. Object storage
 - D. HDFS storage
 - E. Linked storage

Summary



Recommendations

- Huawei official websites
 - Enterprise service: <https://e.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/en/index.html>
 - Online learning: <https://learning.huawei.com/en/>
- Popular tools
 - HedEx Lite
 - Network Documentation Tool Center
 - Information Query Assistant

Acronyms and Abbreviations

DME: Data Management Engine

HPDA: High Performance Data Analytics

RAID: Redundant Array of Independent Disks. It is a technology that provides a disk group (logical disks) consisting of multiple disks (physical disks) combined in different modes. The disk group features higher storage performance over a single disk and supports data redundancy.

ROW: Redirect on write. A core technology used to implement file system snapshots. When the source file system receives a data write request and data in the source file system needs to be modified, the storage system specifies a new storage location in the storage pool for the new data and directs the modified data block to the new storage location.

OLTP: Online transaction processing

OLAP: Online analytical processing

NDMP: Network Data Management Protocol. It is an open protocol for network-based backup of file service systems that allows platform-independent data storage.

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.



Storage System Operation Management



Foreword

- This course describes three methods of managing storage systems: OceanStor DeviceManager, common line interface (CLI), and UltraPath, as well as management content and related operations.

Objectives

Upon completion of this course, you will be able to know:

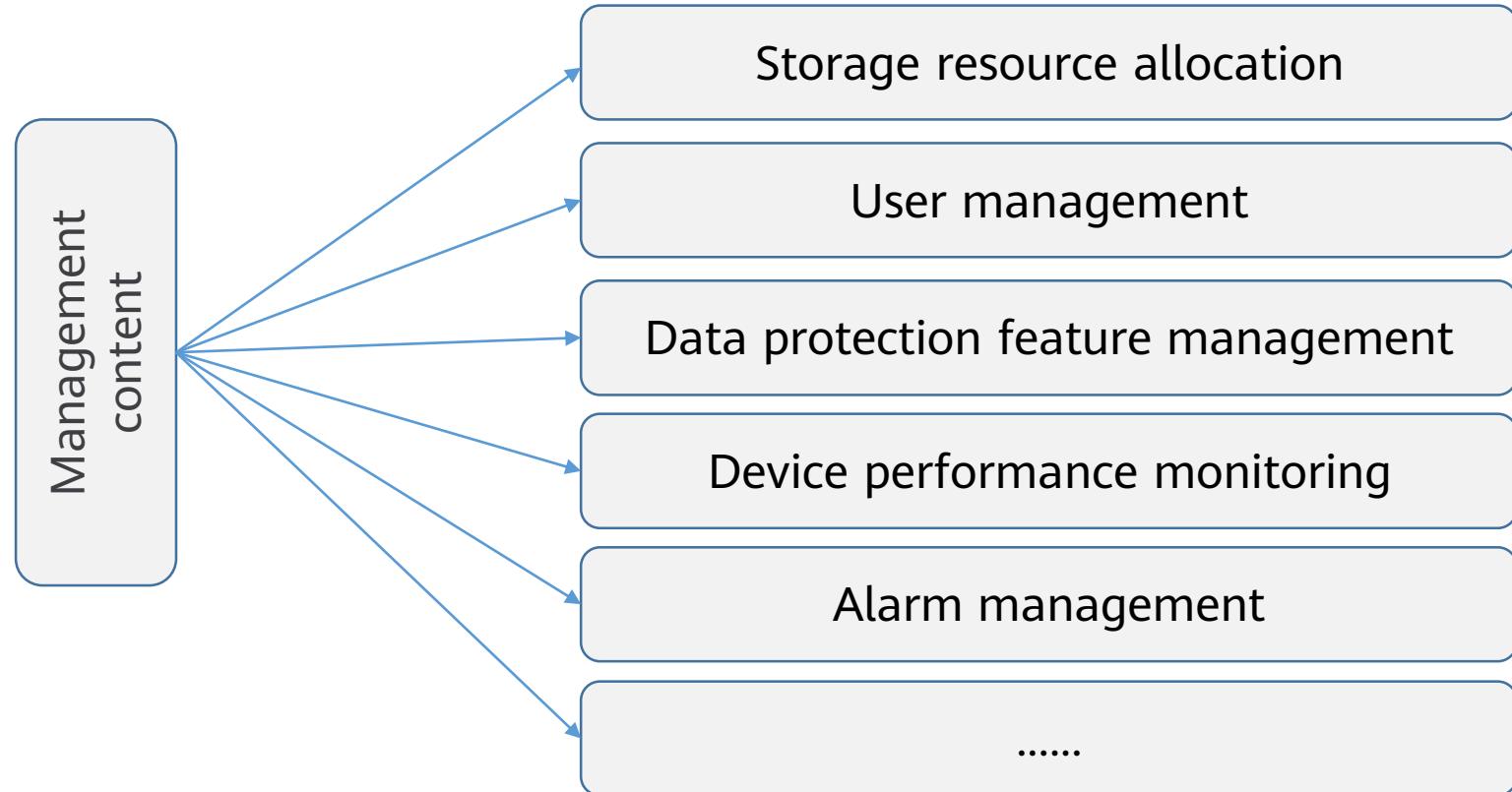
- DeviceManager, CLI, and UltraPath storage system management tools
- Basic management operations of the storage system

Contents

- 1. Storage Management Overview**
2. Storage Management Tools
3. Basic Management Operations

Storage Management Definition

- Storage management allows users to use management tools to query, set, manage, and maintain storage systems.



Common Storage System Access Mode



In what ways can I **log in to a storage system?**

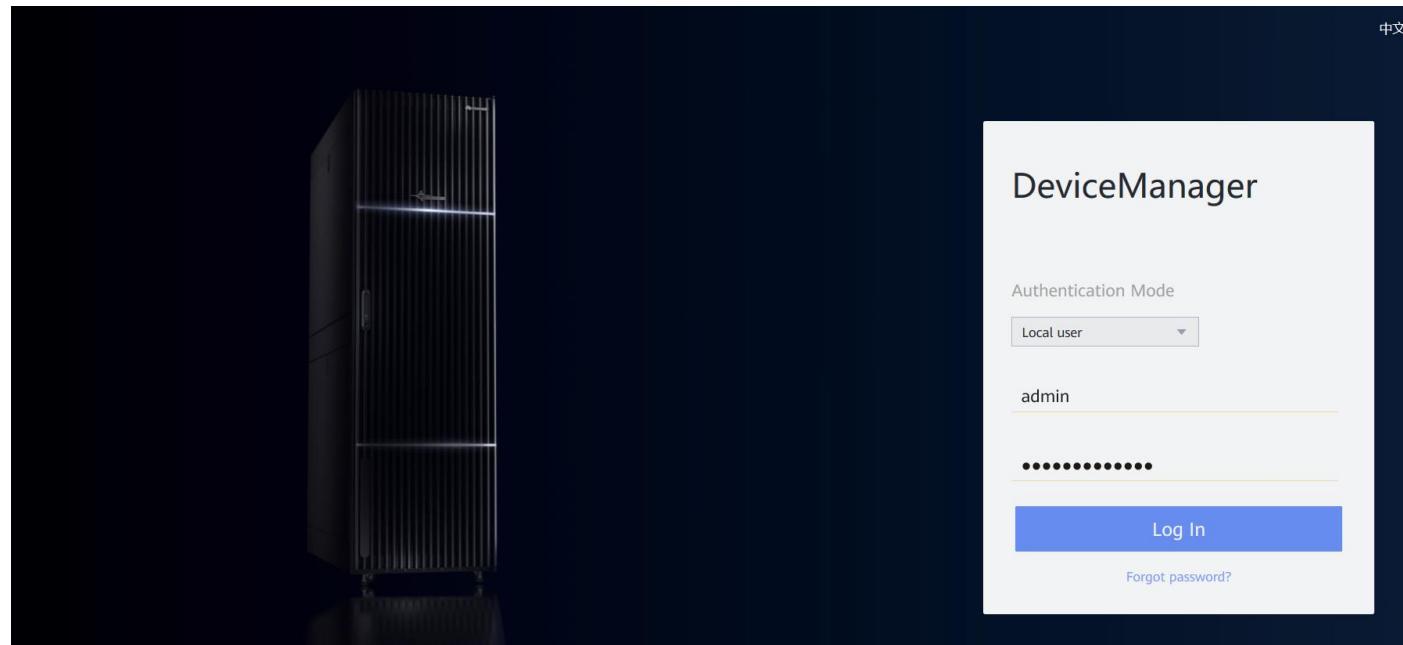
Log in to DeviceManager.

Log in to the CLI.



Introduction to DeviceManager

- DeviceManager is storage management software designed by Huawei for a single storage system. It can help you easily configure, manage, and maintain storage devices.
- Main software functions include storage resource allocation, user management, data protection feature management, device performance monitoring, and alarm management.



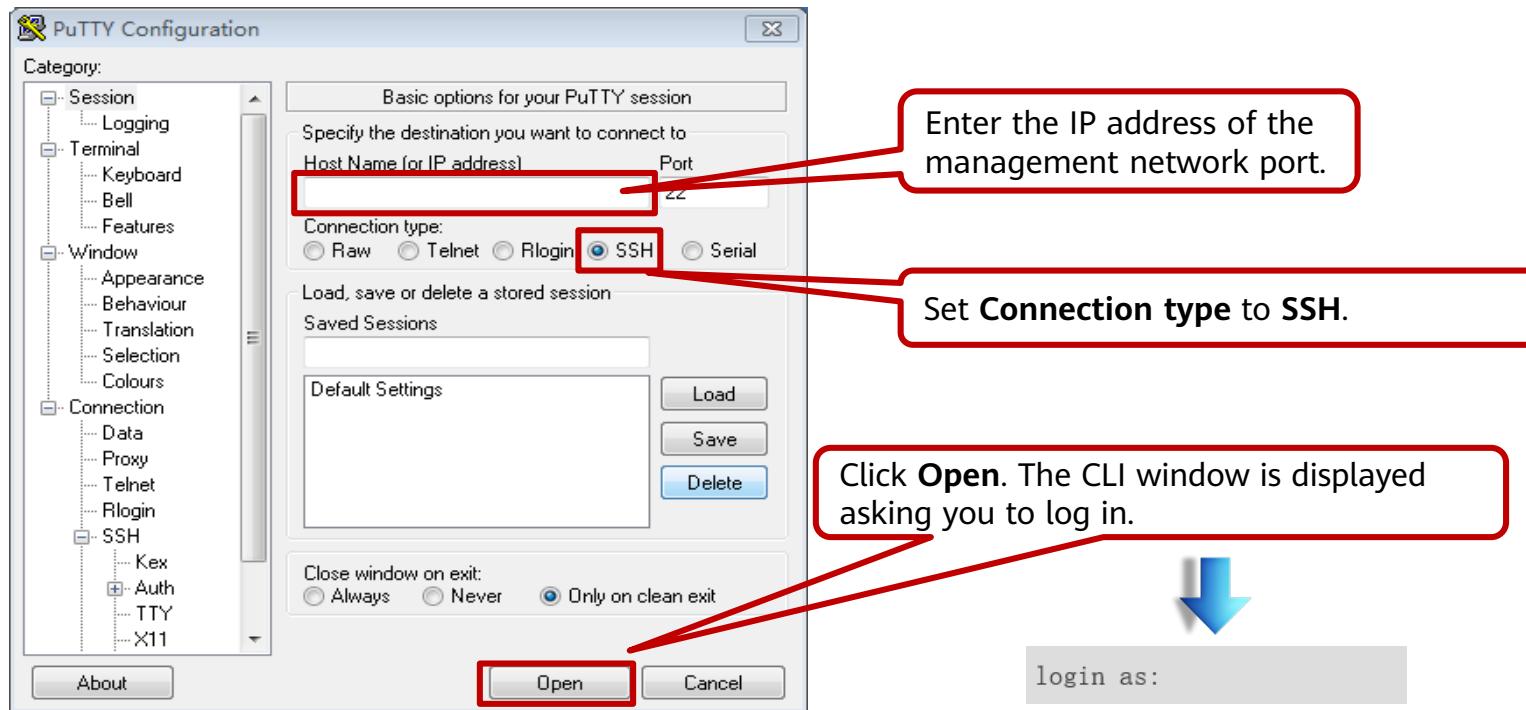
Logging In Using DeviceManager



You must add port number **8088** after the IP address of the management network port. Otherwise, the login fails.
Format: **https://xxx.xxx.xxx.xxx:8088**

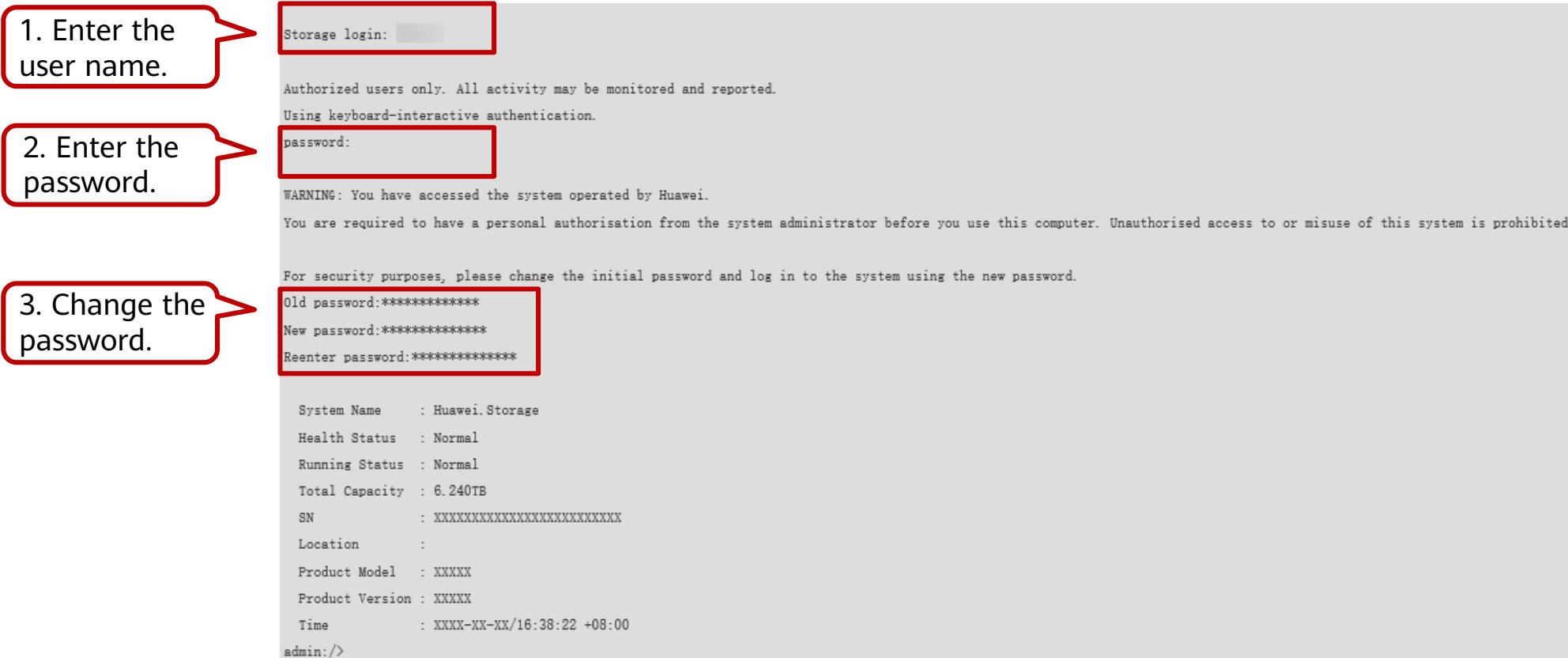
Introduction to the CLI

- CLI allows you to manage and maintain the storage system. Configuration commands are entered on the keyboard and compiled and executed by programs. The command output is displayed in text or graphic format on the CLI.
- Terminal software is required for logging in to the CLI. PuTTY is used as an example.



Logging In Using the CLI

- Enter the user name and password as prompted. The system asks you to change the password upon the first login. Change the password immediately to ensure system security. The following information is displayed when the login is successful:



The diagram illustrates the three steps of logging in via the CLI:

1. Enter the user name. (Red callout points to the 'Storage login:' prompt.)
2. Enter the password. (Red callout points to the 'password:' prompt.)
3. Change the password. (Red callout points to the password change prompts: 'Old password:', 'New password:', and 'Reenter password:').

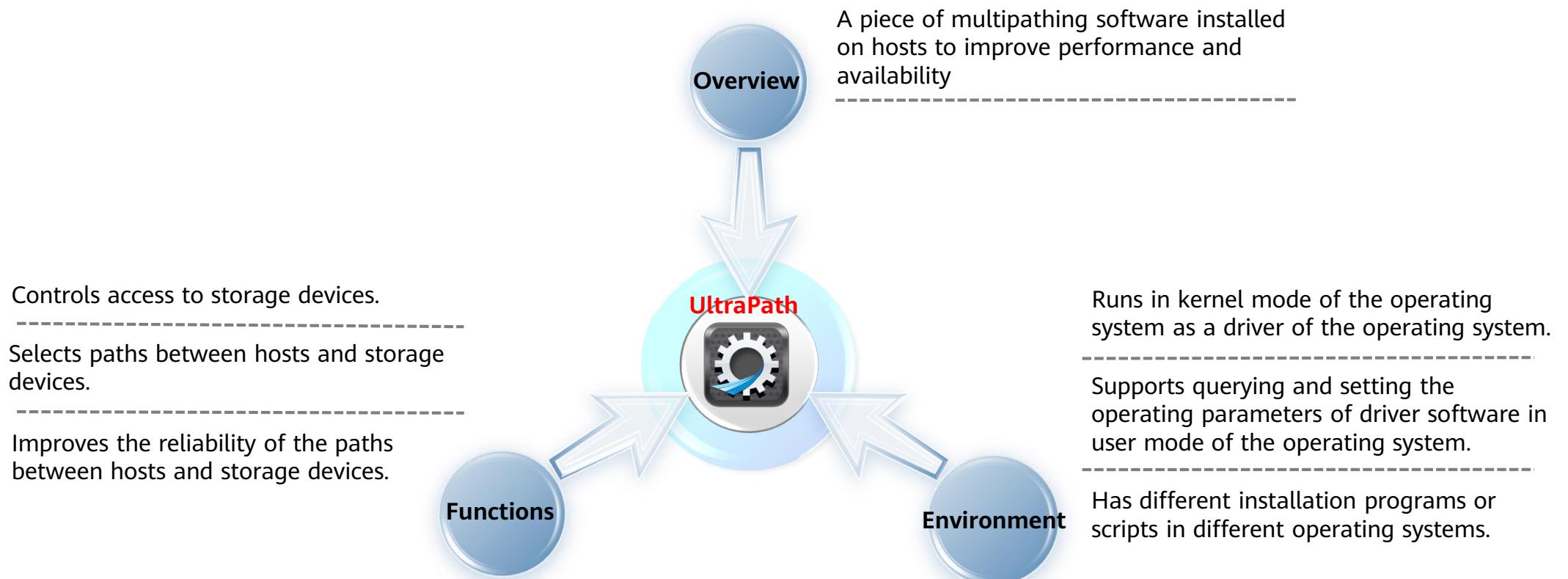
```
Storage login: [REDACTED]
Authorized users only. All activity may be monitored and reported.
Using keyboard-interactive authentication.
password: [REDACTED]
WARNING: You have accessed the system operated by Huawei.
You are required to have a personal authorisation from the system administrator before you use this computer. Unauthorised access to or misuse of this system is prohibited.

For security purposes, please change the initial password and log in to the system using the new password.
Old password:*****
New password:*****
Reenter password:*****
```

```
System Name      : Huawei.Storage
Health Status    : Normal
Running Status   : Normal
Total Capacity   : 6.240TB
SN               : XXXXXXXXXXXXXXXXXXXXXXXXX
Location        :
Product Model   : XXXXX
Product Version : XXXXX
Time             : XXXX-XX-XX/16:38:22 +08:00
admin:/>
```

Introduction to UltraPath

- OceanStor UltraPath is the multipathing software developed by Huawei. Its functions include masking of redundant LUNs, optimum path selection, I/O load balancing, and failover and fallback. These functions enable your storage network to be intelligent, stable, and fast.



Contents

1. Storage Management Overview
- 2. Storage Management Tools**
3. Basic Management Operations

DeviceManager GUI (1)

The screenshot shows the DeviceManager GUI Home page. At the top, there is a navigation bar with tabs: Home (selected), Services, Data Protection, Insight, System, and Settings. On the far right, there are search, filter, and user-related icons.

Alarms: A circular gauge indicates 6 total alarms, with 2 Critical, 3 Major, and 1 Warning. To the left, a device summary for an OceanStor Dorado 6000 V6 is shown, including its model, version (6.1.5RC1), and ESN.

Common Operations: Buttons for Create LUN Group, Create Host, Create File System, and Create NFS Share.

Capacity: Shows a reduction ratio of 1.003 : 1, with a total of 2.073 TB, used 1.494 GB, and free 2.072 TB. It also displays pre-savings and post-savings data.

Data Reduction: A chart showing the reduction ratio across different categories: All, Block, and File.

Capacity Trend: A trend graph with a windmill icon, showing historical monitoring data retention.

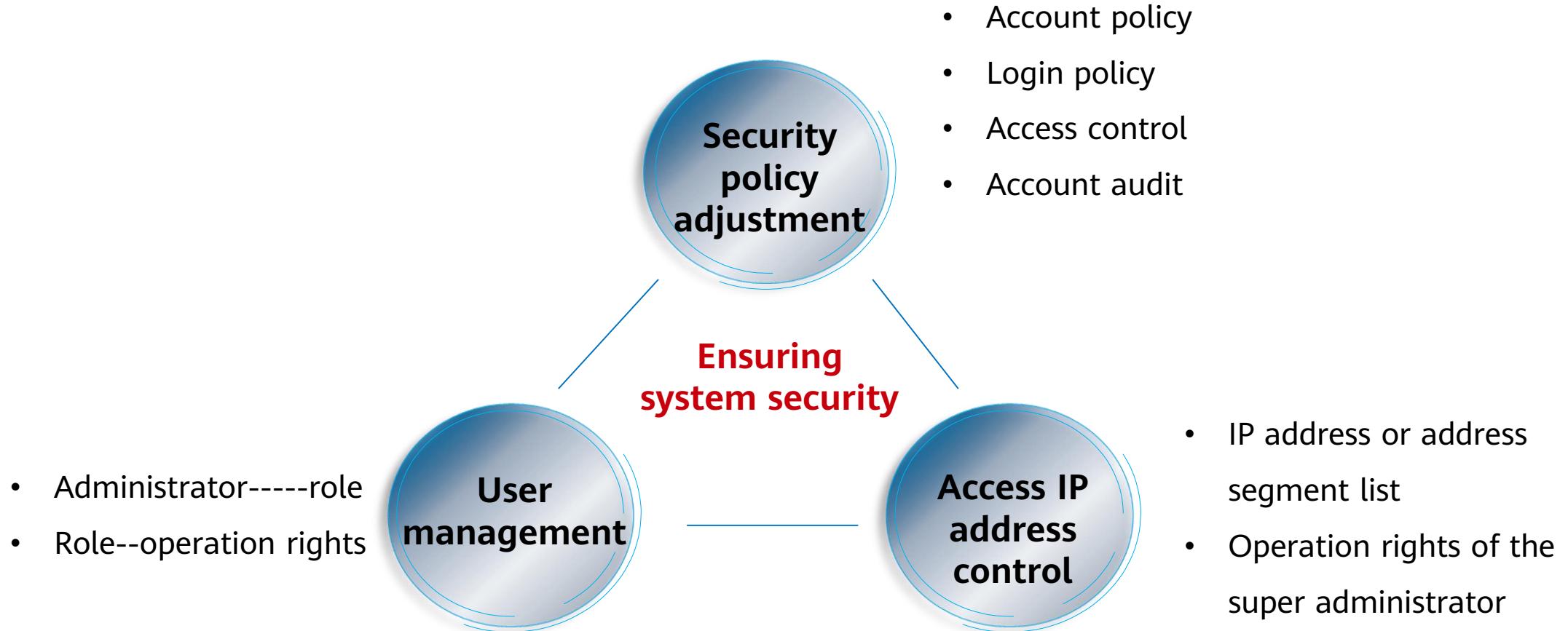
Network: A network diagram showing Physical Ports (8 ETH), Logical Ports (ISCSI, NFS, CIFS), and Resources (LUNs, LUN Groups, Hosts, Host Groups, File Systems).

Note: The GUI may vary slightly depending on the product version and model. The actual GUI prevails.

DeviceManager GUI (2)

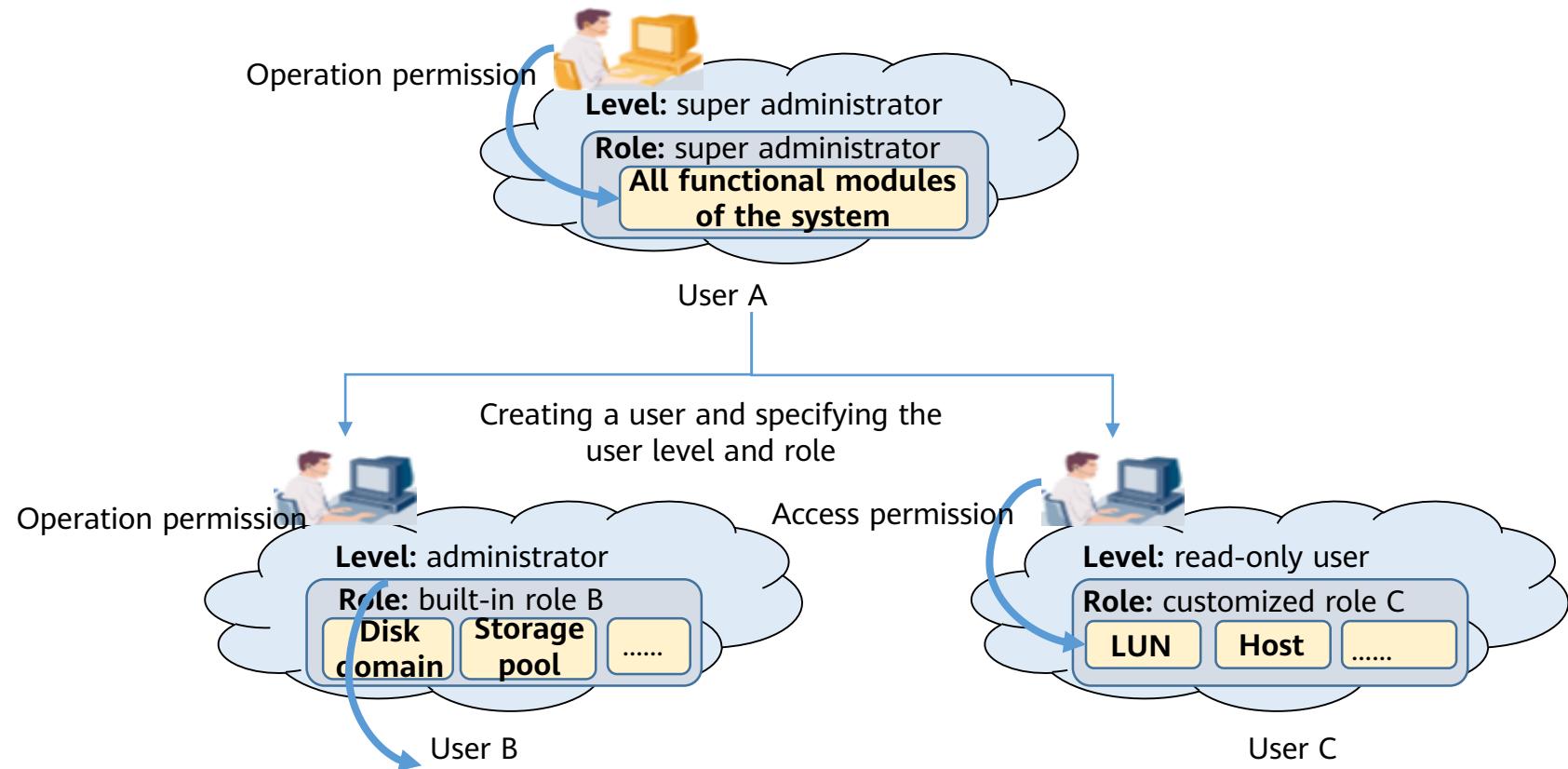
No.	Name	Description
1	Function pane	Displays available functions related to the current operation.
2	Navigation bar	Lists all functional modules of the storage system.
3	Alarm and task statistics area	The alarm statistics area displays the number of alarms by severity and helps you learn about the running status of the system. The task statistics area displays all the tasks executed by users. You can check whether the tasks are executed successfully.
4	Device management area	In the device management area, you can view and modify device information, and power off or restart devices.
5	Logout and language area	The logout and language area provide buttons of logout and language. DeviceManager supports two languages: English and simplified Chinese.

Managing the Access Permission of a Storage System



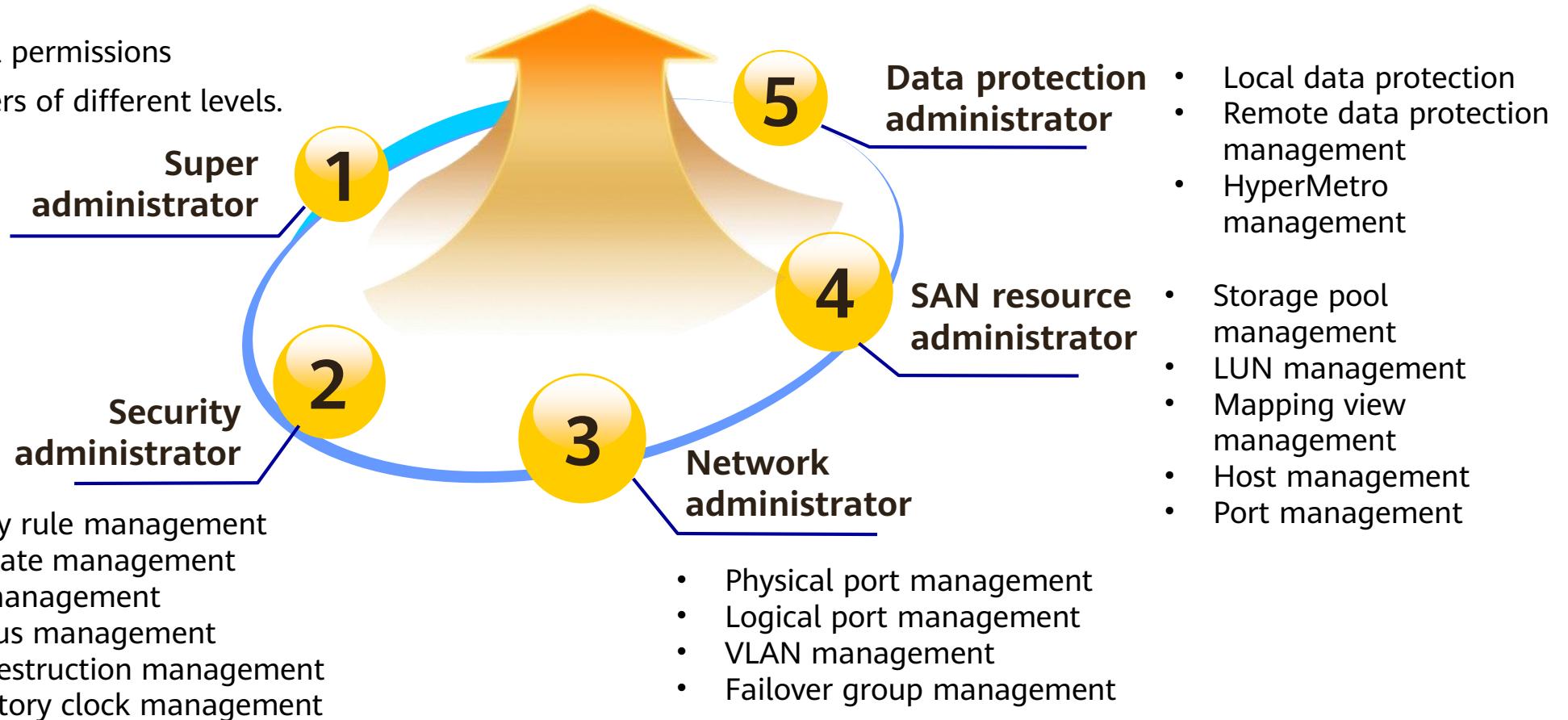
Storage System User Management

- To prevent misoperations from compromising the storage system stability and service data security, the storage system defines user levels and roles to determine user permission and scope of permission.

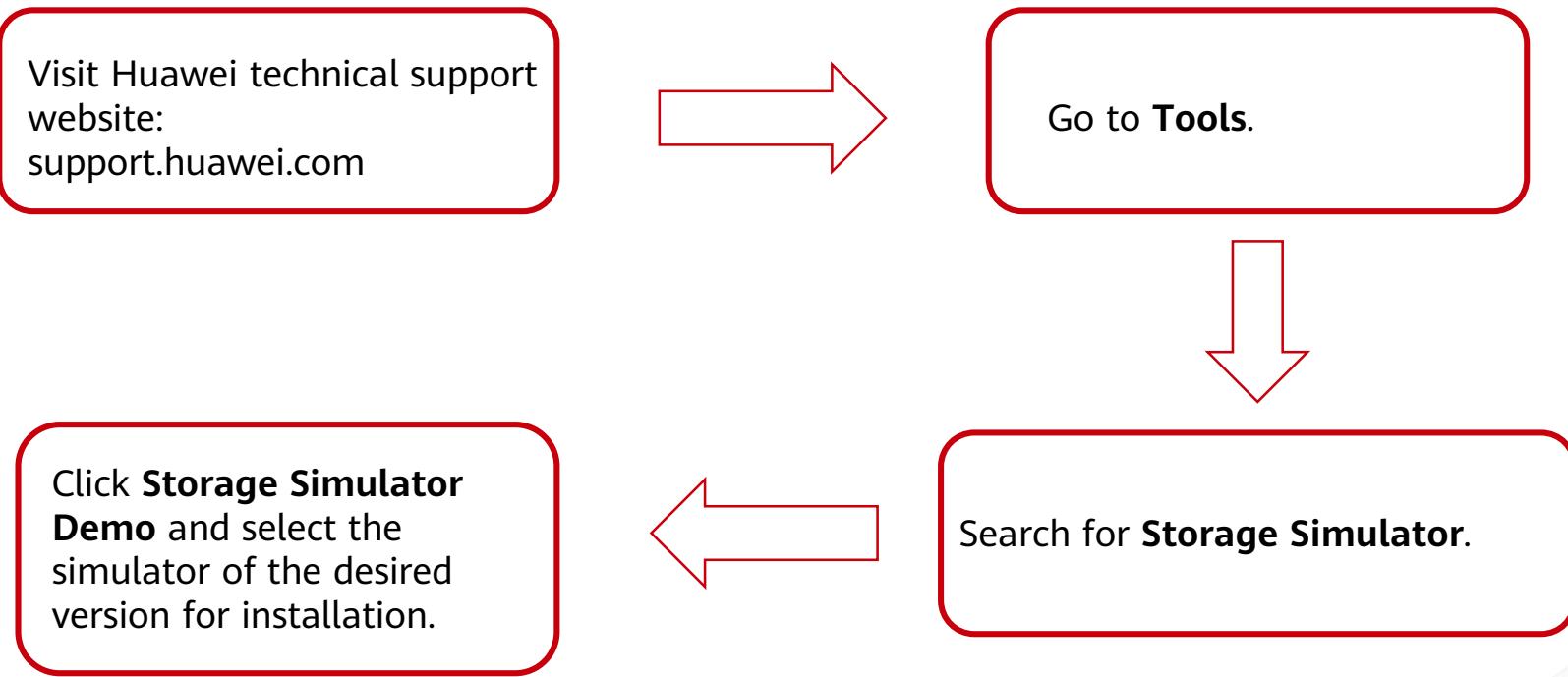


Roles and Permissions of a User

- Full control permissions
- Creates users of different levels.



Downloading a DeviceManager Demo



CLI Format Conventions (1)

- You are required to follow the format conventions when you use the CLI commands. Typical command formats are:

```
change storage_pool relocation_speed relocation_speed=?
```



1 2 3

- First field: operation that you want to perform, for example, change (modify) and show (query)
- Second field: object of an operation, for example, storage_pool (storage pool) and host (host)
- Third field (available only in some commands): object attribute, for example, relocation_speed (migration rate)
- Other fields: other parameters required

CLI Format Conventions (2)

- For example, `change user user_name=? { level=? | action=? }`
 - `change user` keeps unchanged.
 - `user_name=?`, mandatory; For `level=?` and `action=?`, one of them can be selected.
 - For parameter `level=?`, `level=` remains unchanged. The value of `?` must be an optional value, for example, `level=admin`.
- Correct command example: `admin:/>change user user_name=newuser level=admin`

Format	Description
Boldface	The keywords of a command are in boldface .
<i>Italics</i>	The arguments of a command line, which will be replaced by actual values, are in <i>italics</i> .
[]	Items in square brackets ([]) are optional.
{ x y ... }	Optional items are grouped in braces ({ }) and separated by vertical bars (). One item must be selected.
[x y ...]	Optional items are grouped in square brackets ([]) and separated by vertical bars (). Only one item or no item can be selected.
{ x y ... } *	Optional items are grouped in braces ({ }) and separated by vertical bars (). At least one item must be selected, and at most all items can be selected.
[x y ...] *	Optional items are grouped in square brackets ([]) and separated by vertical bars (). Several items or no item can be selected.

CLI Command Completion

- On the CLI, you can press **Tab** or the space bar to use the command completion function.
- The difference between the two keys is as follows: The space key is used to supplement only the current field, whereas the **Tab** key is used to supplement all possible values.

Press **Tab** once to display the available starting segments of a command line.

```
admin:/>/Press "Tab"  
^  
add      change     create  
delete   exit       export  
help     import    poweroff  
poweron  reboot    remove  
scan     show      swap
```

After the starting segment is determined and completed, press **Tab** once to display the available adjacent segments of the starting segment.

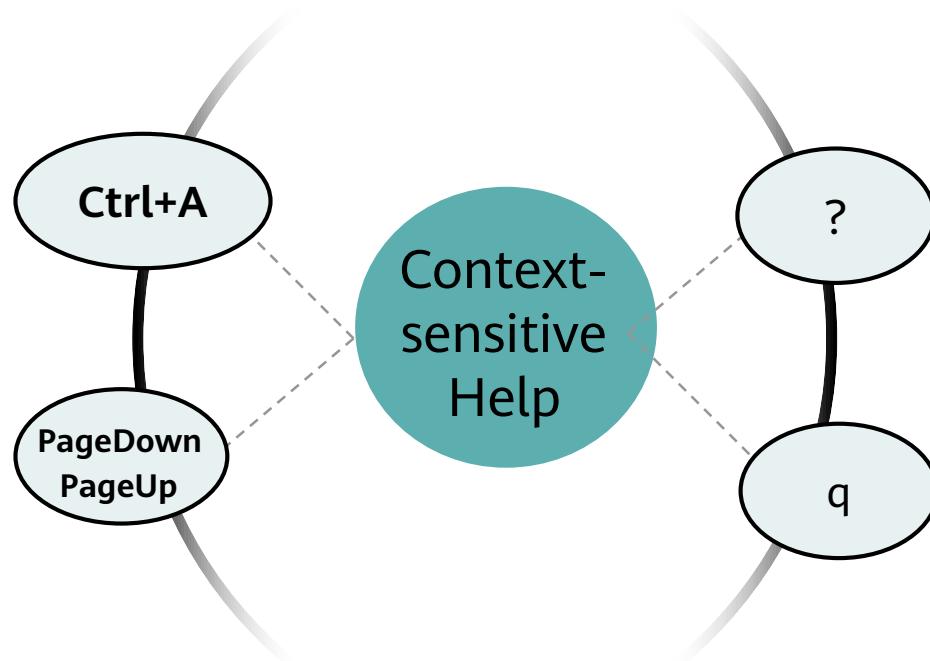
```
admin:/>add//Press "Tab"  
cache_partition  clone      consistency_group  
disk_domain     host      host_group  
lun_copy        lun_group mapping_view  
notification    port      port_group  
remote_device   security_rule smartqos_policy  
snmp           storage_pool
```

When all the fields required by the command are entered and the conditions for running the command are met, the system prompts that the command can be run after you press **Tab**. In this case, you can press **Enter** to run the command.

```
admin:/>add port ipv4_route eth_port_id=0 type=net target_ip=192.168.3.0 mask=255.255.255.0 gateway=10.0.0.1//Press Tab  
Command is executable now.
```

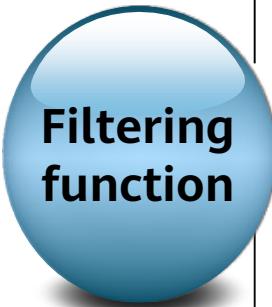
Context-Sensitive Help

- Press **Ctrl+A** to view the optional values of certain parameters in certain commands. Generally, these values need to be obtained from the system.
- You can turn pages on the context-sensitive help page.



- Enter a question mark (?) to query the basic instruction of CLI operations and detailed description of command parameters.
- After entering the first field of the command and a space, enter a question mark (?). You can query all available next fields and the detailed description of each field.
- Exit the context-sensitive help page.

CLI Command Filtering



Purpose

Redundant information is deleted, and valid content is displayed as required.

How to Use

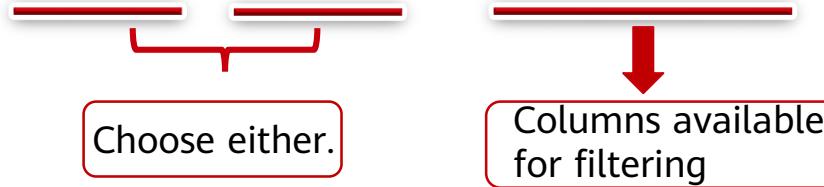
After entering the complete query command, enter | and press **Tab** or the space bar.

Related Commands

- filterColumn column filtering command
- filterRow row filtering command

CLI Column Filtering Command - filterColumn

```
show xxx|filterColumn { exclude | include } columnList=?
```



exclude: Filter out information that does not need to be displayed.

Include: Only the columns to be displayed are reserved.

If multiple columns are involved, they are separated by commas (,).

```
admin:/>show bbu general|filterColumn exclude//Press "Tab"  
<columnList=?>    column list separated by comma, select one or more  
                      separated by comma, the spaces are replaced with \s in the  
                      parameter list.  
columnList=Inter\sID  
columnList=Health\sStatus  
columnList=Current\sVoltage(V)  
columnList=Firmware\sVersion  
columnList=Owning\sController  
columnList=ID  
columnList=Running\sStatus  
columnList=Number\sOf\sDischarges  
columnList=Delivered\sOn  
columnList=Electronic\sLabel
```

```
admin:/>show bbu general |filterColumn include  
columnList=Inter\slID, ID
```

Inter	ID
0.0A.0	CTE0.0
0.0A.1	CTE0.1

CLI Row Filtering Command - filterRow

```
show xxx|filterRow column=? predict=? [ predict2=? ] value=? [ logicOp=? ]
```



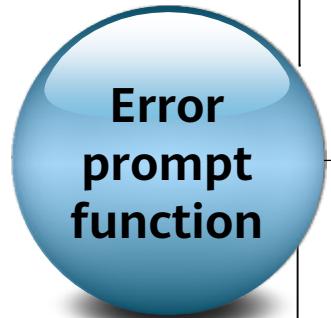
predict=?

- not: The **logicOp** is not.
- equal_to: a value equal to **value=?**
- greater_than: a value greater than **value=?**
- greater_equal: a value equal to or greater than **value=?**
- less_than: a value less than **value=?**
- less_equal: a value less than or equal to **value=?**
- match: regular expression matching **value=?**

logicOp=?

- and: Multiple columns that meet the condition are displayed.
- or: Any column that meets the condition is displayed.

Error Prompt Function



Purpose

Specify the position of the input error in the command and provide the correct field for reference.

How to Use

When the format of the entered command is incorrect, the system displays the error location with symbol ^.

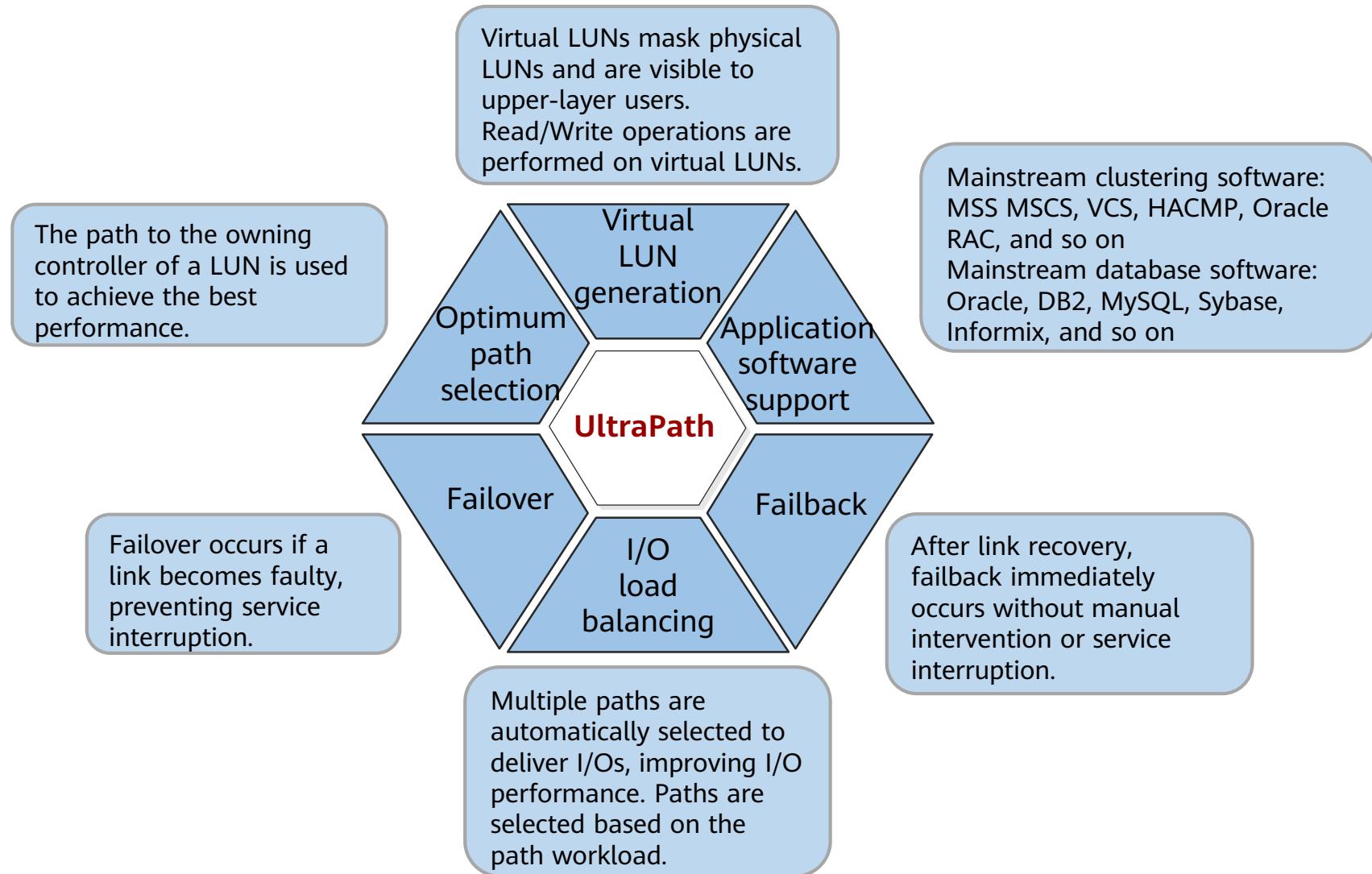
Note

When multiple errors occur in the command, the system displays only the first error.

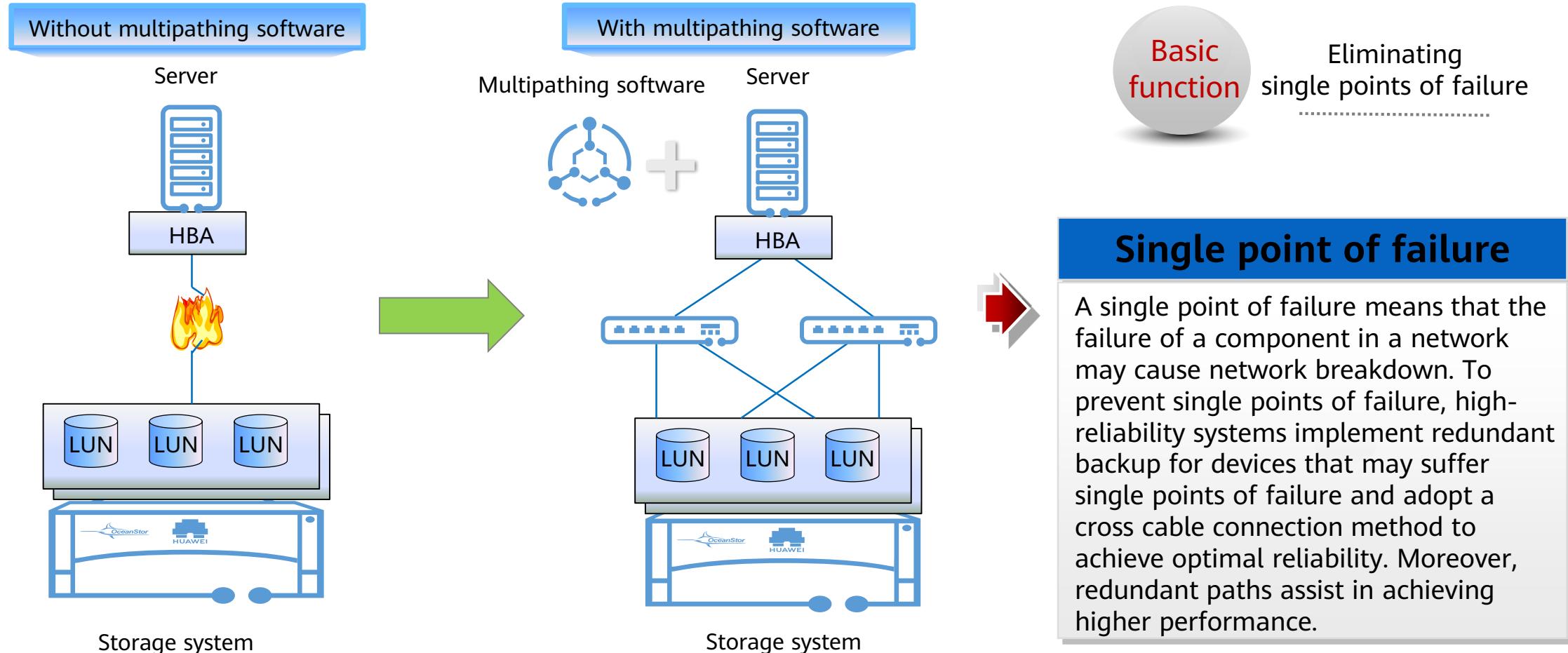
```
admin:/>add part
          ^
port      port_group

admin:/>add part
```

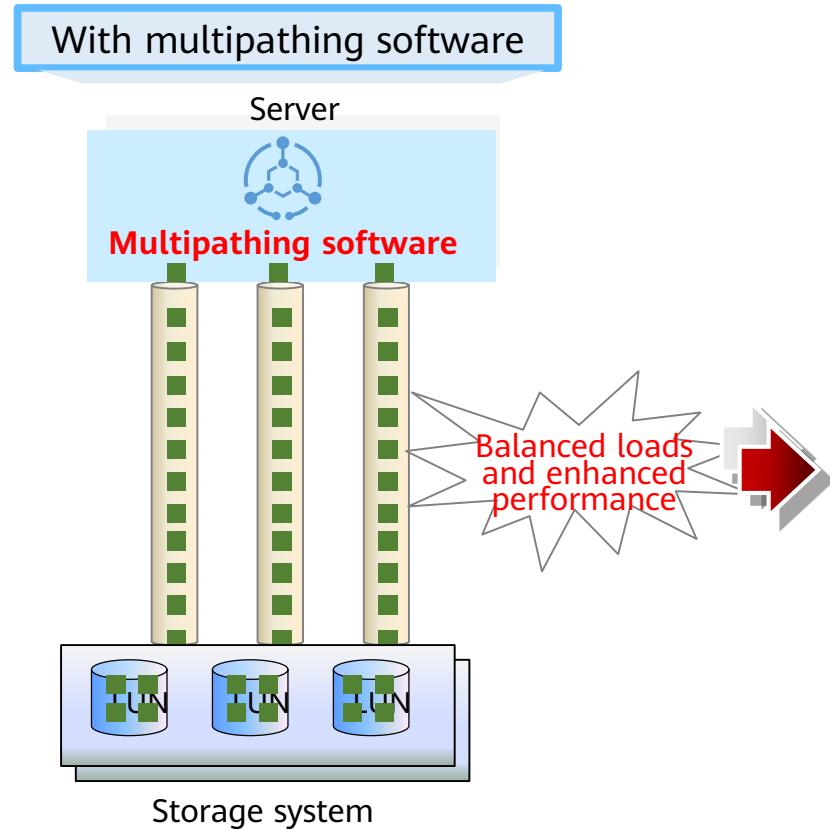
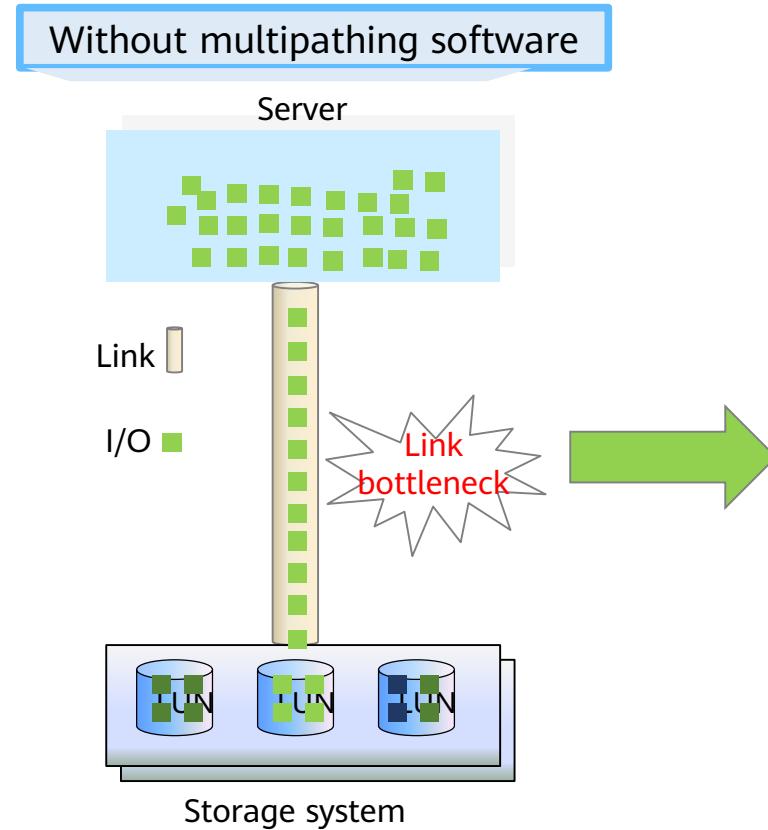
Major Functions of UltraPath



Positioning of Multipathing Software



Positioning of Multipathing Software



Basic function

Load balancing

Load balancing

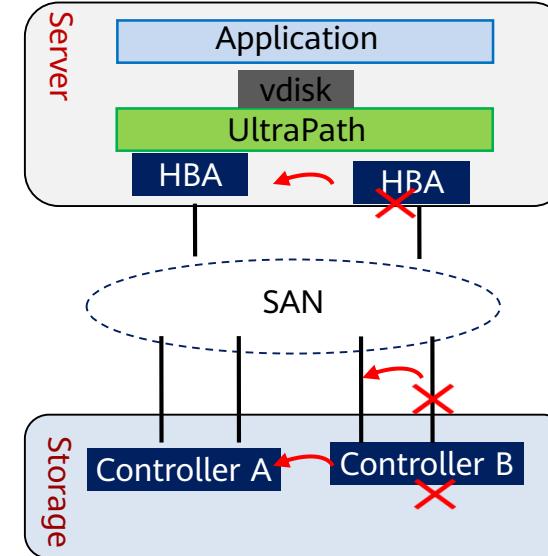
Load balancing is another critical function of multipathing software. With load balancing, the system uses the bandwidth of multiple links to improve overall throughput. Common load balancing algorithms include round-robin, minimum queue depth, and minimum task.

Positioning of Multipathing Software

Positioning

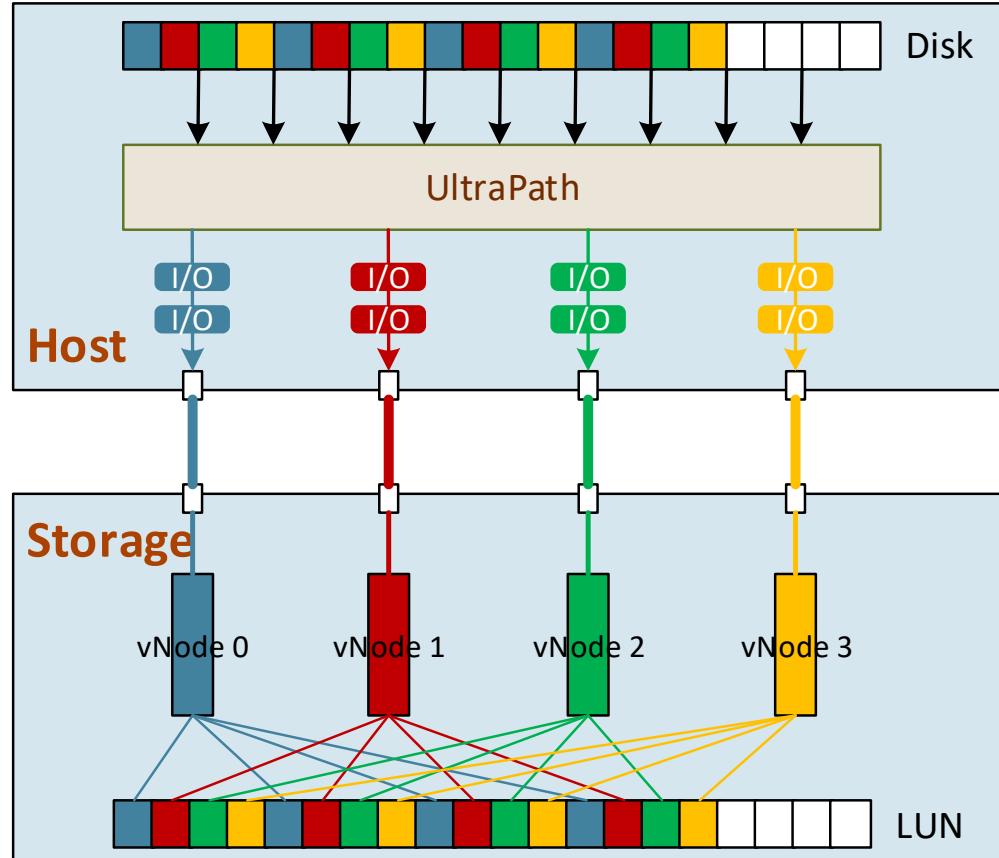
UltraPath is a type of filter driver software running in the host kernel. It can block and process disk creation/deletion and I/O delivery of operating systems.

UltraPath implements a reliable use of redundant paths. If a path fails or cannot meet the performance requirement, UltraPath automatically and transparently transfers I/Os to other available paths to ensure that I/Os are transmitted effectively and reliably. As shown in the figure on the right, UltraPath can handle many faults such as HBA faults, link faults, and controller faults.



Basic Function	Severity	Description
Failover	High	If a path is faulty, I/Os on the path are automatically transferred to another available path.
Fallback	High	After the faulty path recovers, I/Os are automatically transferred back to the path.
Load balancing	High	The bandwidths of multiple links are used, improving the overall system throughput.

Active-Active Architecture with Full Load Balancing in OceanStor V6



Even distribution of unhomed LUNs

Data on LUNs is divided into 64 MB slices. The slices are distributed to different virtual nodes based on the hash result (LUN ID + LBA).

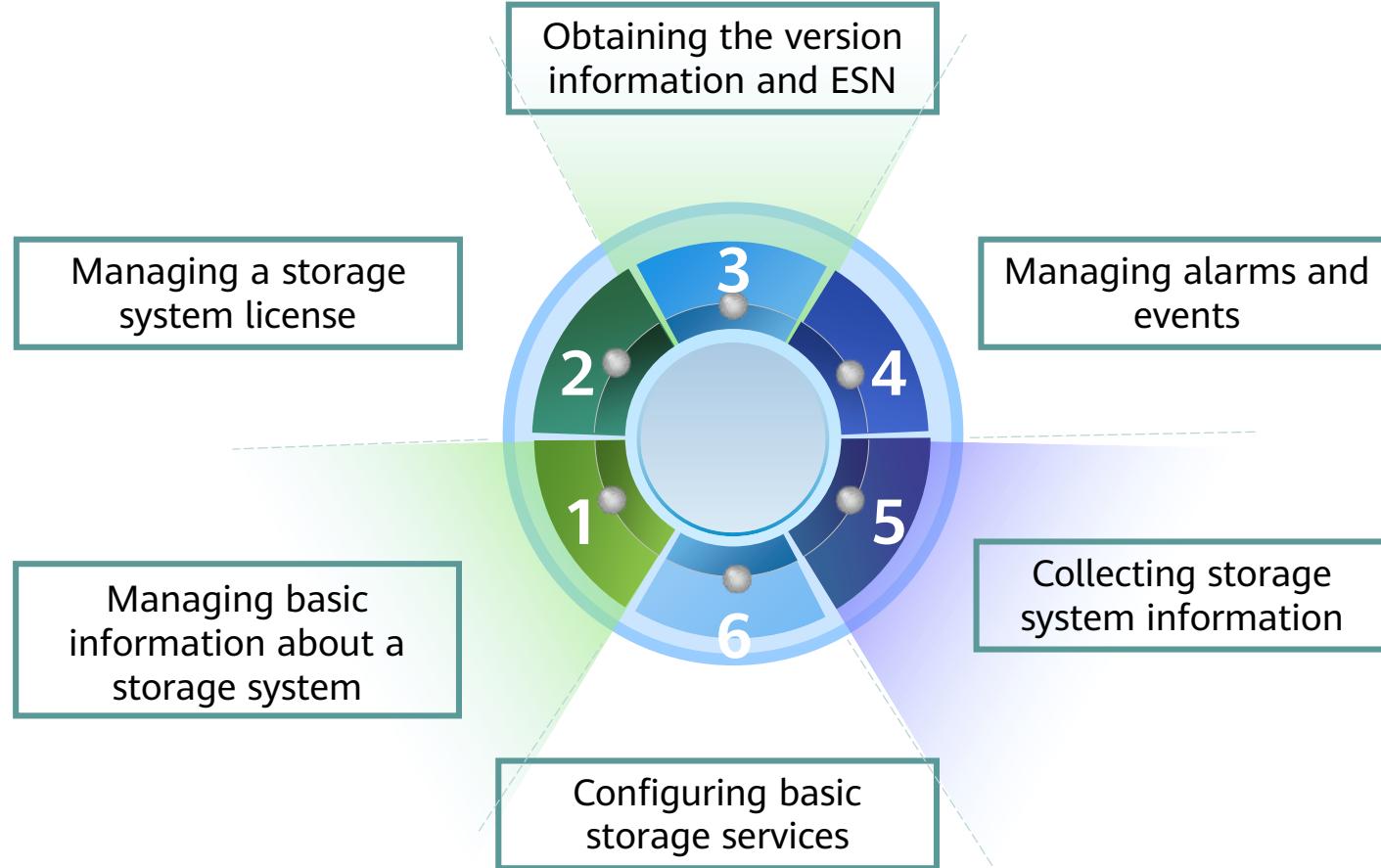
Load balancing

UltraPath interacts and negotiates with the storage system, calculates the hash result (LUN ID + LBA) for each delivered I/O, obtains the corresponding virtual node, and selects the physical path corresponding to the virtual node to deliver the I/O. This reduces cross-CPU distribution in a storage system and improves end-to-end performance.

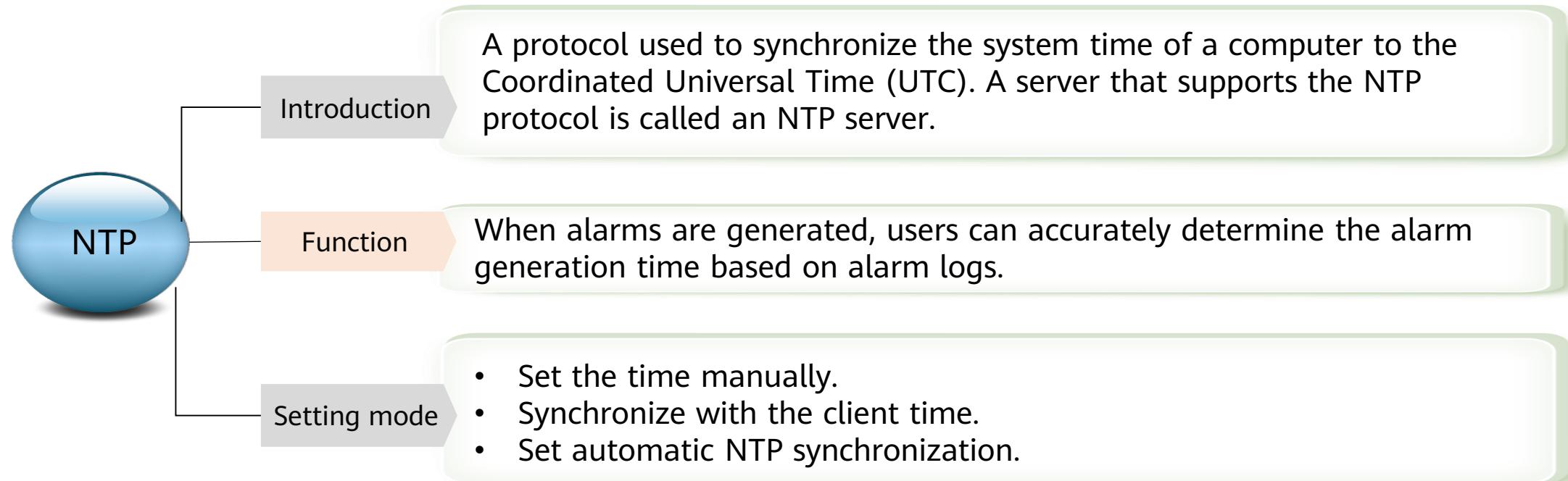
Contents

1. Storage Management Overview
2. Storage Management Tools
- 3. Basic Management Operations**

Basic Management Operations



Managing Basic Information About a Storage System - Setting the Device Time (1)



Managing Basic Information About a Storage System - Setting the Device Time (2)

- Managing the device time on DeviceManager

The screenshot shows the 'Device Time' configuration page. It displays the current time as '2022-06-20 15:52:34 UTC+08:00 (PRC)'. Below this, there is a 'Change Mode' section with three options: 'Synchronize with client time', 'Change manually', and 'Synchronize with NTP server time'. The third option is selected. At the bottom, there is a note: '* Auto NTP Sync' followed by a 'Configure' link.

- Managing the device time on the CLI

- The **change ntp_server config** command is used to automatically synchronize the storage system time with the NTP server time.
- The **show system ntp** command is used to query NTP settings.
- The **show ntp status** command is used to query the NTP status.
- The **show ntp_server general** command is used to query the settings of the time synchronization function.

Managing Device Licenses (1)



Introduction

Permission credentials for using various value-added features (such as snapshot, remote replication, clone, and SmartQoS)

Precautions

During routine device management, you need to check whether the license file is available.

Using DeviceManager to manage licenses

Depending on whether a license has been imported or activated, the license operation displayed in the **License Management** area can be **Import License**, **Activate License**, or **Update License**.

For an activated license file, DeviceManager provides two control modes:

- Running time-based control: displays the expiration time of the license.
- Capacity-based control: displays the used/total capacity of the license.

Managing Device Licenses (2)

Using CLI to manage licenses

- The **export license** command is used to export a license file.

Example: `export license ip=? user=? password=? license_path=? [port=?] [protocol=?]`

- The **import license** command is used to import a license file.

Example: `import license ip=? user=? password=? license_path=? [port=?] [protocol=?]`

- The **show license** command is used to query the function configuration of the imported license file in the system.

Example: `show license`

- The **show license_active** command is used to query information about active licenses.

Example: `show license_active`

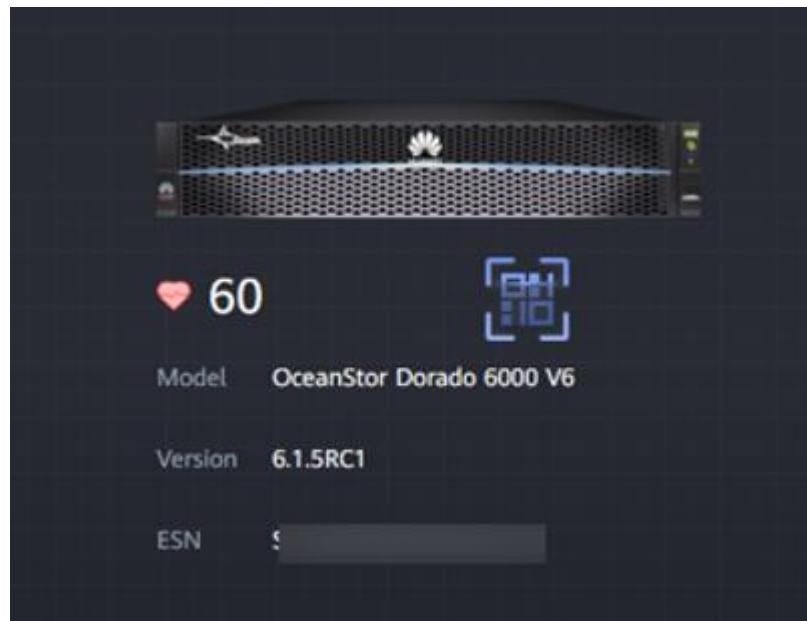
Obtaining the Current Version Information of the Device



Function

The matching software version can be accurately determined based on the system version.

- Obtain the current system version information on DeviceManager.



- Log in to the CLI as a super administrator.
 - Run the **show system general** command.
 - Product Version** indicates the version of the current storage system.

```
admin:/>show system general
System Name      : XXX.Storage
Health Status    : Normal
Running Status   : Normal
Total Capacity   : 3.186TB
SN               : 210235G6EHZ0CX0000XX
Location         :
Product Model    : XXXX
Product Version  : VX00R00XCXX
High Water Level(%) : 80
Low Water Level(%) : 20
WWN               : XXXX
Time              :
Patch Version    : SPCXXX SPHXXX
Description       :
```

Obtaining the Device ESN



Introduction

Character string that uniquely identifies a device.

Application scenarios

Scenarios such as license application, device repair, and eService service configuration

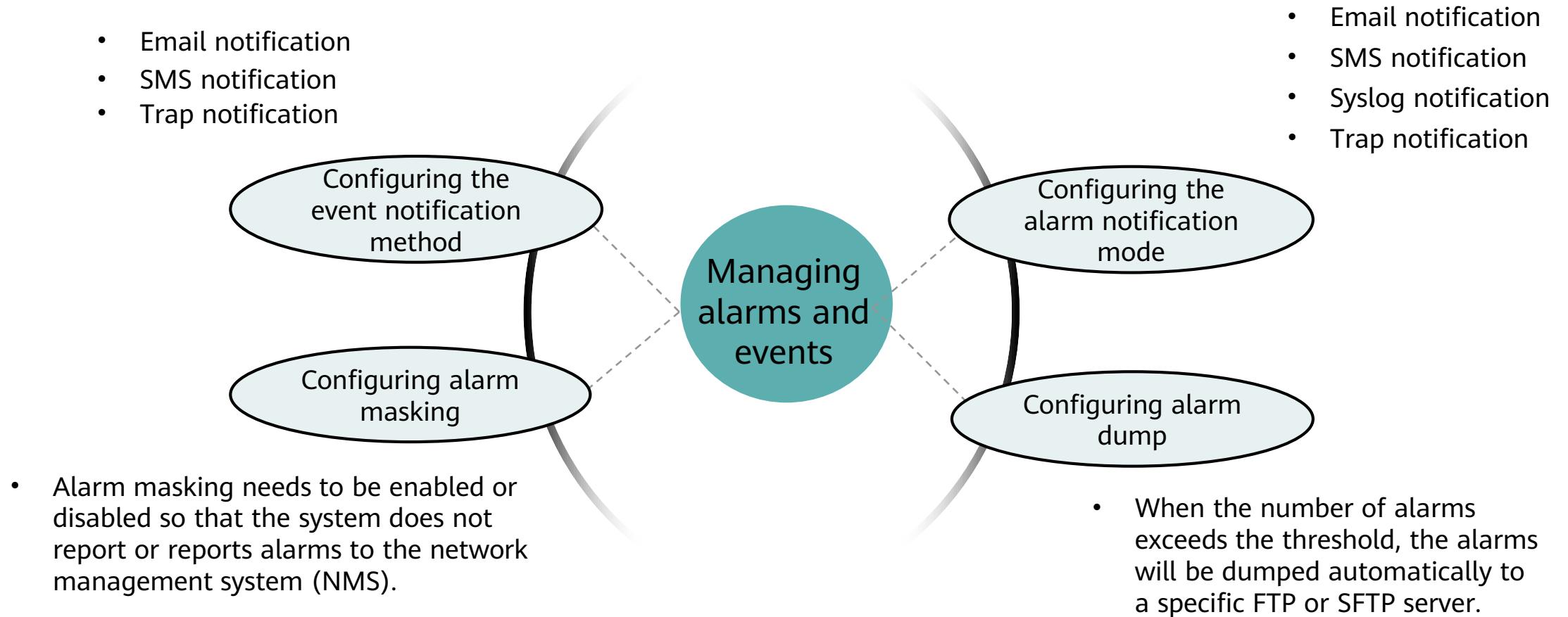
- Obtain the ESN using DeviceManager.



- Obtain the ESN using the CLI.
>> Run the **show system general** command.

```
admin:/>show system general
System Name      : XXX.Storage
Health Status    : Normal
Running Status   : Normal
Total Capacity   : 3.186TB
SN               : 210235G6EHZ0CX0000XX
Location         :
Product Model    : XXXX
Product Version  : VX00R00XCXX
High Water Level(%) : 80
Low Water Level(%) : 20
WWN              : XXXX
Time              :
Patch Version    : SPCXXX SPHXXX
Description       :
```

Managing Alarms and Events



Collecting Storage System Information (1)



- Purpose**
 - Prevent storage system faults and other unpredictable disasters from damaging the storage system.
 - Know the storage system operating status.
- How**
 - Regularly export and securely save the system data for fault locating and analysis.
- System data**
 - Configuration information, system logs, disk logs, and diagnosis files

Collecting Storage System Information (2)

Collecting storage system configuration data using DeviceManager

On DeviceManager:

- You can export the configuration information to collect the information about the current running status of the system.
- You can download **Recent logs** or **All logs** to collect configuration information, event information, and debugging logs on the storage device.
- You can download **DHA Runtime Log List** or **HSSD Log List** to collect disk run logs, I/O statistics and service life, and S.M.A.R.T. logs.
- You can export the diagnosis file to collect fault information of the device.

Collecting storage system configuration data using the CLI

Log in to the CLI of the storage system as the super administrator and run the following command to export the configuration file to an FTP or SFTP server:

```
export configuration_data ip=? user=? password=? db_file=? [ port=? ] [ protocol=? ] [ clean_device_file=? ]
```

Configuring Basic Storage Services



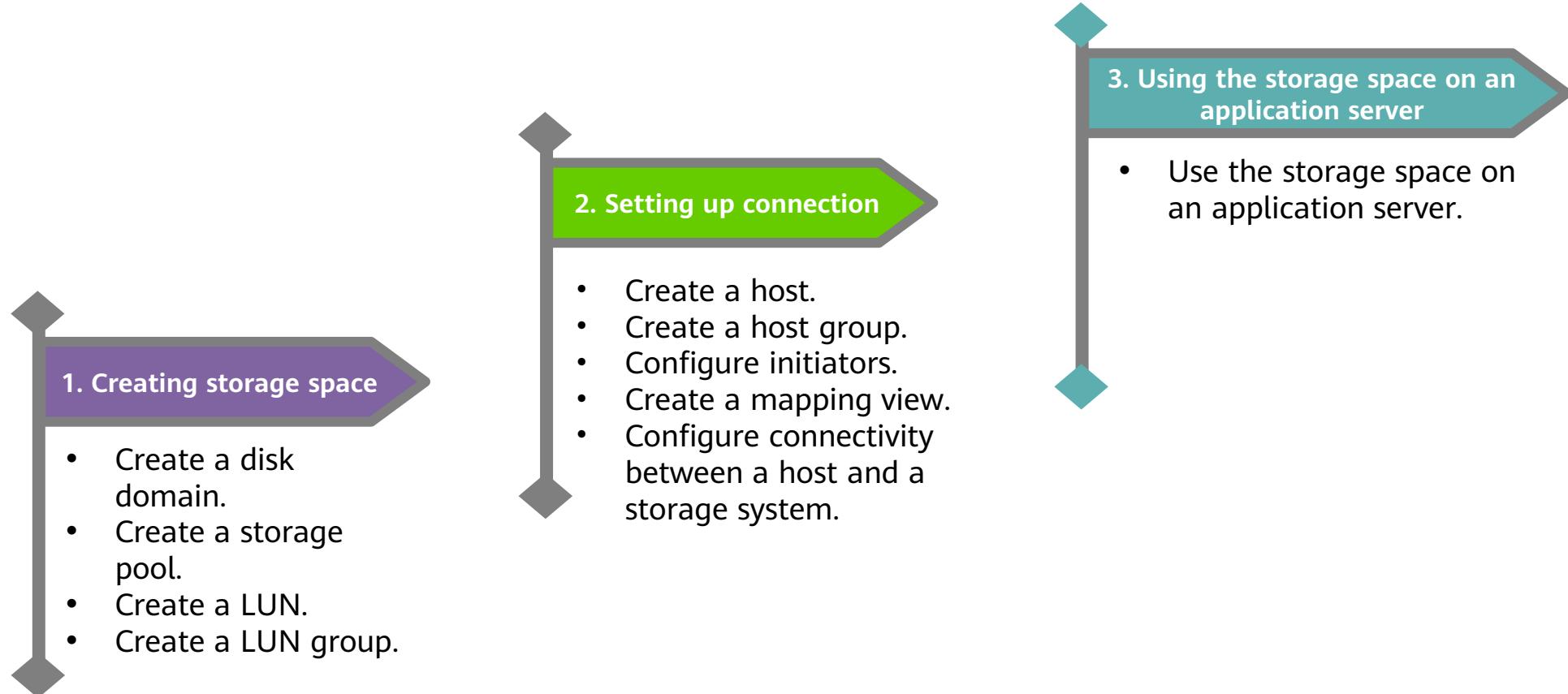
Function

- The storage space provided by the storage system is divided into multiple LUNs.
- Map LUNs to an application server.
- The application server can use the storage space provided by the storage system.

Using DeviceManager to configure basic storage services

- Creating a storage pool: DeviceManager allows you to create a storage pool in either recommended or custom mode.
- Allocating storage resources by creating LUN groups or file systems.

Configuring Basic Storage Services Using the CLI



Basic UltraPath Configuration Guide

Using the CLI to configure UltraPath:

For example, in Windows, choose **Start > All Programs > UltraPath > upadm** or enter **upadm** on the CLI of the Windows operating system. Then, run the CLI commands to configure UltraPath.

Note: Windows is used as an example to explain basic configuration commands for Huawei UltraPath. The configuration commands for other operating systems are similar. For details, see the user guide of the corresponding operating system.

UltraPath Parameter Settings in Typical Application Scenarios

In most scenarios, default settings of UltraPath are recommended. In some scenarios, you can configure UltraPath as instructed by the following:

`upadm set workingmode={0/1}`

- It specifies the load balancing mode at the storage controller level. **0** indicates load balancing between controllers. **1** indicates load balancing within a controller.
- The default setting is load balancing within a controller. UltraPath selects paths to deliver I/Os based on the owning controller of each LUN.
- When the inter-controller load balancing mode is used, UltraPath delivers I/Os to all paths. This increases latency due to I/O forwarding between controllers.

Typical Scenario	Recommended Configuration
The transmission paths between hosts and storage systems become a performance bottleneck.	0 : load balancing between controllers
Other scenarios	1 : default setting, load balancing within a controller

UltraPath Parameter Settings in Typical Application Scenarios

upadm set loadbalancemode={*round-robin/min-queue-depth/min-task*}

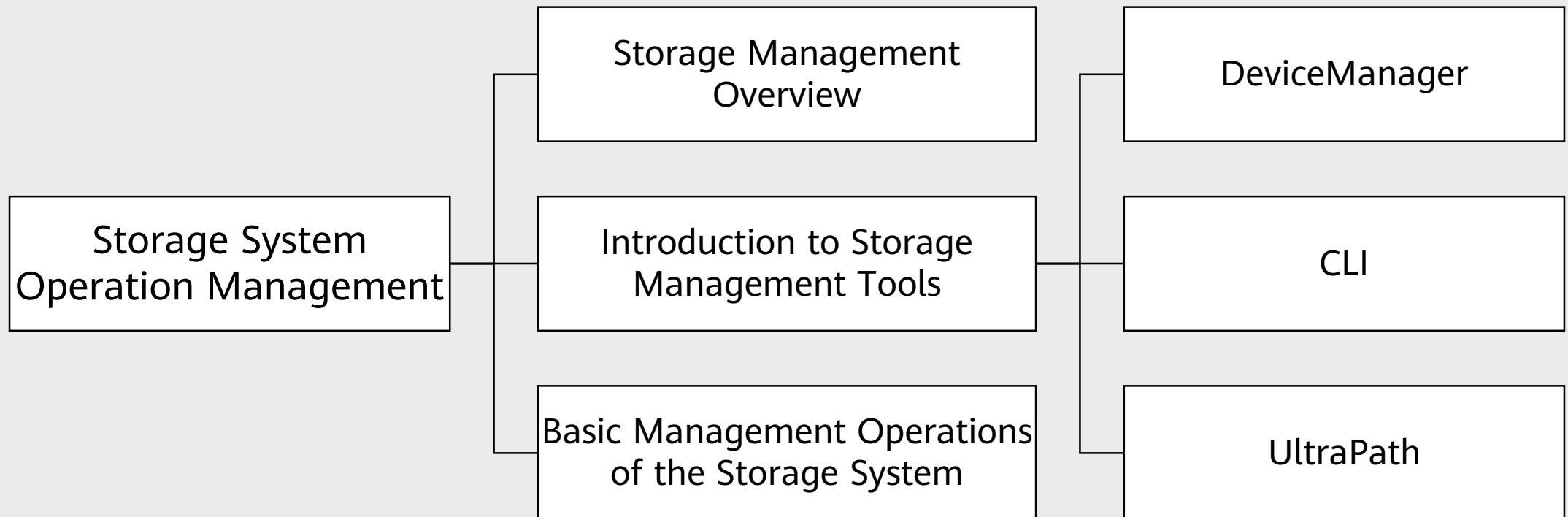
- It specifies the load balancing mode at the link level. The value can be **round-robin**, **min-queue-depth**, and **min-task**.
- The default algorithm is **min-queue-depth**. UltraPath selects the path that has the least number of I/Os from all available paths to deliver I/Os.
- When **round-robin** is used, UltraPath selects all available paths between the application server and storage system one by one to deliver I/Os.
 - When **min-task** is used, UltraPath selects the path that has the least I/O data volume from all available paths to deliver I/Os.

Typical Scenario	Recommended Configuration
The service I/O models delivered by hosts have small differences and I/Os need to be balanced on each path.	round-robin : round robin algorithm
The service I/Os delivered by hosts are large data blocks.	min-task : minimum task algorithm
Other scenarios	min-queue-depth : default setting, minimum queue depth algorithm

Quiz

1. (Single-answer question) The management IP address of a storage device is 192.168.5.12. Engineer A needs to enter () in the address box of the browser to log in to the storage device.
 - A. 192.168.5.12
 - B. http://192.168.5.12
 - C. https://192.168.5.12
 - D. https://192.168.5.12:8088
2. (True or false) DeviceManager can monitor the performance of controllers, front-end ports, and back-end ports. ()

Summary



Recommendations

- Huawei official websites:
 - Enterprise business: <https://e.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/en/index.html>
 - Online learning: <https://learning.huawei.com/en/>
- Popular tools
 - HedEx Lite
 - Network documentation tool center
 - Information query assistant

Acronyms and Abbreviations

LUN: Logical Unit Number. It is used to identify a logical unit, which is a device addressed by SCSI.

SAN: Storage Area Network

VLAN: Virtual Local Area Network. It is a group of hosts with a common set of requirements that communicate as if they were attached to the same broadcast domain, regardless of their physical location. VLAN membership can be configured through software instead of physically relocating devices or connections.

NTP: Network Time Protocol. It is an application layer protocol used to synchronize the time between the distributed time server and the client.

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.



Storage Resource Tuning Technologies and Applications



Foreword

- This course describes storage resource tuning technologies, including SmartThin, SmartTier, SmartCache, SmartAcceleration, SmartQoS, SmartDedupe, SmartCompression, SmartVirtualization, and SmartMigration, as well as their service characteristics, implementation principles, and application scenarios.

Objectives

- On completion of this course, you will be able to understand the service characteristics, implementation principles, and application scenarios of the following features:
 - SmartThin
 - SmartTier&SmartCache
 - SmartAcceleration
 - SmartQoS
 - SmartDedupe&SmartCompression
 - SmartVirtualization
 - SmartMigration

Contents

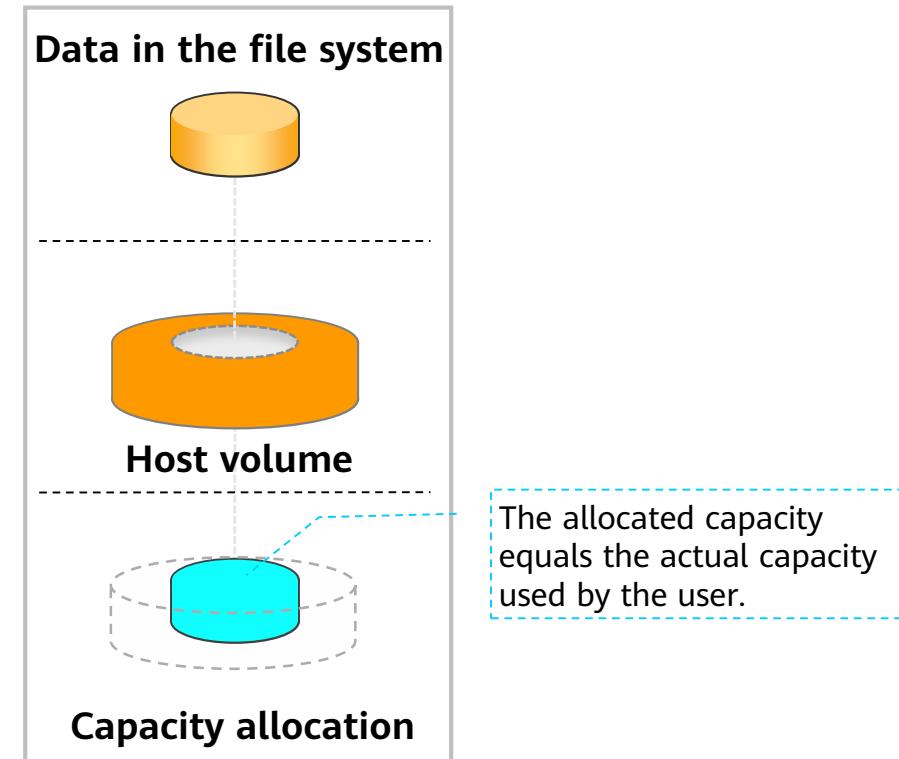
- **SmartThin**
- SmartTier&SmartCache
- SmartAcceleration
- SmartQoS
- SmartDedupe&SmartCompression
- SmartVirtualization
- SmartMigration

Overview

- The traditional deployment of a storage system has the following problems:
 - Adverse impact or even interruption on services when expanding the storage space
 - Uneven storage space utilization
 - Low storage efficiency
- SmartThin
 - SmartThin uses the on-demand storage space allocation policy to improve storage resource utilization and meet service requirements to a greater extent.

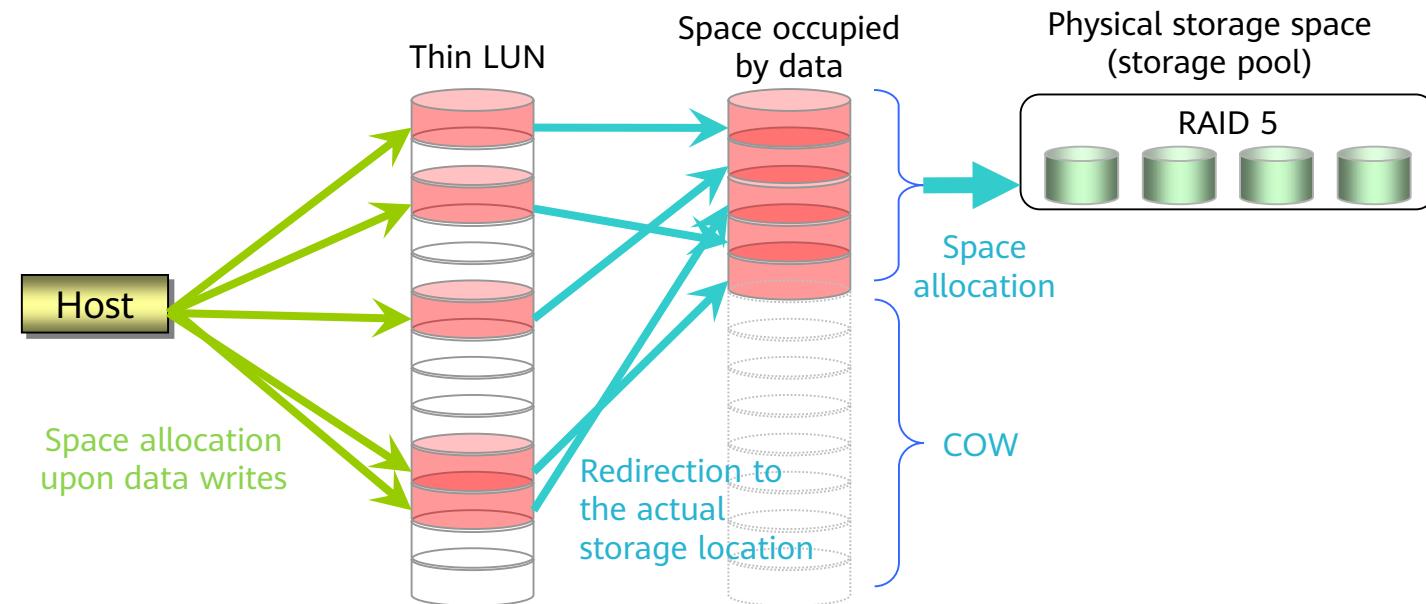
Thin LUN

- Definition: A thin LUN is a logical disk that can be accessed by hosts. It dynamically allocates storage resources from the storage pool according to the actual capacity requirements of users.
 - ✓ Data collection: From the perspective of a storage system, a thin LUN is a LUN that can be mapped to a host.
 - ✓ Full availability: data can be read and written properly.
 - ✓ Dynamic allocation: Resources are allocated while data is being written.

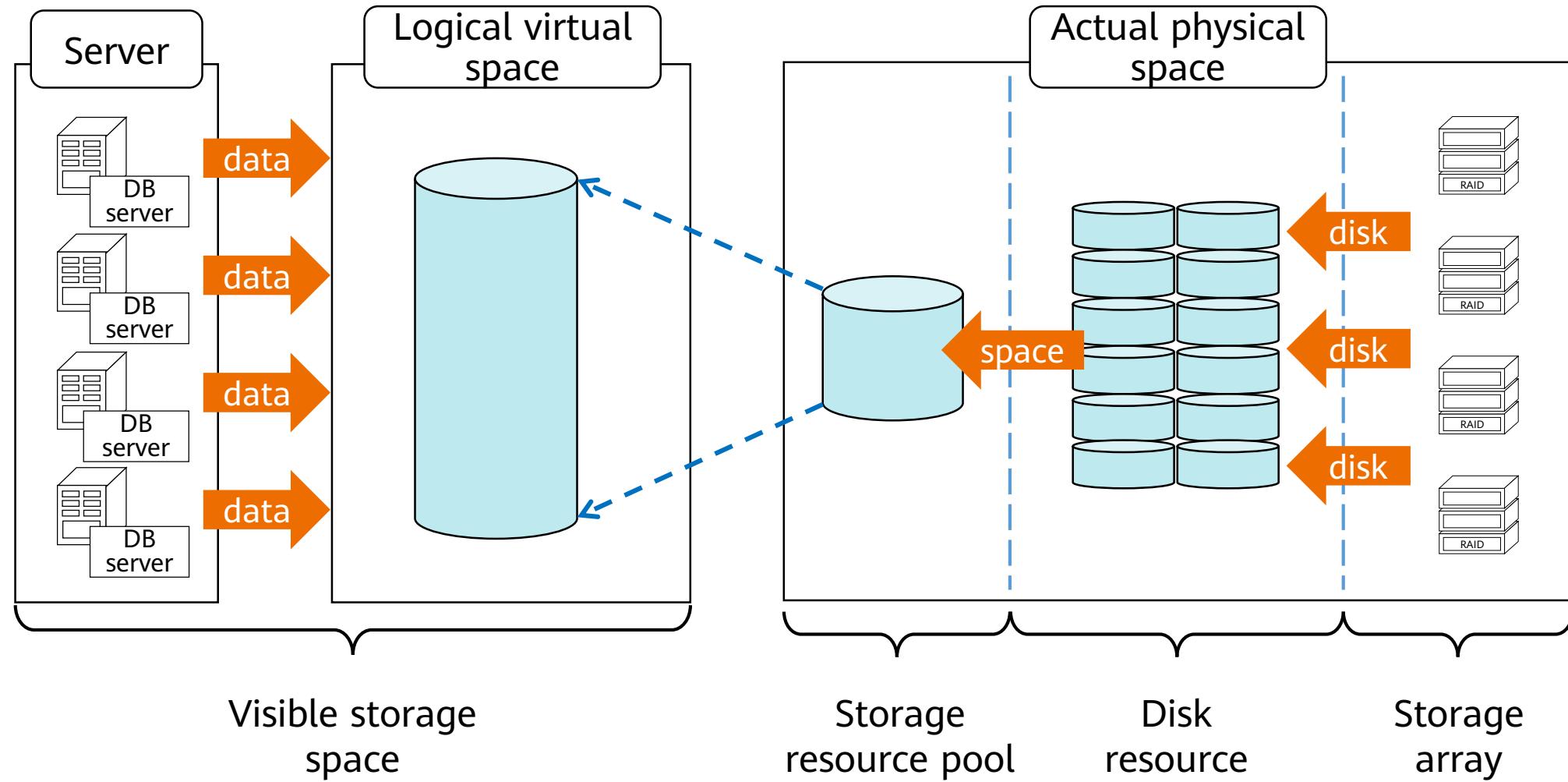


Storage Virtualization

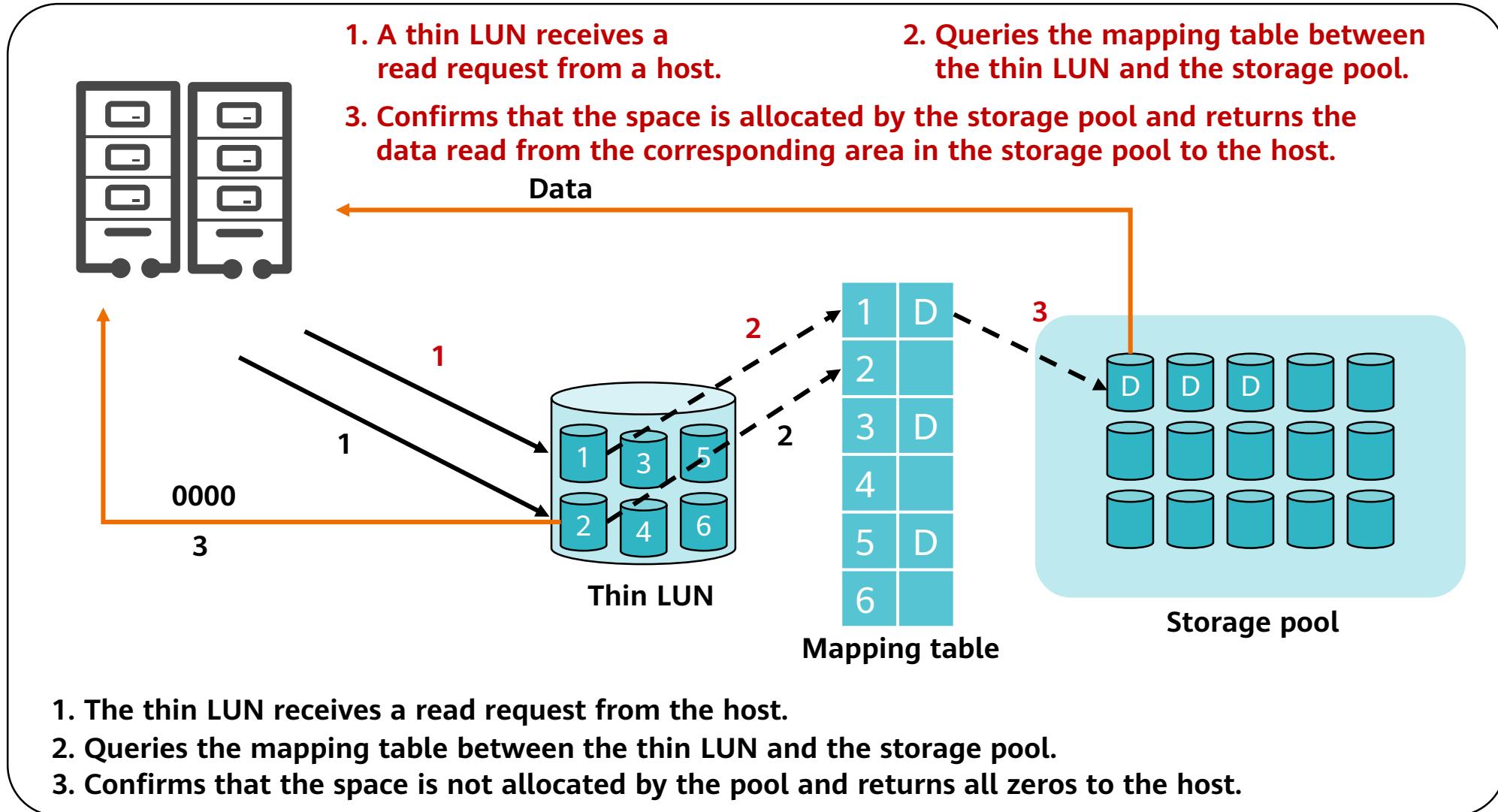
- Capacity-on-write (COW): Storage space is allocated from engines upon data writes based on load balancing rules.
- Direct-on-time (DOT): Data reads from and writes to a thin LUN are redirected.



Working Principles of SmartThin

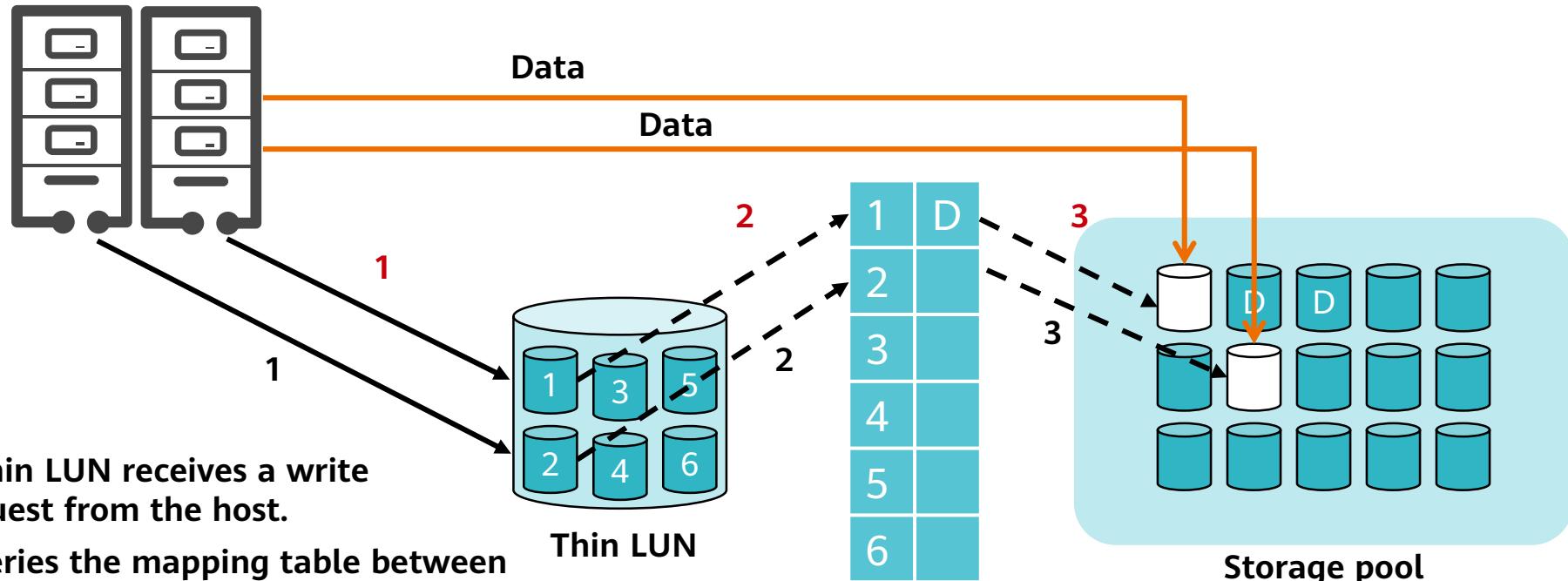


SmartThin Read Process



SmartThin Write Process

1. A thin LUN receives a write request from a host.
2. Queries the mapping table between the thin LUN and the storage pool.
3. Confirms that the space is allocated by the pool and performs the write process on the corresponding area in the storage pool. If the write request asks for releasing space, the space is released.



1. A thin LUN receives a write request from the host.
2. Queries the mapping table between the thin LUN and the storage pool.
3. If the space is not allocated by the pool, the storage system allocates the space first. And then performs write process on the corresponding area in the storage pool. If the write request asks for releasing space, a message is returned to the host.

Application Scenarios

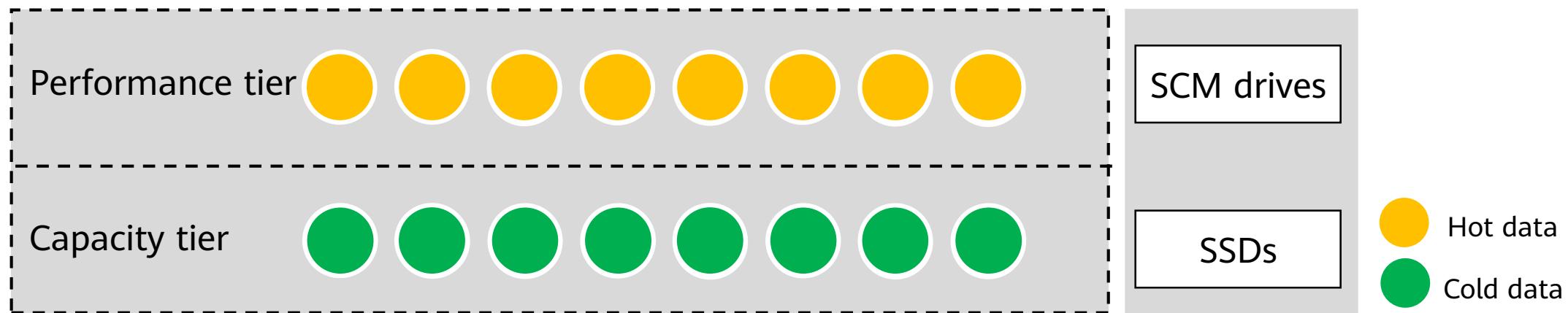
- SmartThin can help core system services that require high service continuity, such as bank transaction systems, expand system capacity online without interrupting ongoing services, such as back transaction systems.
- SmartThin can assist with on-demand physical space allocation for services where the growth of application system data is hard to be accurately evaluated, such as email services and web disk services, preventing a space waste.
- SmartThin can assist with physical space contention for mixed services that have diverse storage requirements, such as carriers' services, to achieve optimized space configuration.

Contents

- SmartThin
- **SmartTier&SmartCache**
- SmartAcceleration
- SmartQoS
- SmartDedupe&SmartCompression
- SmartVirtualization
- SmartMigration

SmartTier

- Tiered storage migrates hot data, cold data, and data with different values in use to specific storage media to effectively balance performance.
- SmartTier is also called intelligent data tiering. SmartTier provides intelligent data storage management. It automatically matches data of different activity levels with storage media of different characteristics by collecting and analyzing data activity levels.

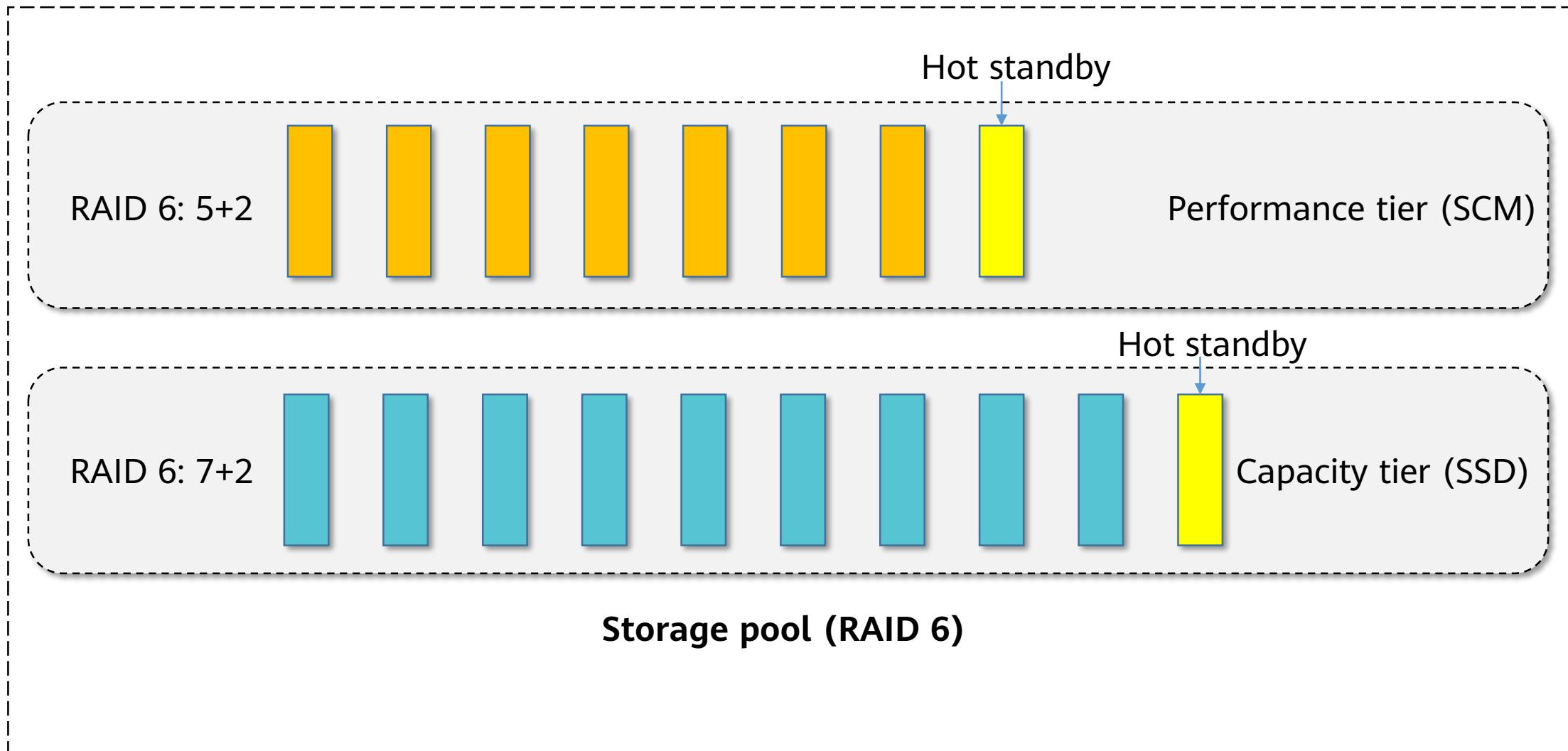


Dividing Storage Tiers

- In the same storage pool, a storage tier is a collection of storage media with the same performance. SmartTier divides storage media into performance and capacity tiers. The performance tier that consists of SCM drives delivers higher performance than the capacity tier that consists of SSDs. Each storage tier respectively uses the same type of disks and RAID policy.

Storage Tier	Disk Type	Disk Feature	Data Feature
Performance tier	SCM drive	Very short response time and high cost per gigabyte.	Suitable for storing frequently accessed data.
Capacity tier	SSD	Short response time and moderate cost per gigabyte.	Suitable for storing less frequently accessed data.

Managing Member Disks



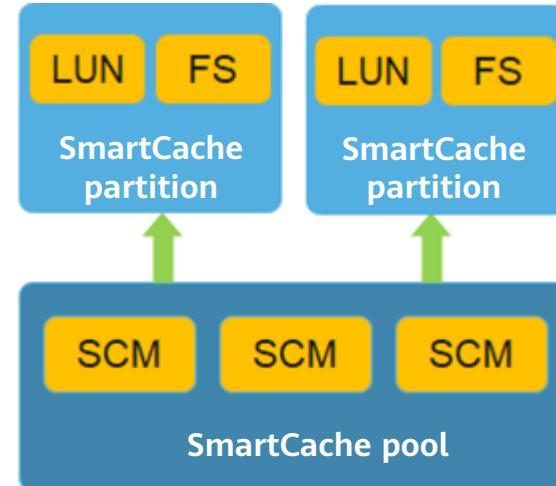
Data Migration

- SCM drives provide better performance than SSDs. With SmartTier available, new data from hosts is preferentially written to the performance tier for better performance. Cold data that is not accessed for a long time is migrated to the capacity tier in the background.
- Data management and migration policies are as follows:
 - **Metadata:** Metadata for new writes and garbage collection (GC) is preferentially stored at the performance tier (SCM) to ensure high-performance access in large-capacity scenarios. Space is allocated from the capacity tier only when the space of the performance tier is insufficient.
 - **User data:** User data is preferentially stored at the performance tier (SCM). When the size of data in the performance tier reaches a certain level, cold data is migrated to the capacity tier (SSD) in the background. This ensures that the performance tier stores more hot data and metadata, allowing faster data accessibility.

SmartCache

Advantages of SCM:

- Lower latency and higher IOPS than SSDs
- Larger capacity and lower price than DRAM
- Suitable for using as cache



- **SmartCache:**

Based on the short read response time of SCM drives, SmartCache uses SCM drives to compose a SmartCache pool and caches frequently-read data to the SmartCache pool. This shortens the response time for reading hot data, improving system performance.

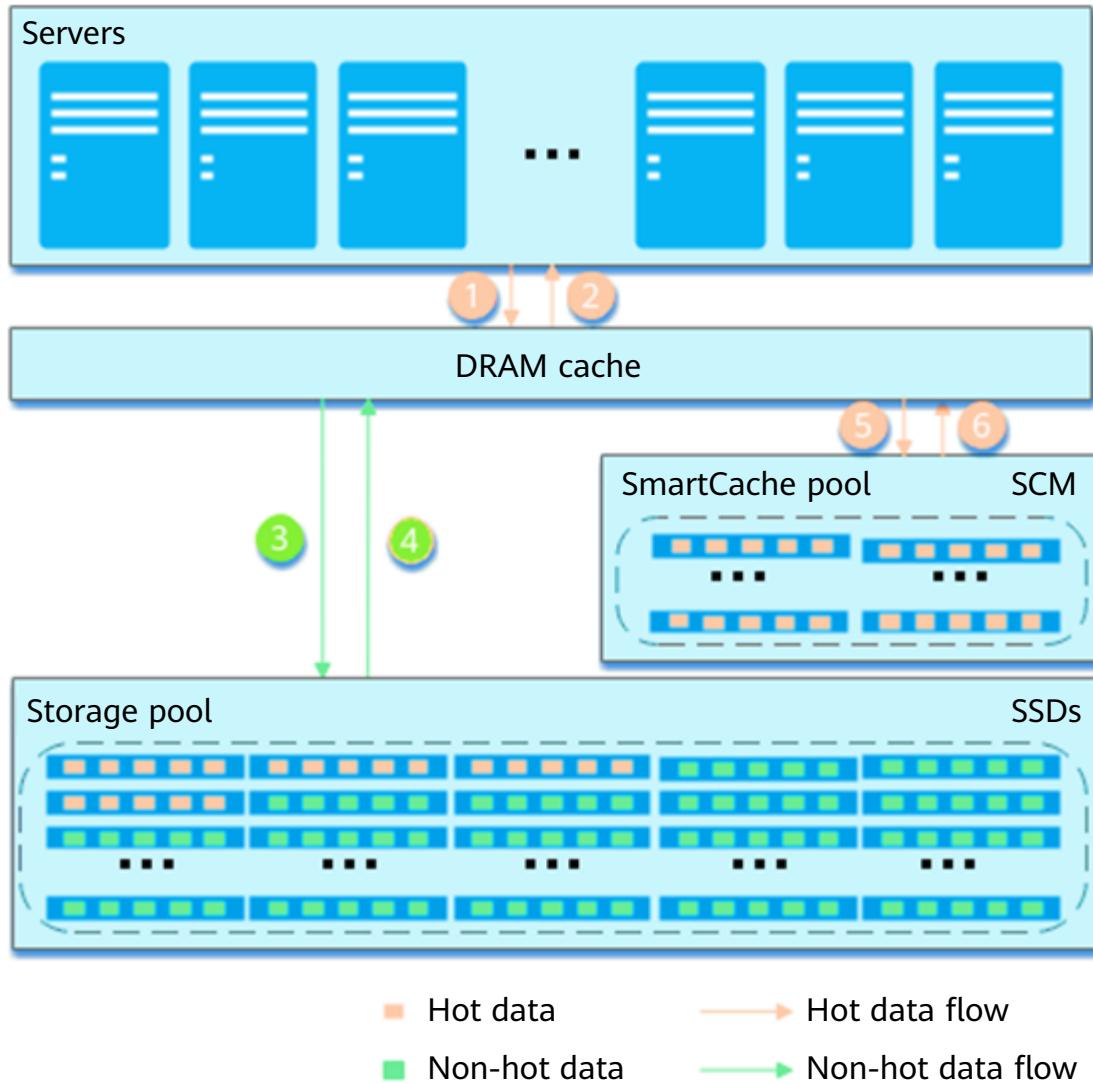
- **SmartCache pool:**

A SmartCache pool consists of SCM drives and is used as a complement of DRAM cache to store hot data.

- **SmartCache partition:**

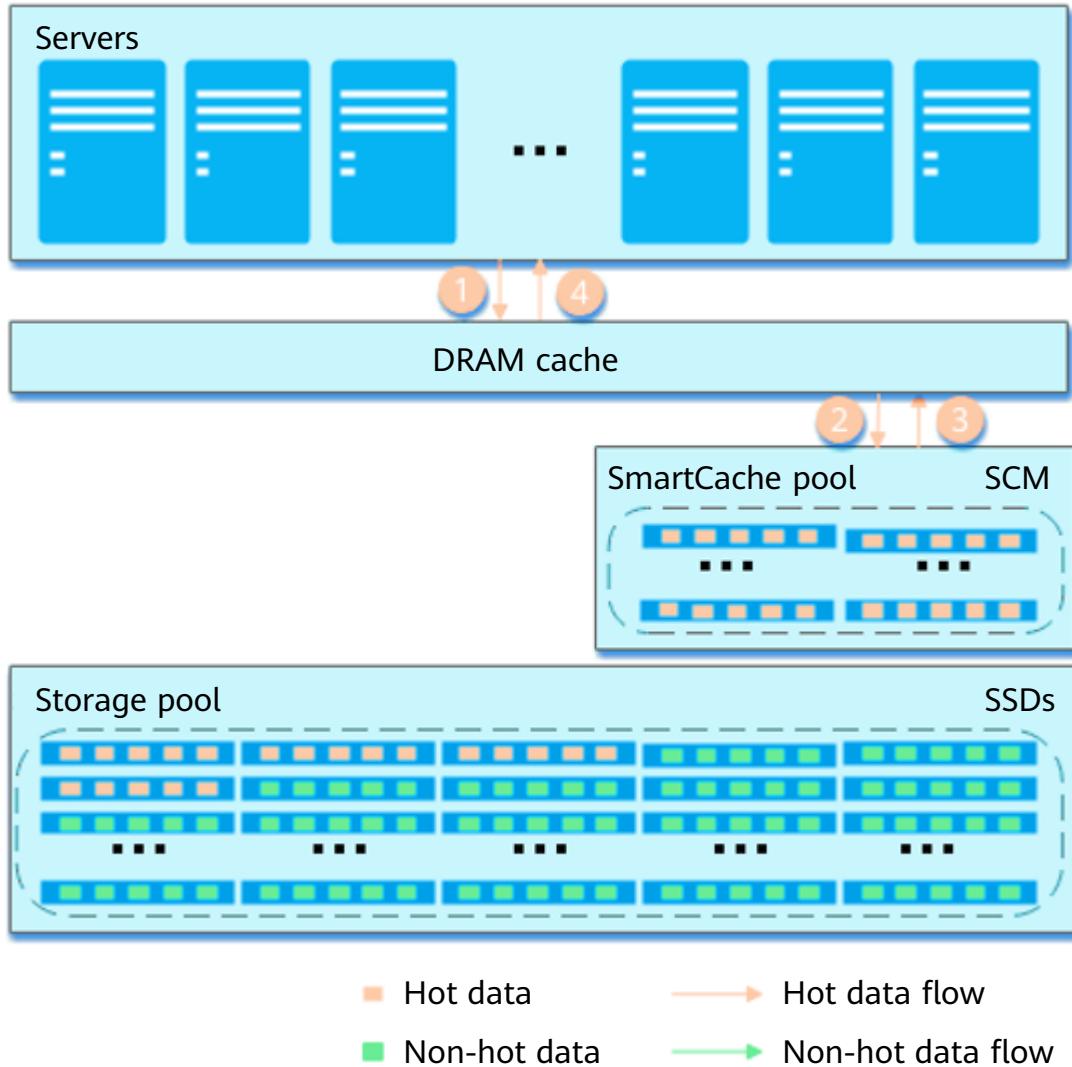
The SmartCache partition is a logical concept based on the SmartCache pool. It is used to store LUNs and file system services.

SmartCache Write Process



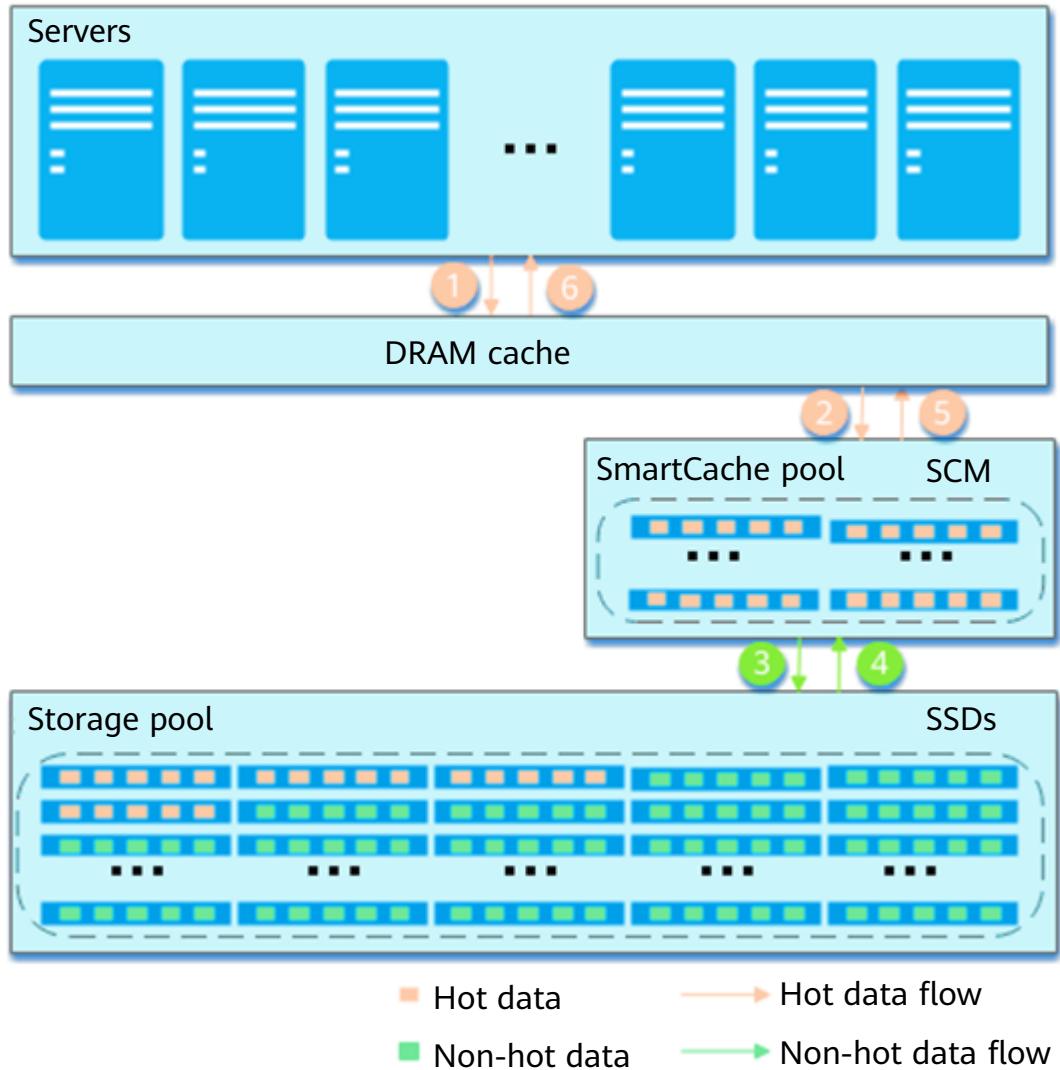
1. After receiving a write I/O request to a LUN or file system from a server, the storage system sends data to the DRAM cache.
2. After the data is written to the DRAM cache, an I/O response is returned to the server.
3. The DRAM cache sends the data to the storage pool management module.
4. Data is stored on SSDs, and an I/O response is returned.
5. The DRAM cache sends data copies to the SmartCache pool. After the data is filtered by the cold and hot data identification algorithm, the identified hot data is written to the SCM media, and the metadata of the mapping between the data and SCM media is created in the memory.
6. Data is cached to the SmartCache pool, and an I/O response is returned.

SmartCache Read Hit



1. A read I/O request from an application server is first delivered to the DRAM cache before arriving at LUNs or file systems.
2. If the requested data is not found in the DRAM cache, the read I/O request is further delivered to the SmartCache pool.
3. If the requested data is found in the SmartCache pool, the read I/O request is delivered to SCM drives. Data is read from the SCM drives and returned to the DRAM cache.
4. The DRAM cache returns the data to the application server.

SmartCache Read Miss



1. A read I/O request from an application server is first delivered to the DRAM cache before arriving at LUNs or file systems.
2. If the requested data is not found in the DRAM cache, the DRAM cache forwards the read I/O request to the SmartCache pool.
3. If the requested data is not found in the SmartCache pool either, the SmartCache module forwards the read I/O request to the storage pool management module to read the data from SSDs.
4. Data is read from the SSDs and returned to the SmartCache module.
5. The SmartCache module returns the data to the DRAM cache.
6. The DRAM cache returns the data to the application server.

Highlights

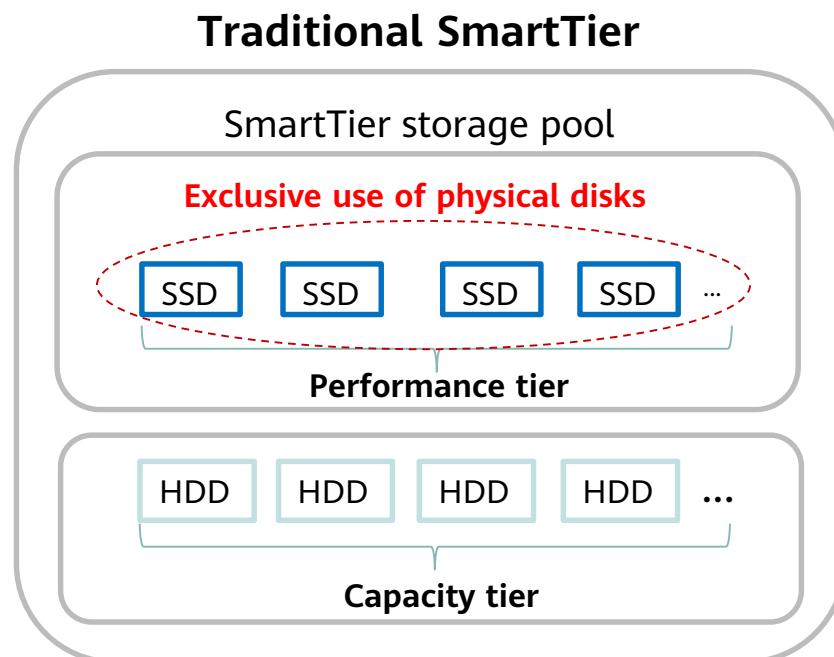
- **Dynamic Capacity Expansion**
 - You can add SCM media to a SmartCache pool for dynamic capacity expansion.
 - You are advised to configure SCM resources of the same quantity and capacity for controllers in the same controller enclosure. This ensures balanced acceleration performance for LUN or file system services of multiple controllers in the same controller enclosure.
- **Flexible Policy Configuration**
 - You can enable or disable SmartCache for specific LUNs or file systems without interrupting services.
 - You can add one or more LUNs or file systems to a specified SmartCache partition to shorten the data read response time.
- **Adaptive Switch**
 - When detecting that the SmartCache policy is inefficient (for example, when the SmartCache hit ratio is low or the CPU is busy), the system stops allowing I/Os to enter the SmartCache pool to reduce the impact on services. In this case, I/Os are not sent to the SmartCache pool and the requested data will not be found in the SmartCache pool. In addition, to ensure data consistency, data in the SmartCache pool is cleared.
 - When detecting a scenario suitable for SmartCache, the system automatically restores to the previous state.

Contents

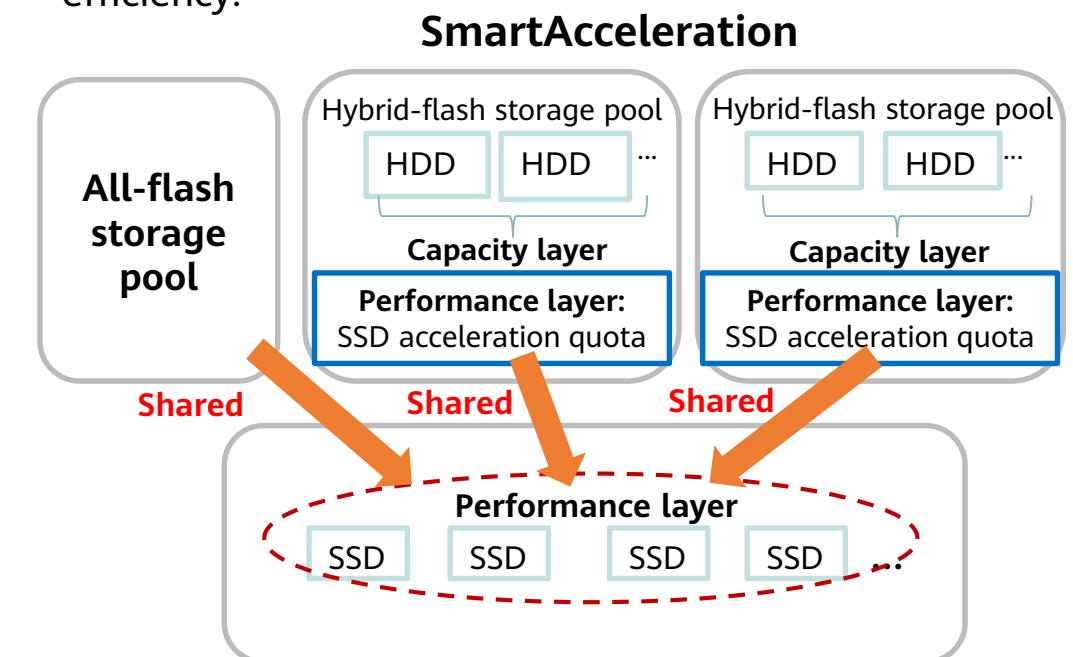
- SmartThin
- SmartTier&SmartCache
- **SmartAcceleration**
- SmartQoS
- SmartDedupe&SmartCompression
- SmartVirtualization
- SmartMigration

SmartAcceleration

- SSDs at the performance tier are exclusively used and cannot be shared by multiple disk domains.
- Performance tier configurations are difficult to adjust, lacking flexibility.
- Tiers and caches are independent of each other, and physical disks must be planned separately for both tiers and caches.

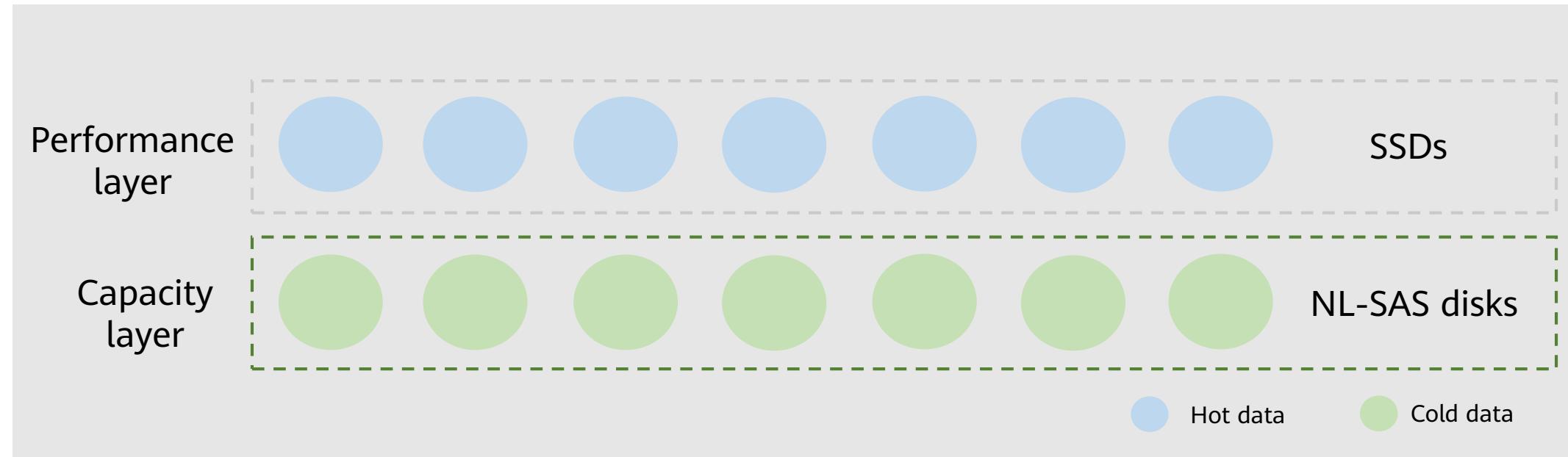


- The performance layer is shared by multiple disk domains, improving utilization.
- The performance layer is distributed by the system to each pool based on the recommended quota by default, and can be flexibly scaled.
- The performance layer converges tiers and caches, and is automatically configured for optimal overall efficiency.

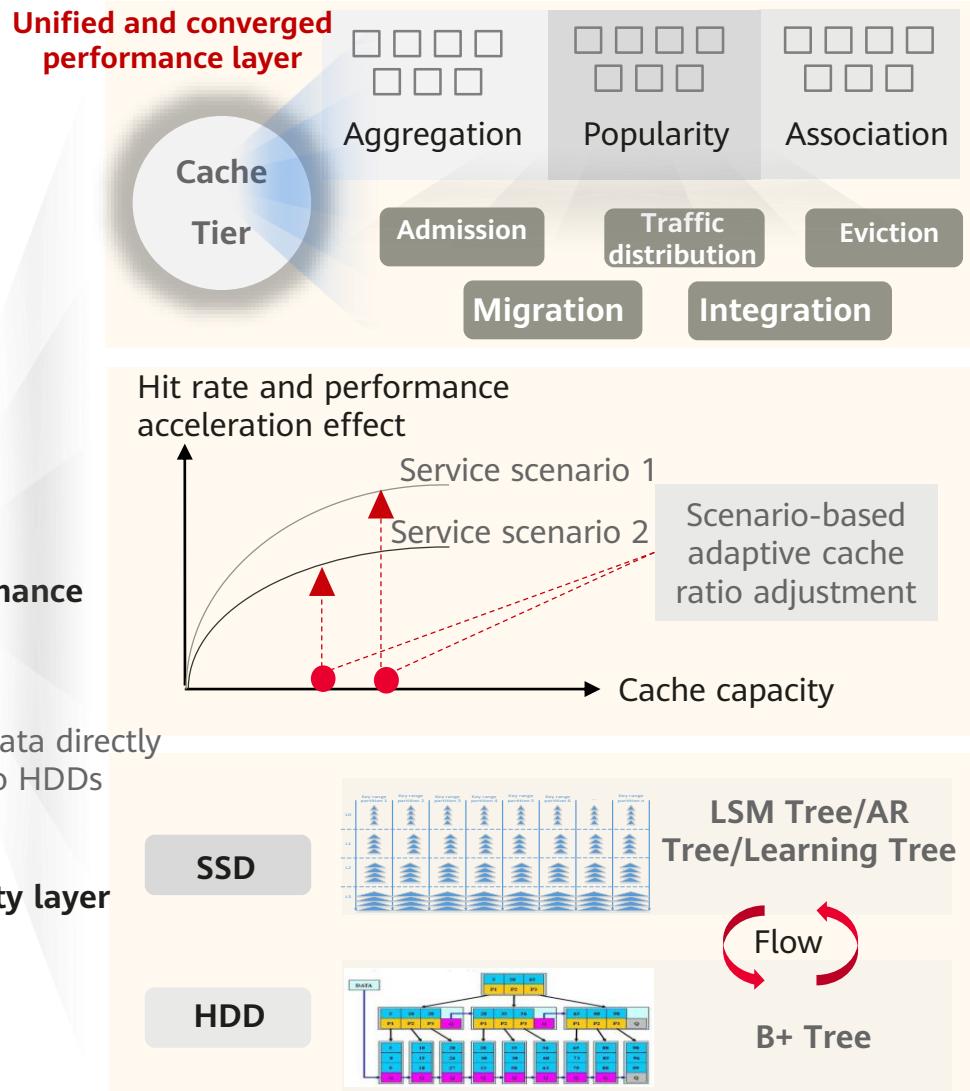
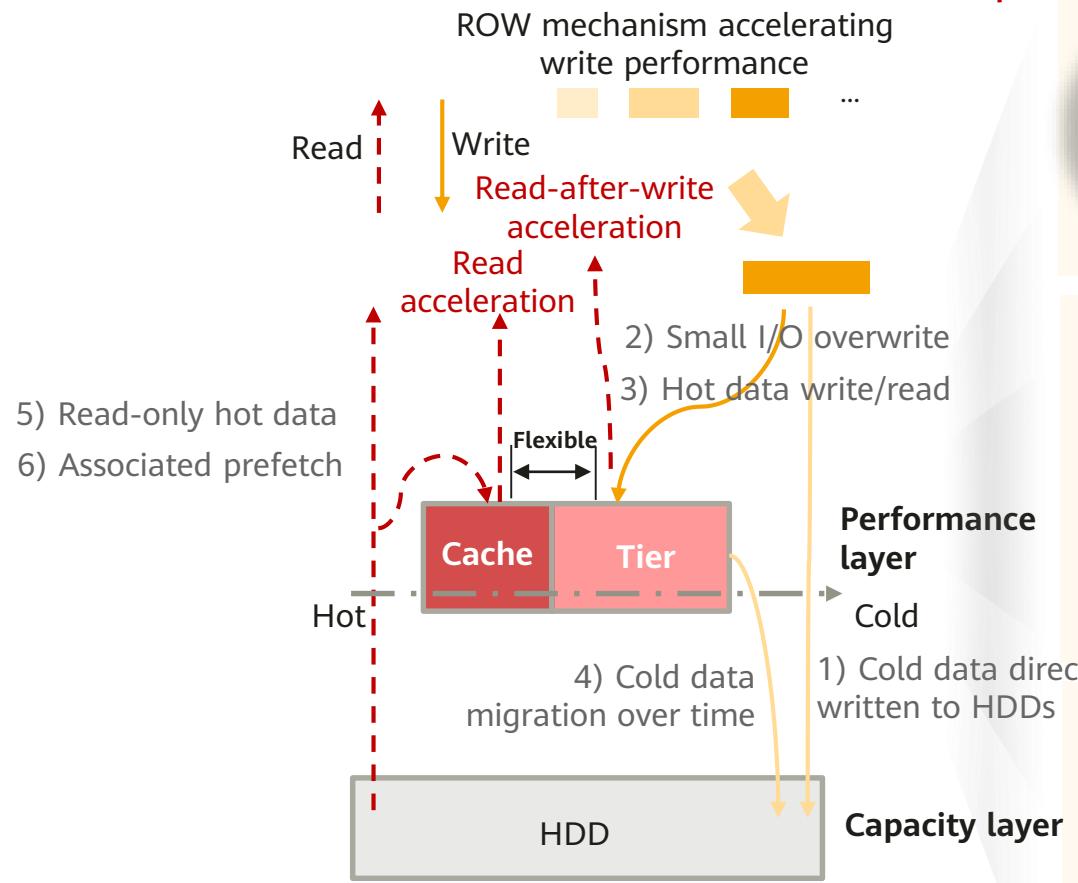


Storage Layer

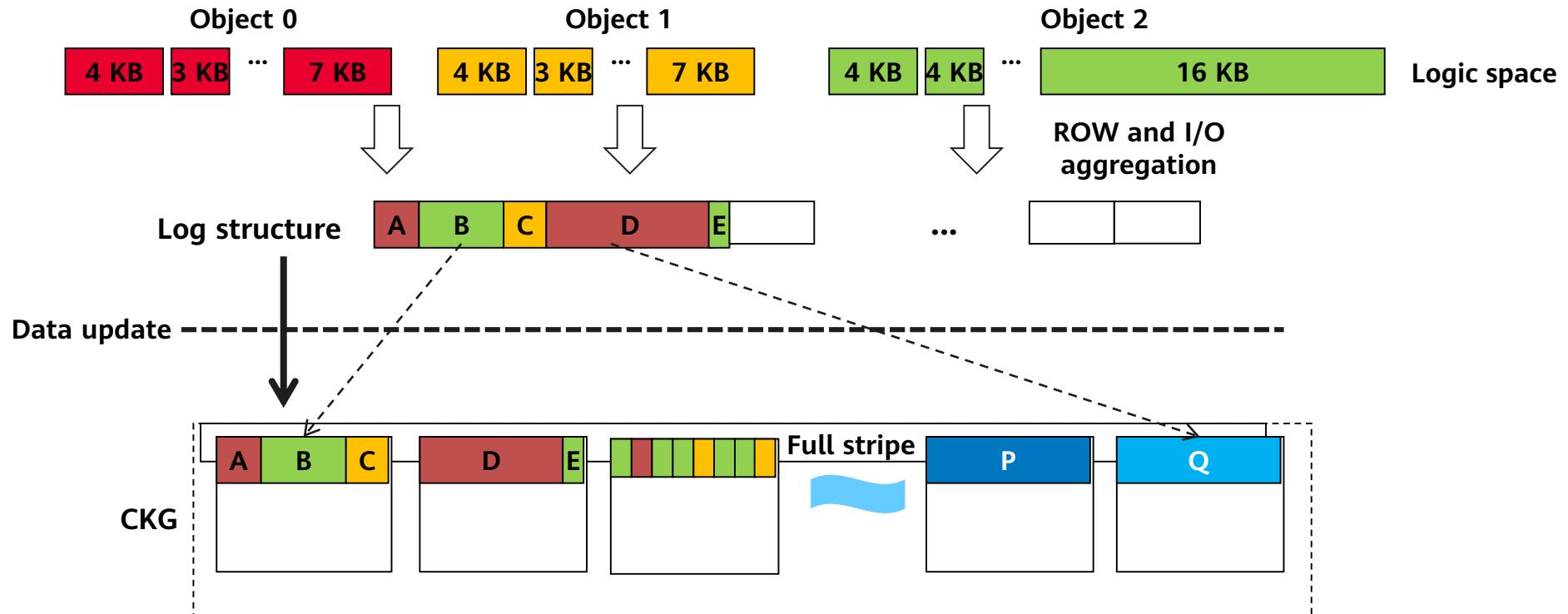
- Random distribution of hot and cold data cannot maximize the characteristics of disks on various media. SmartAcceleration uses a multi-dimensional adaptive hot and cold data sensing algorithm to implement a unified performance layer, where hot and cold data are automatically identified based on their characteristics and then automatically distributed, achieving optimal distribution of hot and cold data.



Unified Performance Layer That Flexibly Integrates Caches and Tiers

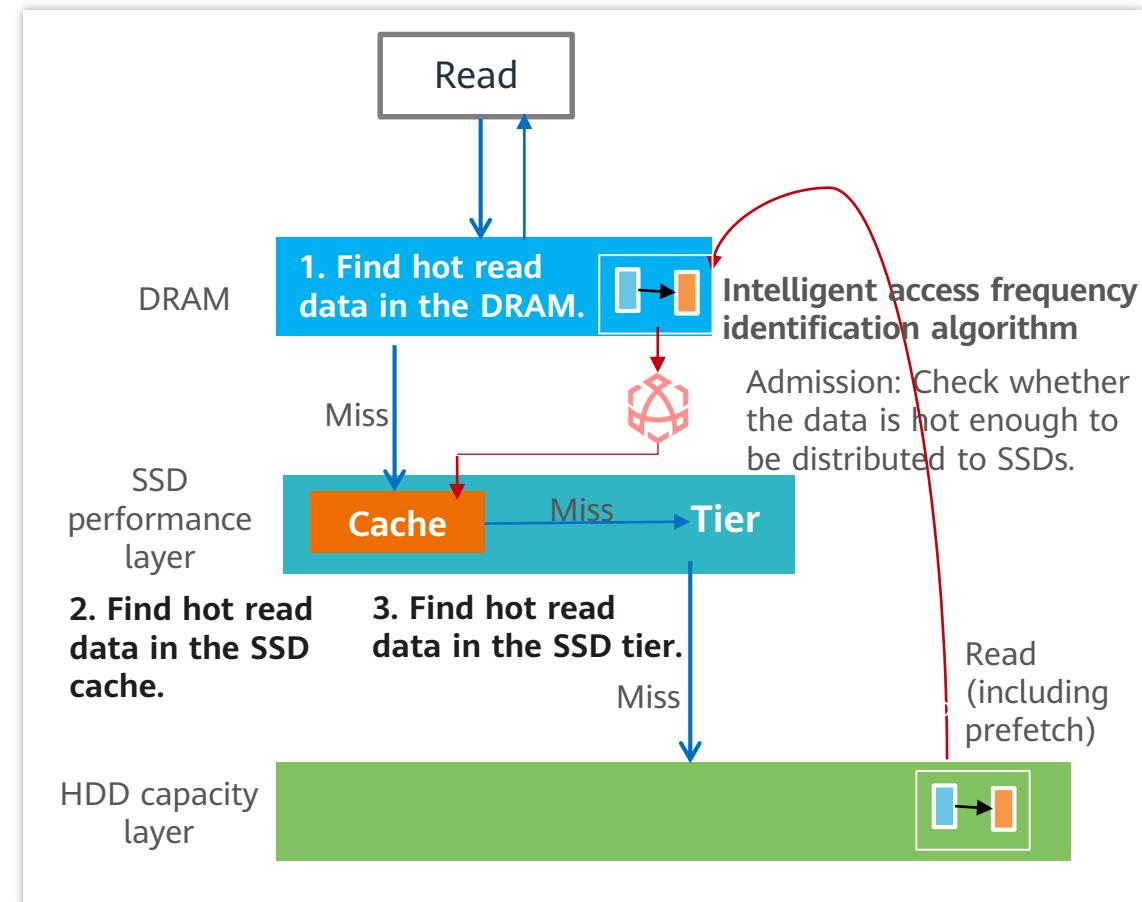
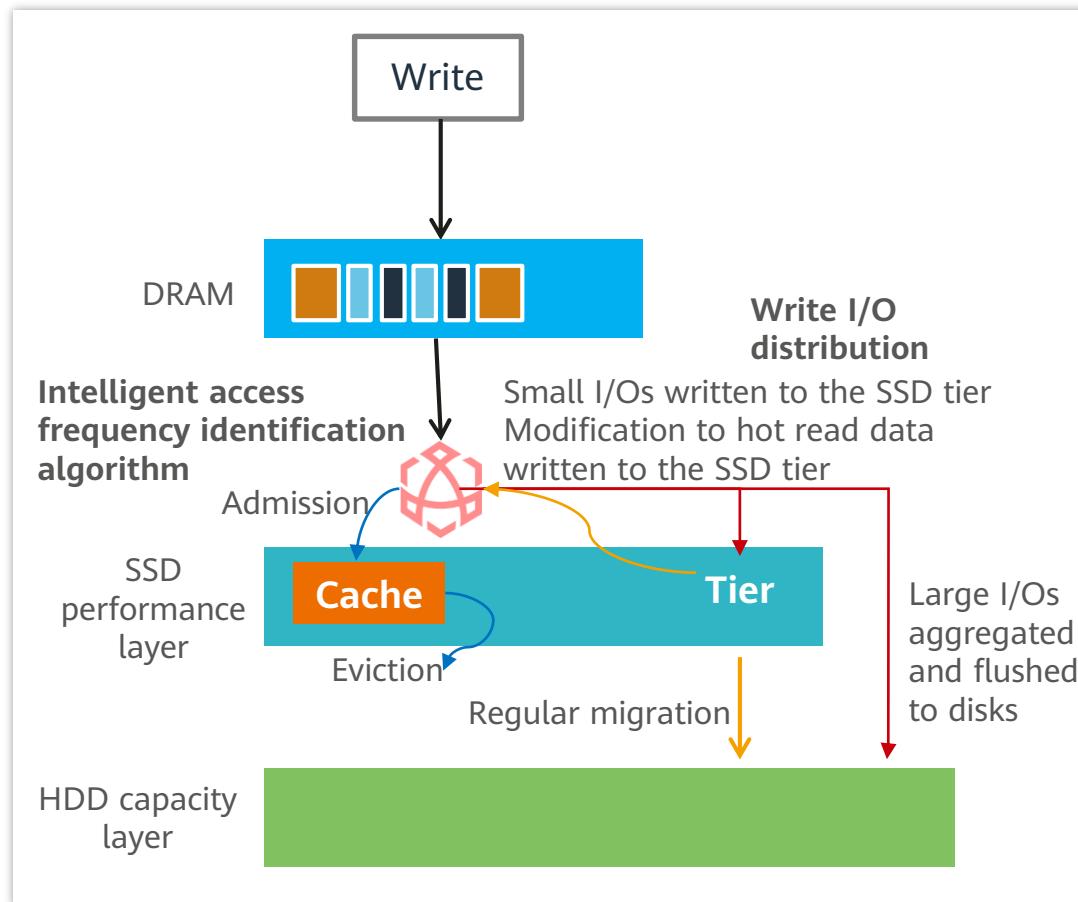


ROW-based Large-Block Sequential Write



Traditional Mode				OceanStor Mode			
Configuration	Read Count	Write Count	Total I/Os	Configuration	Read Count	Write Count	Total I/Os
RAID-5	2	1	4 (3)	RAID-5	0	0	1
RAID-6	3	2	6 (5)	RAID-6	0	0	1
RAID-TP	4	3	8 (7)	RAID-TP	0	0	1

Working Principles of SmartAcceleration



Contents

- SmartThin
- SmartTier&SmartCache
- SmartAcceleration
- **SmartQoS**
- SmartDedupe&SmartCompression
- SmartVirtualization
- SmartMigration

Overview

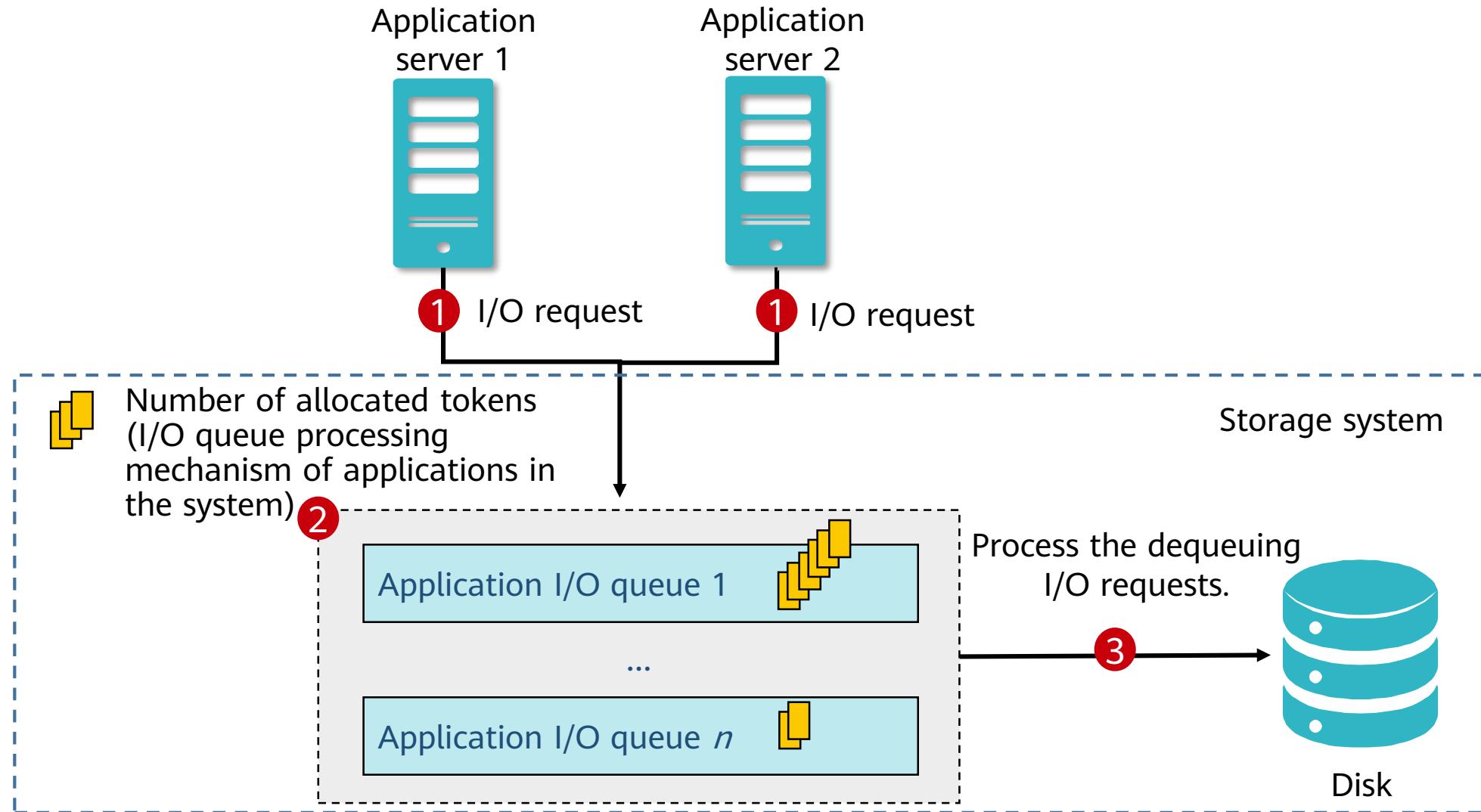


- **SmartQoS:**

SmartQoS is an intelligent service quality control feature. It dynamically allocates storage system resources to meet the performance requirement of certain applications.

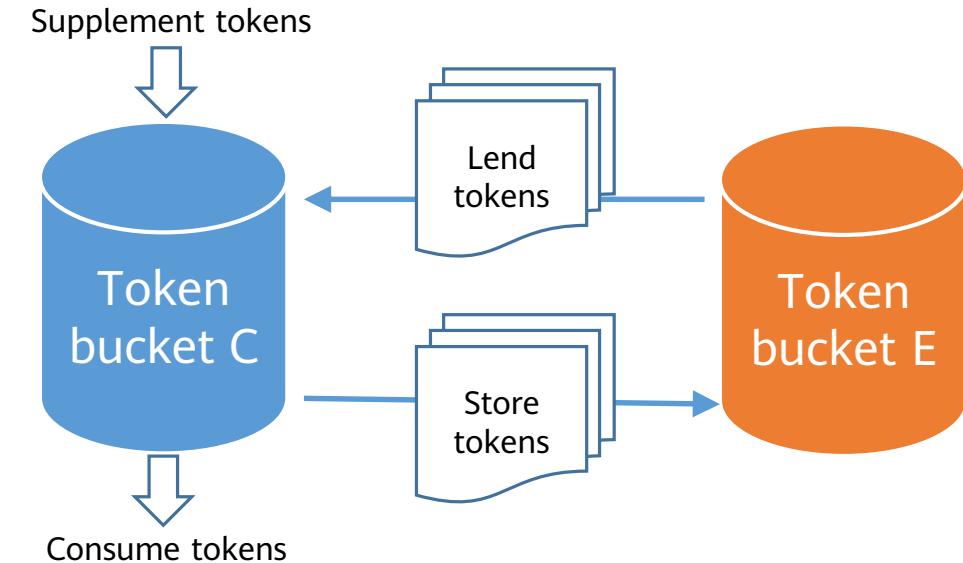
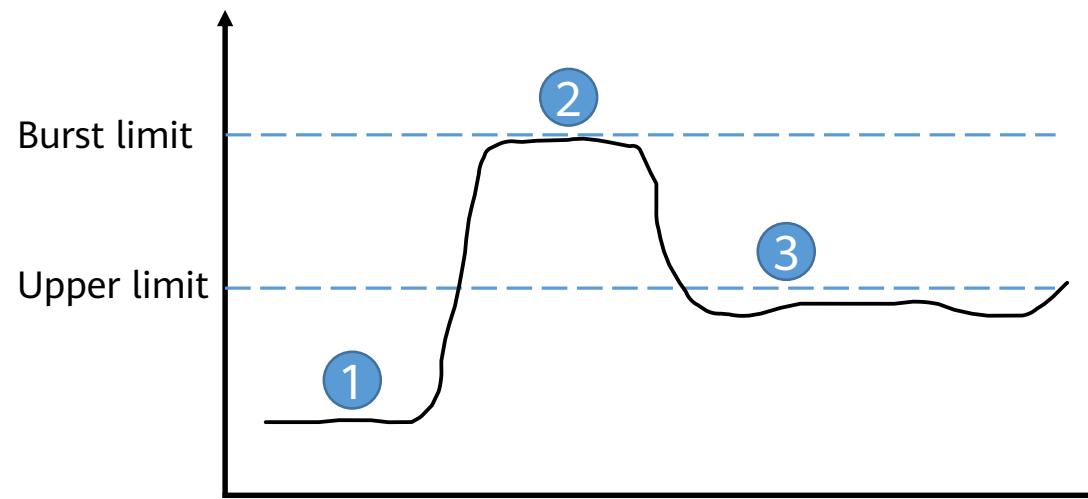
- The controlled objects of SmartQoS include **LUNs, LUN groups, hosts, and file systems.**
- By function, SmartQoS can be classified into traffic control management, burst traffic control management, and lower limit guarantee.
- Other characteristics of SmartQoS include multi-dimensional traffic control, hierarchical policies, and objective distribution.

Upper Limit Traffic Control Management



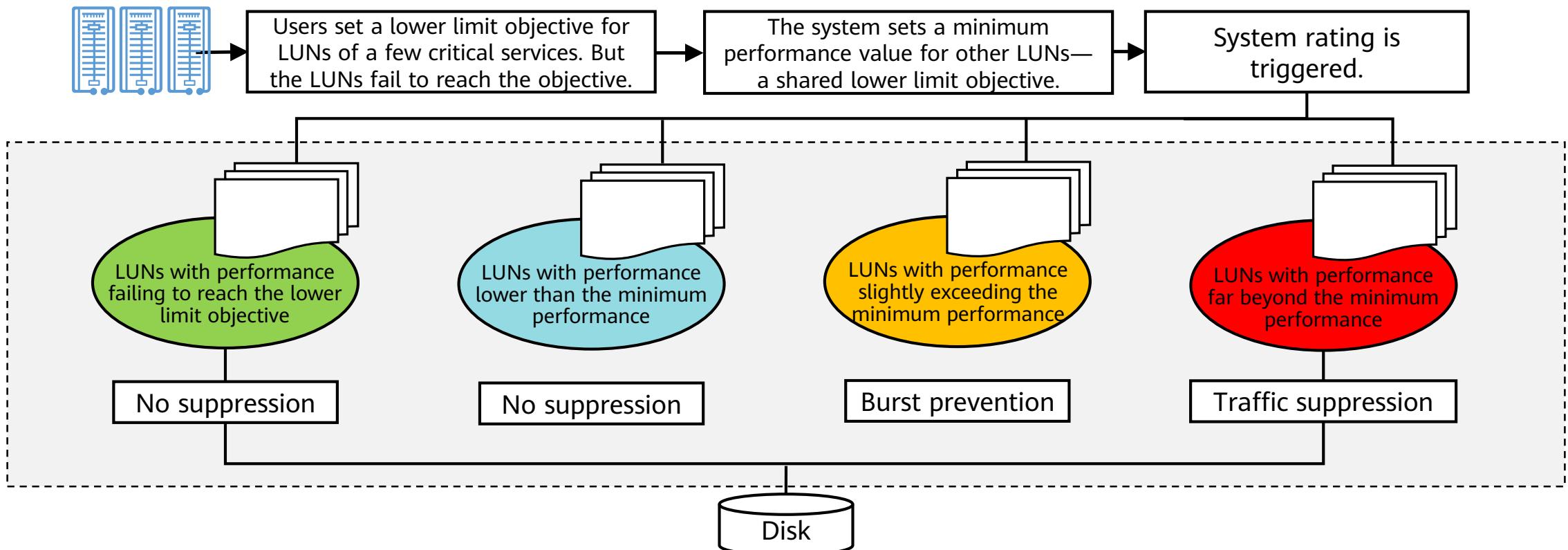
Burst Traffic Control Management

- When configuring the burst value, you must configure the upper limit. That means the instantaneous performance value can reach the burst value, but the average performance value cannot exceed the upper limit.
- The service load in phase 1 is lower than the upper limit, and the burst capabilities are accumulated.
- In phase 2, when the service load increases, the performance can exceed the upper limit and reach the burst threshold. The total burst duration is the smaller value between the accumulated burst duration in phase 1 and the burst duration configured in the QoS policy.
- In phase 3, once the burst ends, the service performance falls below the upper limit.



Lower Limit Guarantee

- When multiple services preempt resources and a small number of critical services have no advantages in resource preemption, the lower limit guarantee function can automatically suppress non-critical services to reserve resources for critical services.



Multi-dimensional Traffic Control

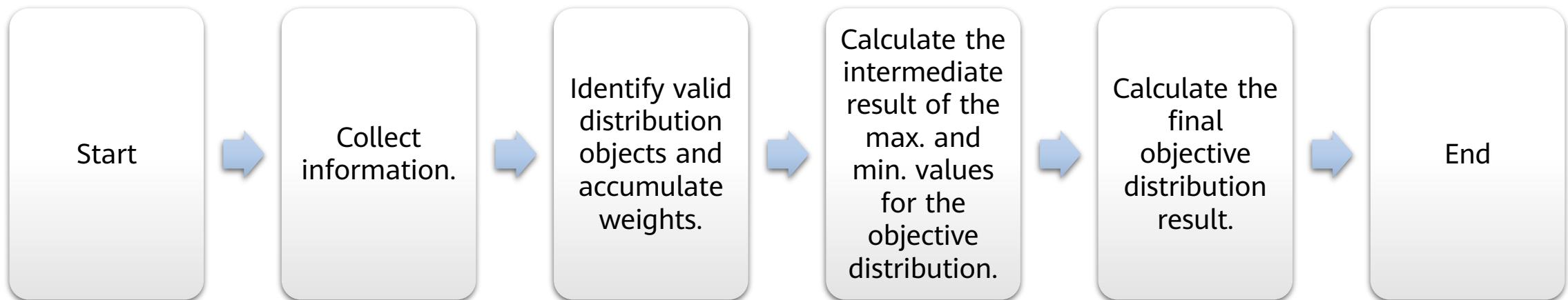
- SmartQoS policies provide multi-dimensional processes to simplify user operations:
 - Traffic control for LUNs
 - Traffic control for LUN groups
 - Traffic control for hosts
 - Traffic control for file systems
- Notes:
 - Only one type of objects can be added to each policy.
 - Only one LUN group can be added to each policy.
 - A LUN, LUN group, host, or file system can be added to only one policy.
 - A LUN can be added to different LUN groups, and different LUN groups can be added to different policies. In this case, the final traffic control value for the LUN is set to the smallest value to ensure that all policies are implemented.
 - Host policies do not support the lower limit guarantee function.

Hierarchical Management

- Hierarchical policy: System capabilities are classified by user, vStore, or subsidiary to prevent interference between different dimensions. Multiple common policies can be added to a hierarchical policy. The total performance of all sub-policies meets the requirements of the hierarchical policy.
- Common policy: Traffic control policies are configured for different services of a single user, vStore, or subsidiary.
- Notes:
 - A hierarchical policy has all the features of a common policy, and a common policy can be added as a sub-policy.
 - Sub-policies with LUNs and LUN groups can be added to the same hierarchical policy.
 - The policy configuration for a LUN/LUN group, a host, and a file system are mutually exclusive. Therefore, they cannot be added to the same hierarchical or common policy.
 - A lower limit guarantee policy cannot be added to a hierarchical policy.

Objective Distribution

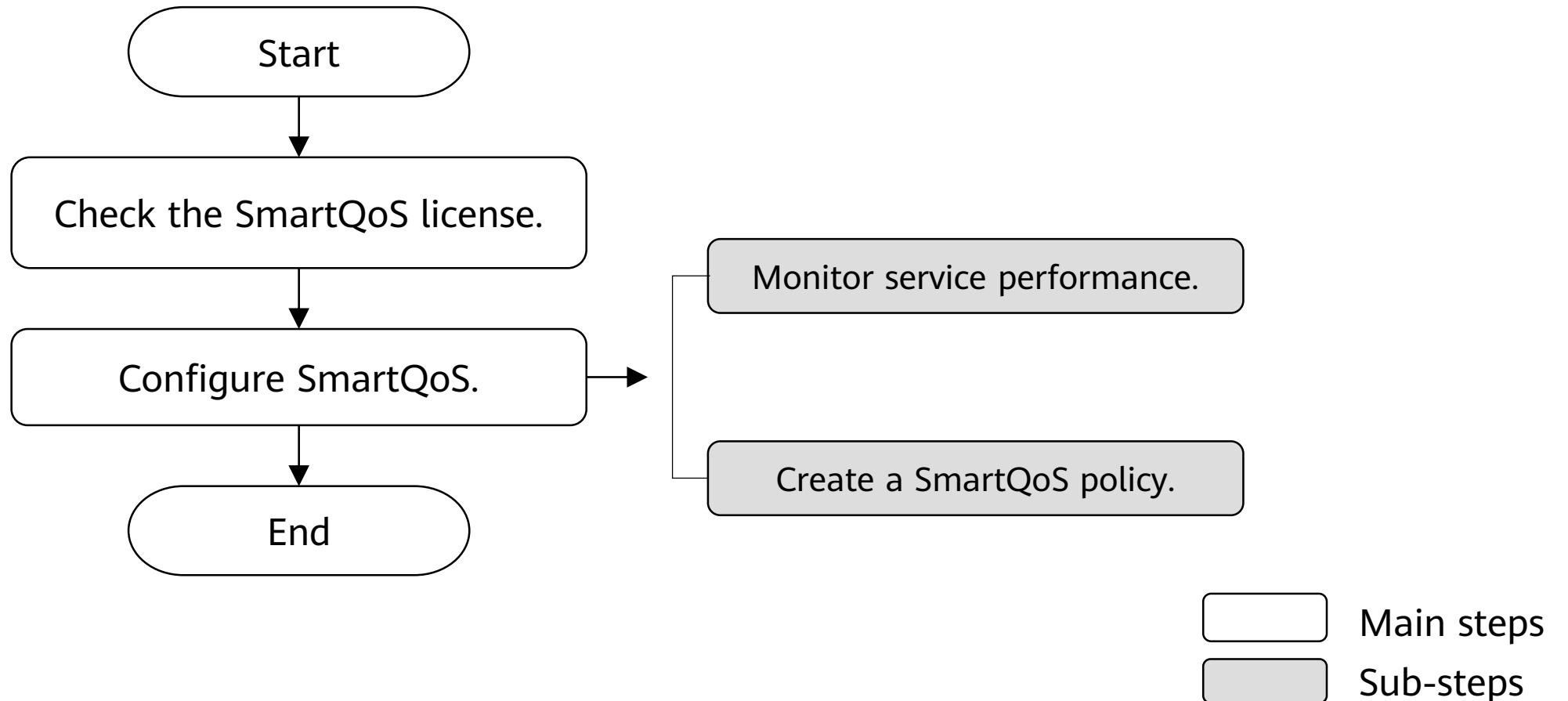
- All objects in a SmartQoS policy share the upper limit objectives. The SmartQoS module periodically collects the performance and requirement statistics of all controlled objects in a traffic control policy, and distributes the traffic control objective to controlled objects.
- Currently, the max-min weight allocation algorithm is used. The figure below shows the overall objective distribution process.



SmartQoS Application Scenarios

- Upper limit traffic control management:
 - Suitable to scenarios with mixed services to prevent services from affecting each other. It can limit the resource usage of low-priority services that have great impacts on the system.
 - Suitable for scenarios where performance restriction is required. For example, when an online service and a backup service coexist, SmartQoS restricts the maximum traffic of the backup service to ensure performance of the online service while ensuring the backup window.
- Burst traffic control management:
 - Suitable for latency-sensitive services, such as database services, VM deployment, and VM startup.
- Lower limit guarantee:
 - Ensures service quality (IOPS, BPS, and latency) for services that have a small quantity but are the most critical.

Configuration Process



Contents

- SmartThin
- SmartTier&SmartCache
- SmartAcceleration
- SmartQoS
- **SmartDedupe&SmartCompression**
- SmartVirtualization
- SmartMigration

SmartDedupe

- SmartDedupe is a data deduplication technology developed by Huawei for saving storage space. It deletes duplicate data in a storage system by searching for duplicate data blocks (generally greater than 4 KB) and saves only one copy of these blocks in the storage system. This technology is widely used in network disks, emails, and backup media devices.
- **Type of deduplication**
 - Inline deduplication: Data is deduplicated when being written to storage media.
 - Post-process deduplication: Data is first written to persistent storage media and then read for deduplication.
 - Fixed-length deduplication: Data is divided into blocks of a fixed size for deduplication.
 - Variable-length deduplication: Data is divided into blocks of variable sizes based on its content for deduplication. This is mainly used in data backup.
 - Similarity-based deduplication: The system divides data into blocks of a fixed size and analyzes the similarity among the blocks. The system deduplicates the identical data blocks and performs combined or delta compression on the similar data blocks.

SmartCompression

- SmartCompression is a data compression technology developed by Huawei. It is a process of encoding information using fewer bits than the original representation.
- **Type of compression**
 - Inline compression: Data is compressed when being written to storage media.
 - Post-process compression: Data is first written to storage media and then read for compression.
 - Software compression: Uses the CPU of a system to perform a compression algorithm.
 - Hardware compression: The compression algorithm logic is fixed to the hardware device, for example, the acceleration engine, FPGA, or ASIC of the Arm CPU. The compression API provided by the hardware device is invoked when the compression algorithm logic is used.
 - Lossy compression: A compression mode in which original data cannot be completely restored after decompression. This mode is commonly used to process audio, video, and images.
 - Lossless compression: A compression mode in which original data can be completely restored after decompression.

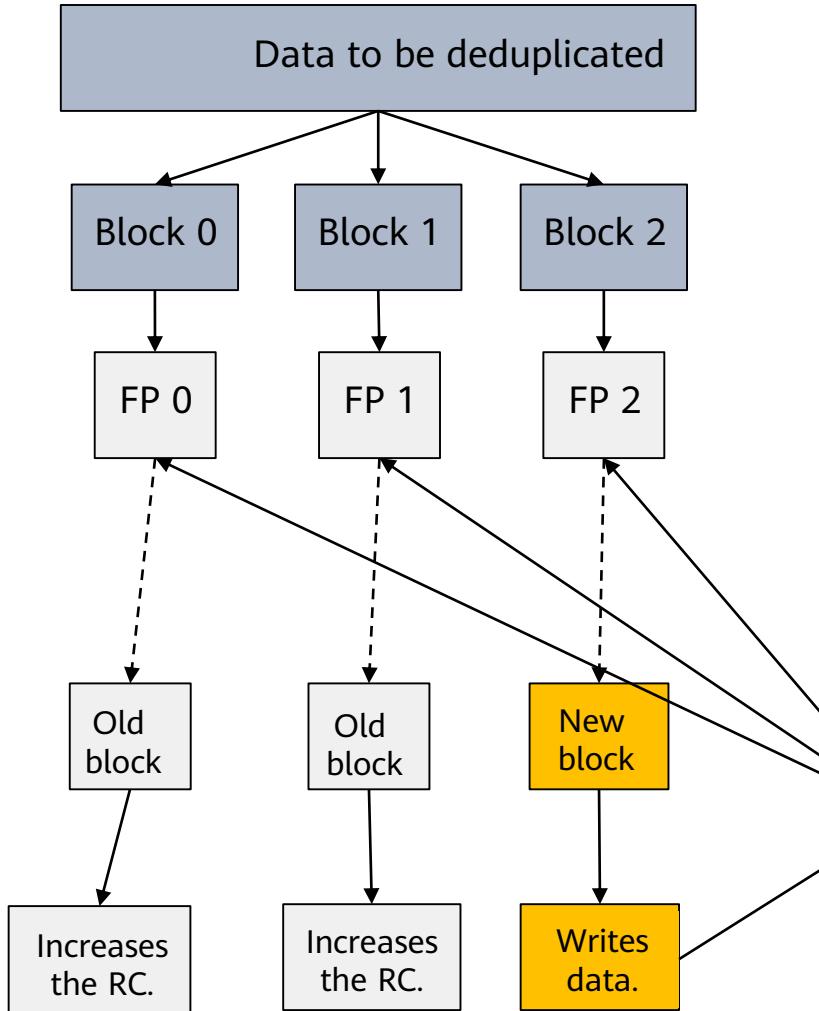
Working Principles of Inline Deduplication

1. Divides data into blocks.

2. Calculates the fingerprints of the data blocks.

3. Checks whether the fingerprints exist in the fingerprint table.

4. If a fingerprint exists in the fingerprint table, this block already exists in the system. If a fingerprint is not found in the fingerprint table, the block is new.

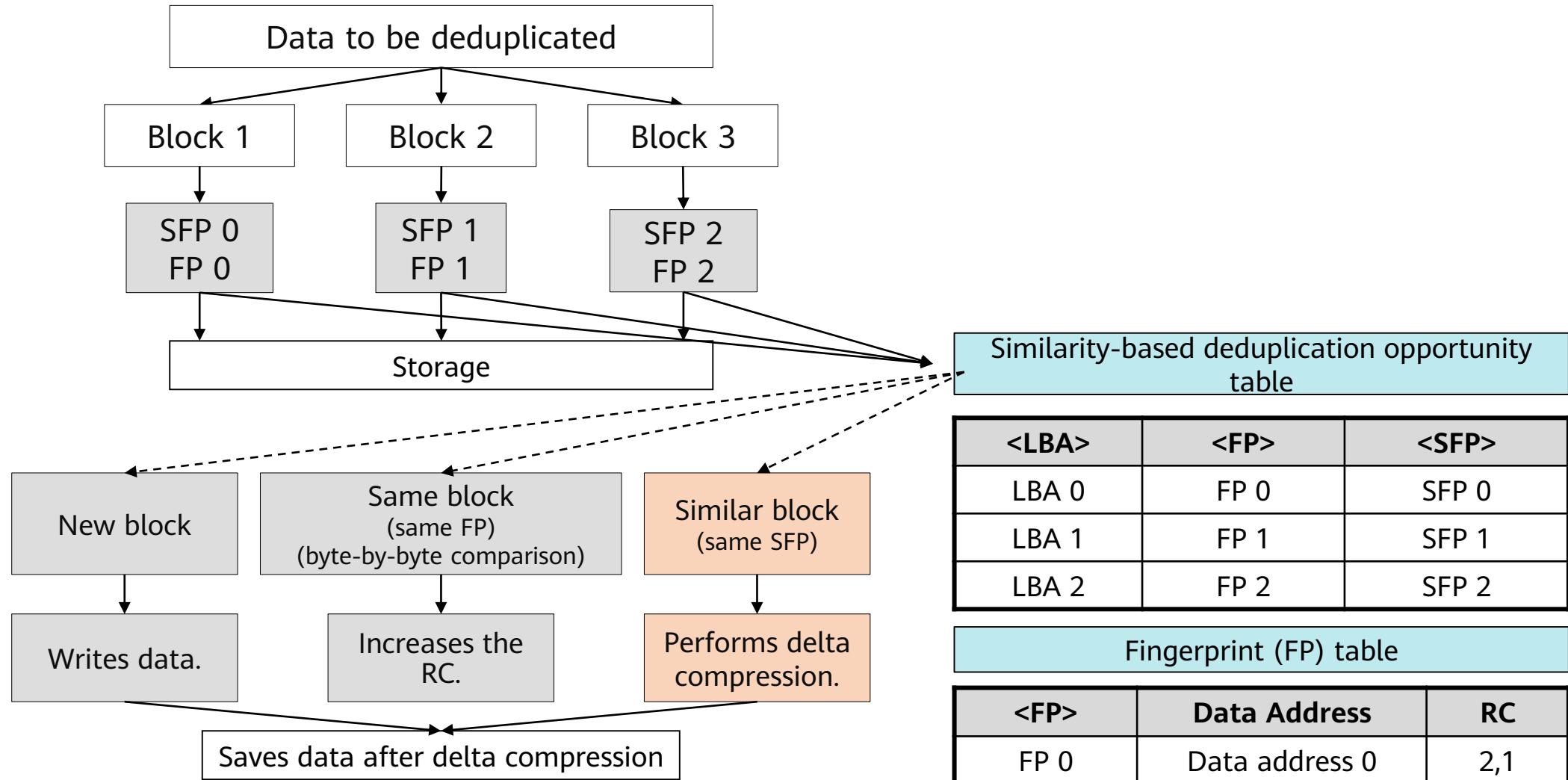


5. For existing blocks, increases the reference count (RC) and returns the addresses of the existing blocks. For new blocks, writes them to the storage space.

6. Adds the fingerprints and storage locations of the new blocks to the fingerprint table.

Fingerprint (FP) table		
FP	Data Address	RC
FP 0 = FP x	Data address 0	+1
FP 1 = FP y	Data address 1	+1
FP 2	Data address 2	1
FP x	Data address x	1->2
FP y	Data address y	1->2

Working Principles of Post-processing Similarity-based Deduplication

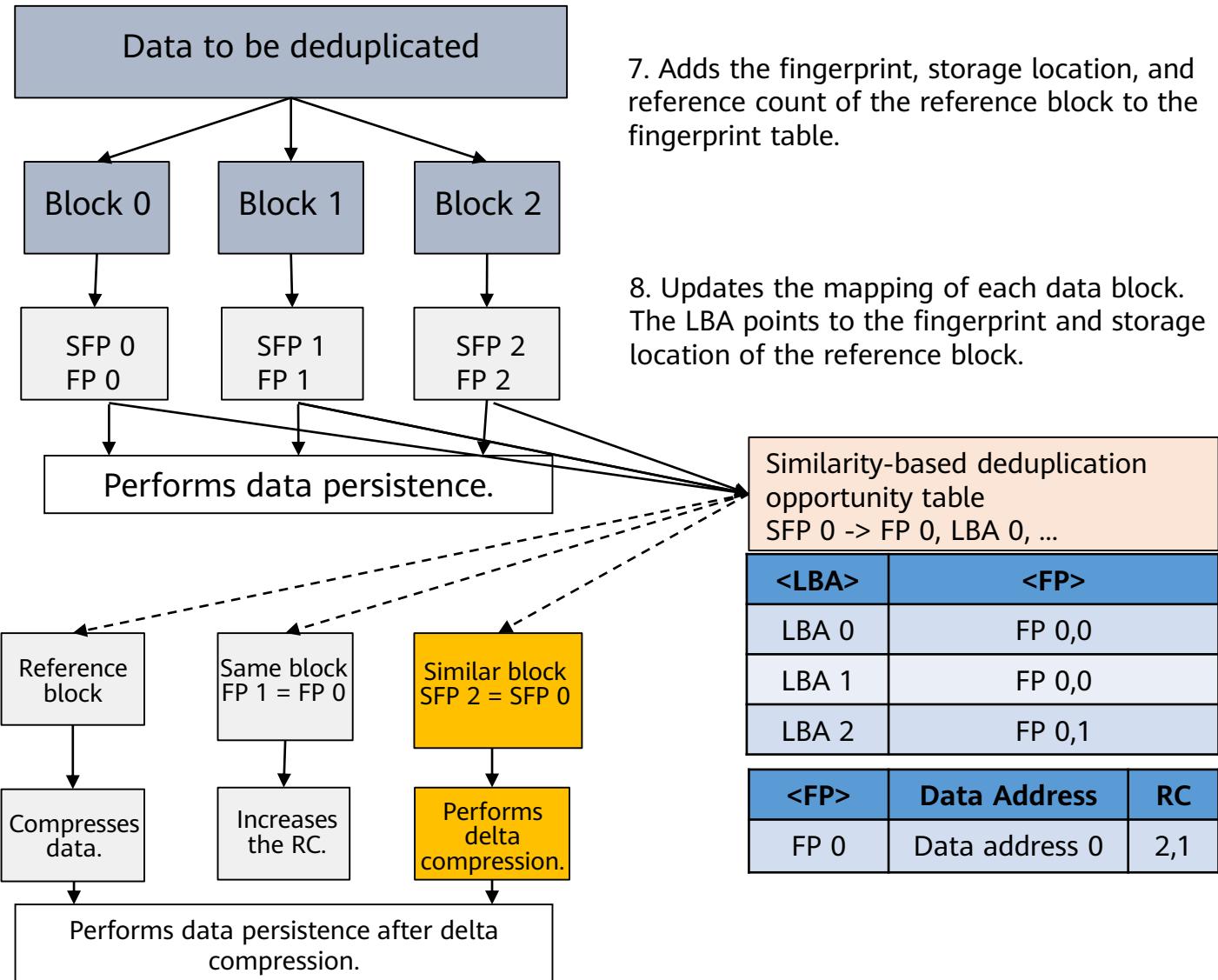


<FP>	Data Address	RC
FP 0	Data address 0	2,1

Fingerprint (FP) table

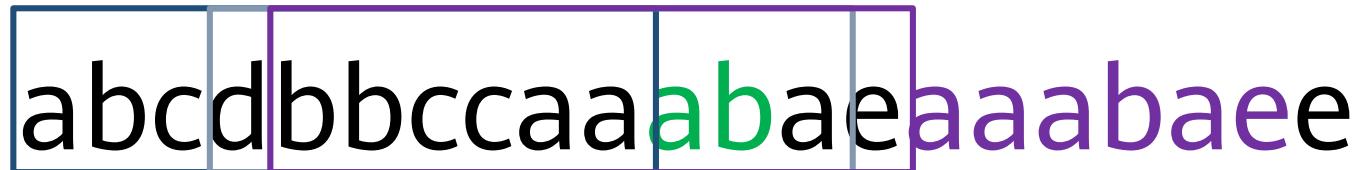
Working Principles of Similarity-based Deduplication in the Background

1. Divides data into blocks.
2. Calculates the fingerprints and similar fingerprints (SFPs) of the data blocks and adds them to the similarity-based deduplication opportunity table.
3. Analyzes the similarity-based deduplication opportunity table in the background and performs similarity-based deduplication on data blocks with the same SFP.
4. Selects a reference block and compares other blocks with the reference block.
5. If the fingerprint of another block is the same as that of the reference block and the two blocks are consistent byte by byte, the block is duplicate. Increases the reference count of the reference block.
6. If another block is different from the reference block, but they have the same SFP, performs delta compression on this block. After delta compression, writes the data to persistent storage media.



Working Principles of SmartCompression

(LZ77 example)

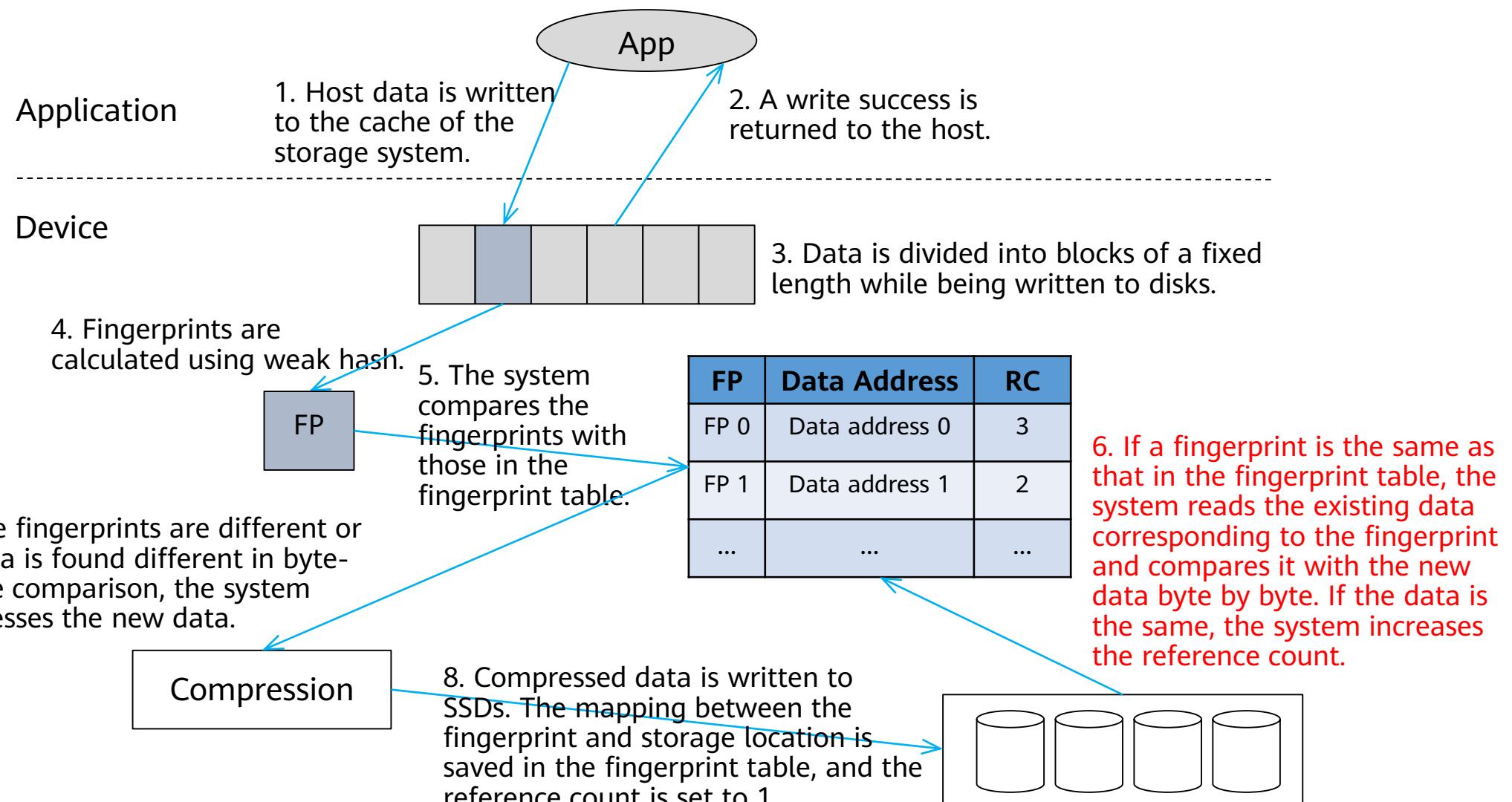


- Assume that a string **abcdbbbccaaabaaeaaabaaee** is to be compressed, the first 10 characters have been compressed, and the window is 10 characters.

Window	Uncompressed	Identical String	Start Position of the Identical String	Length of the Identical String	Last Character in the Identical String	Compression Code
abcdbbbccaa	abaaeaaabaaee	ab	0	2	a	(0,2,a)
bbbccaaaba	eaaabaaee	null	0	0	e	(0,0,e)
bbccaaabaae	aaabaaee	aaabae	4	6	e	(4,6,e)

- The compression result is **(0,2,a)(0,0,e)(4,6,e)**.

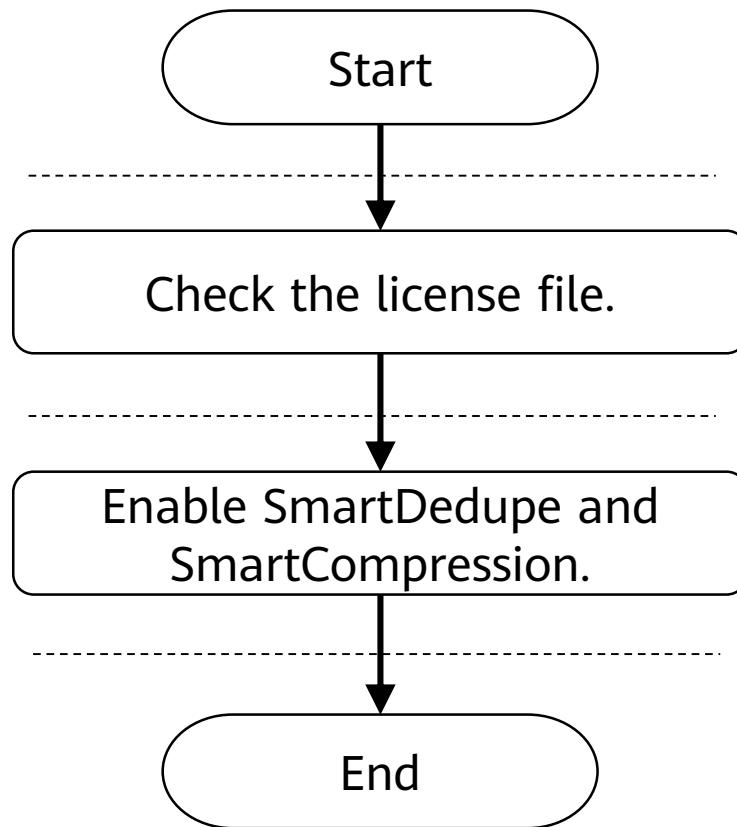
SmartDedupe and SmartCompression Process



Scenarios Where SmartDedupe and SmartCompression Are Used Together

- Using both SmartDedupe and SmartCompression can achieve the best space saving effect.
- The scenarios include:
 - VDI and VSI scenarios
 - Data testing or development systems
 - Storage systems with the file service enabled
 - Engineering data systems

Configuration Process



Contents

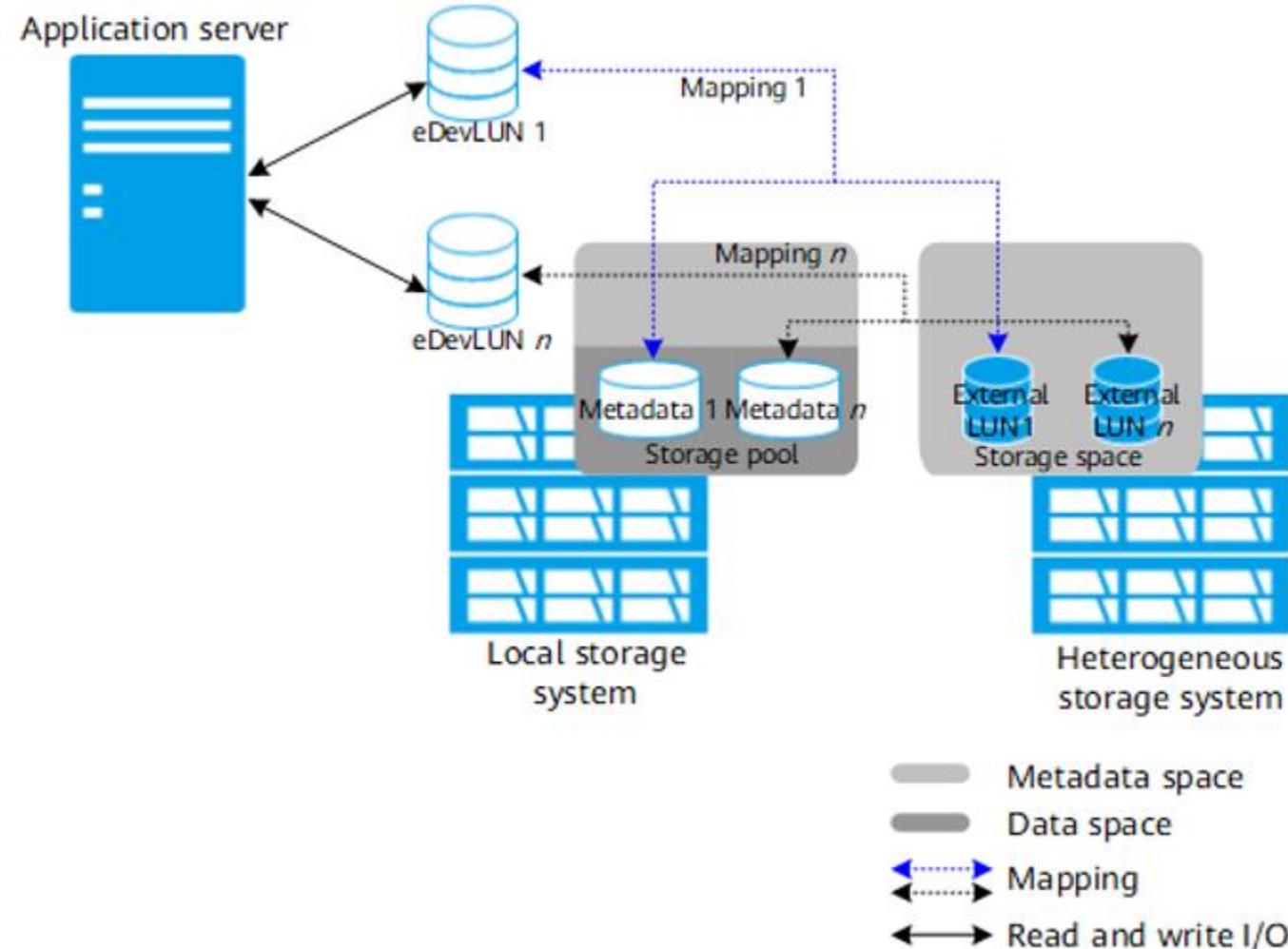
- SmartThin
- SmartTier&SmartCache
- SmartAcceleration
- SmartQoS
- SmartDedupe&SmartCompression
- **SmartVirtualization**
- SmartMigration

Concepts Related to SmartVirtualization

- Question: SmartVirtualization allows external LUNs to provide physical storage space for OceanStor Dorado series storage systems. How can it be implemented?
- What do "local storage system", "heterogeneous storage system", "external LUN", "eDevLUN", "online takeover", "offline takeover", and "hosting" mean?

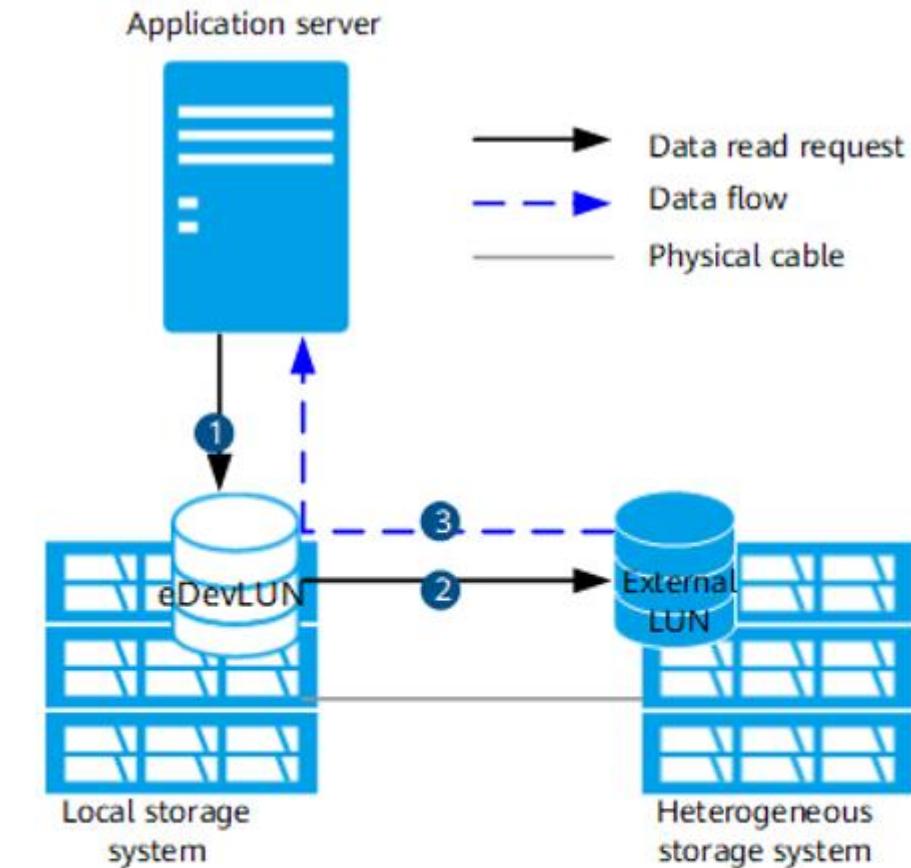
Relationship Between an eDevLUN and an External LUN

- An eDevLUN consists of data and metadata. A mapping is established between data and metadata.



Data Read Process

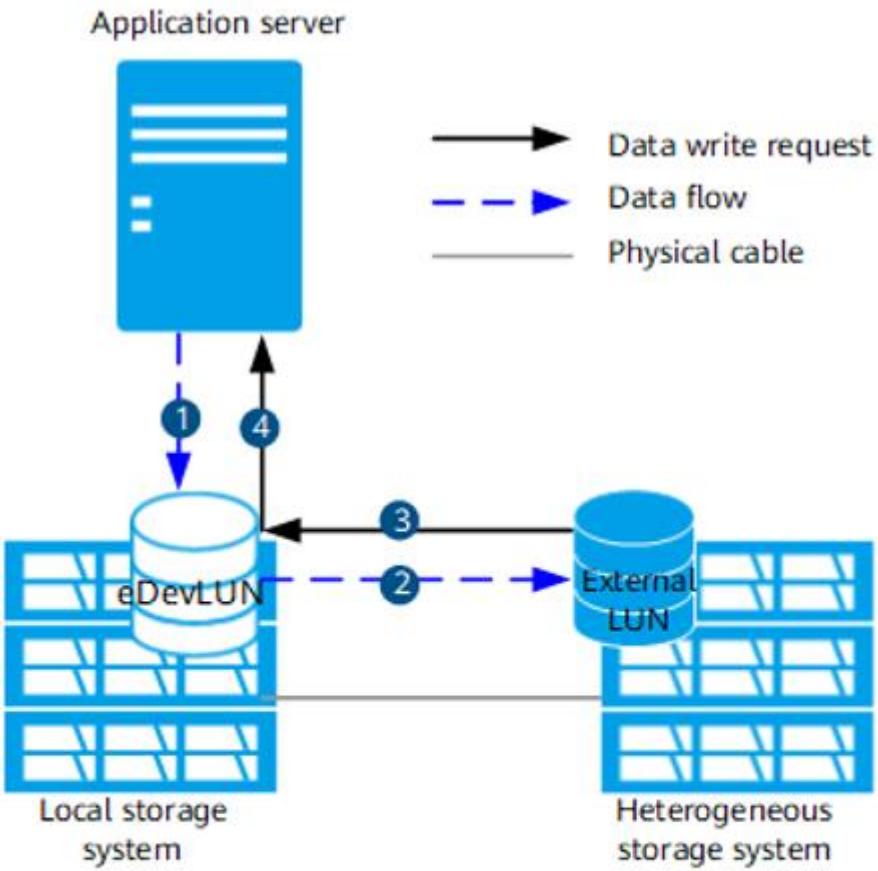
- After an external LUN in a heterogeneous storage system is hosted using SmartVirtualization, when an application server sends a request to read data from the external LUN, the eDevLUN in the local storage system receives the request and reads data from the external LUN



- 1 The application server sends a data read request.
- 2 The local storage system receives the request and reads data from the external LUN.
- 3 Data is returned to the local storage system and then the application server.

Data Write Process

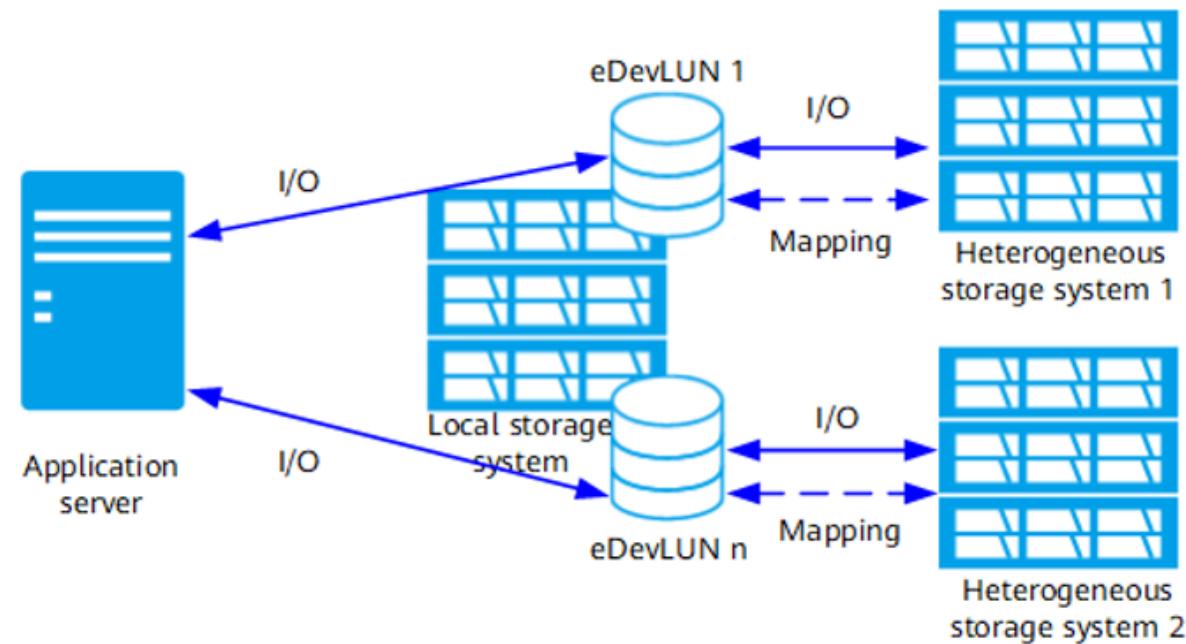
- After an external LUN in a heterogeneous storage system is hosted using SmartVirtualization, the data write process is as follows:



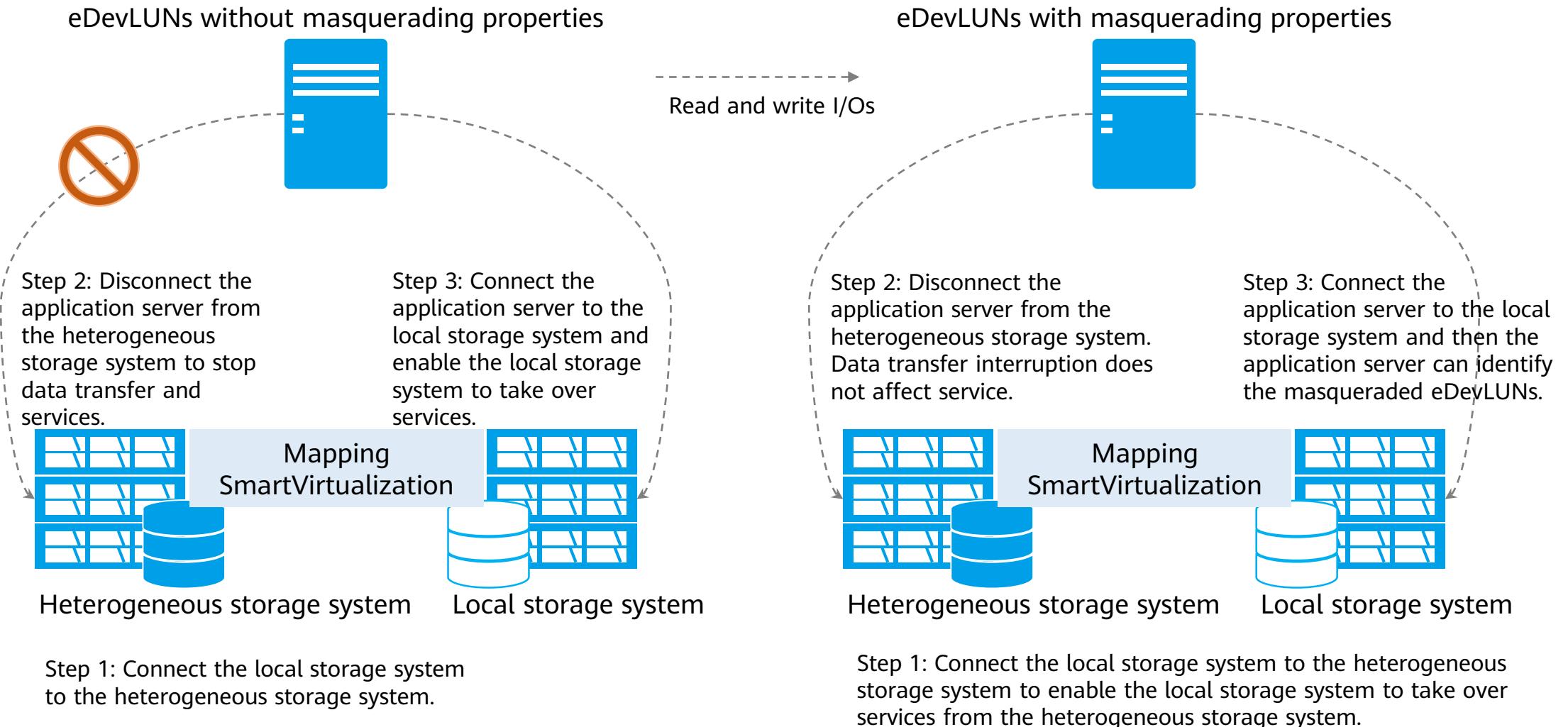
- 1 The application server writes data to the local storage system.
- 2 The local storage system writes the data to the heterogeneous storage system.
- 3 The heterogeneous storage system sends the write success message to the local storage system.
- 4 The local storage system sends the write success message to the application server.

Centralized Management of Storage Resources

- If multiple heterogeneous storage systems have been deployed onsite, the following two challenges may occur:
 - Due to incompatibility issues, the multipathing software on an application server may not be compatible with all heterogeneous storage systems.
 - In a certain network environment (such as a Fibre Channel direct-connection network), one application server can only be connected to one storage system. However, in actual applications, one application server needs to distribute services to multiple storage systems.



Offline Takeover and Online Takeover



Service Data Migration from a Legacy Storage System to a New Storage System

- As services grow continuously, more storage is required for storing increasing data. The legacy storage system cannot provide satisfactory data storage capacity and performance. In this case, you can purchase a storage system that provides a larger capacity and better performance to replace the legacy storage system. As software and hardware of the legacy and new storage systems are different, the services may be interrupted and data may be lost during data migration. SmartVirtualization can mask the differences between storage systems to map an external LUN of the legacy storage system to the new storage system (presented as an eDevLUN on the new storage system). Then SmartMigration can be used to reliably migrate all service data from the legacy storage system to the new storage system while keeping services running.

New storage system



Legacy storage system



Cold Data Migration from a New Storage System to a Legacy Storage System

- After a legacy storage system is replaced with a new storage system, some data in the new storage system is rarely accessed, which is called cold data. If massive cold data is stored in the new storage system, the storage resource utilization of the storage system lowers down, causing a waste of storage space. To reduce operation expenditure (OPEX), SmartVirtualization can work with SmartMigration to migrate the cold data to the legacy heterogeneous storage system.

New storage system



Legacy storage system

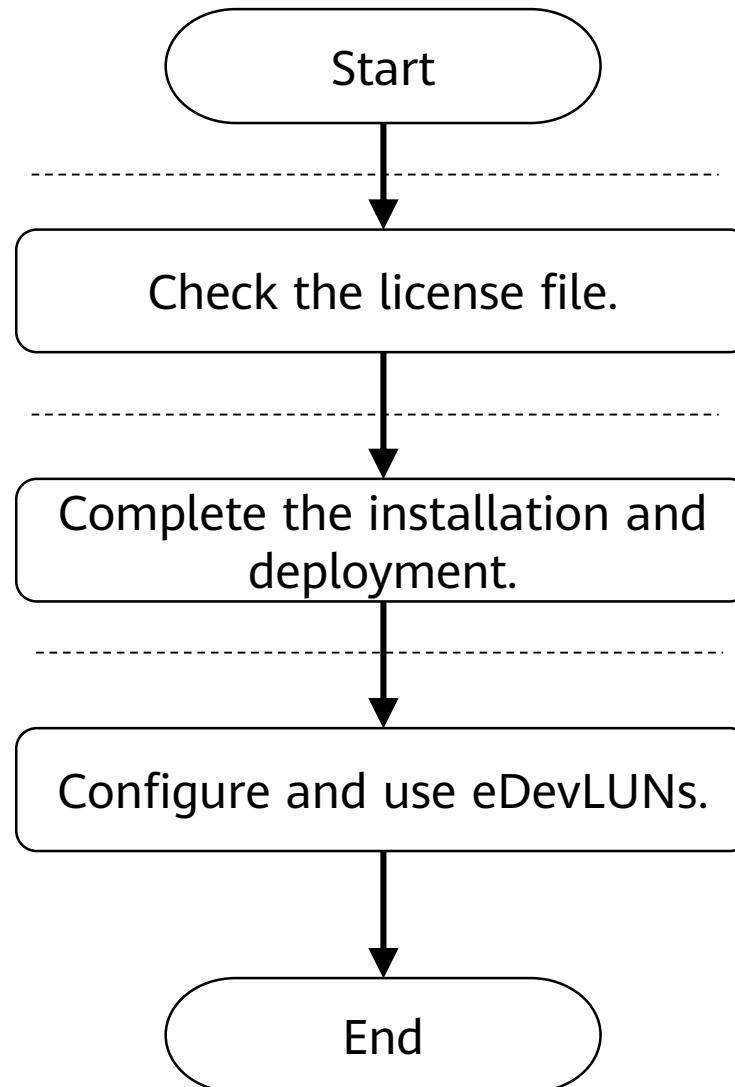


Takeover Mode Selection

- SmartVirtualization provides online and offline modes to take over heterogeneous storage systems. There are three masquerading types for online takeover, which are basic masquerading, extended masquerading, and third-party masquerading. The takeover mode depends on the vendors and versions of the heterogeneous storage systems and multipathing software.

Takeover Mode	Masquerading	Description
Offline takeover	No masquerading	The offline takeover mode is applicable to all compatible Huawei and third-party heterogeneous storage systems. In this mode, services running on the related application servers are stopped temporarily.
Online takeover	Basic masquerading or extended masquerading	The selection of basic masquerading or extended masquerading depends on the vendor and version of the multipathing software and the versions of Huawei heterogeneous storage systems.

Configuration Process

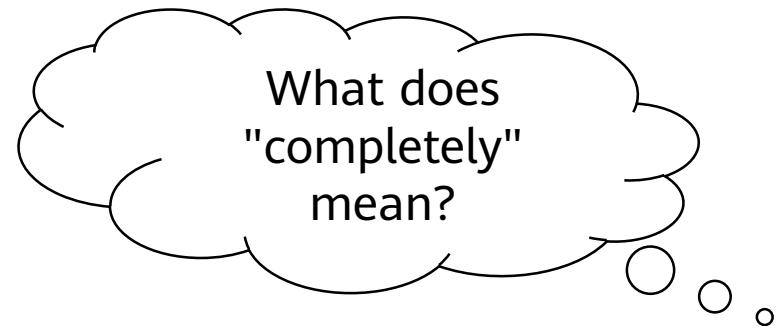


Contents

- SmartThin
- SmartTier&SmartCache
- SmartAcceleration
- SmartQoS
- SmartDedupe&SmartCompression
- SmartVirtualization
- **SmartMigration**

Overview

- SmartMigration is a key technology for service migration. Services on a source LUN can be completely migrated to a target LUN without interrupting host services. The target LUN can totally replace the source LUN to carry services after the replication is complete.



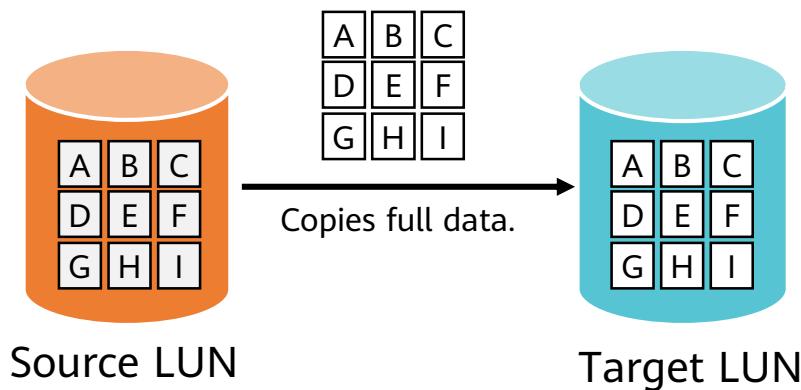
Working Principles of SmartMigration

- SmartMigration is leveraged to adjust service performance or upgrade storage systems by migrating services between LUNs.
- SmartMigration is implemented in two phases:

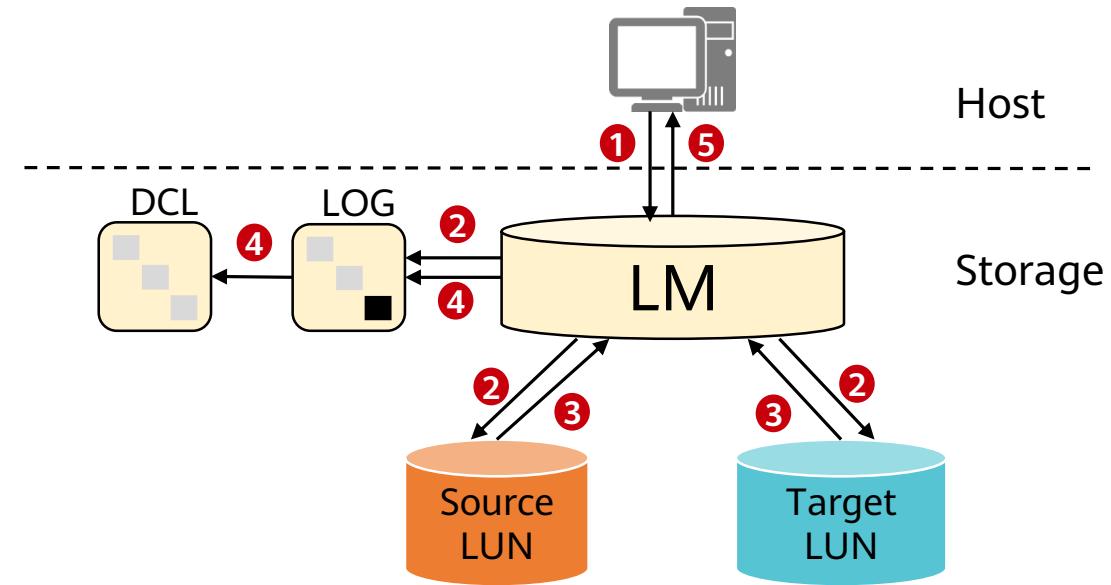


SmartMigration Service Data Synchronization

- After creating a SmartMigration task, create the pair relationship between a source LUN and a target LUN.
- Service data synchronization between the source and target LUNs involves initial synchronization and change synchronization.



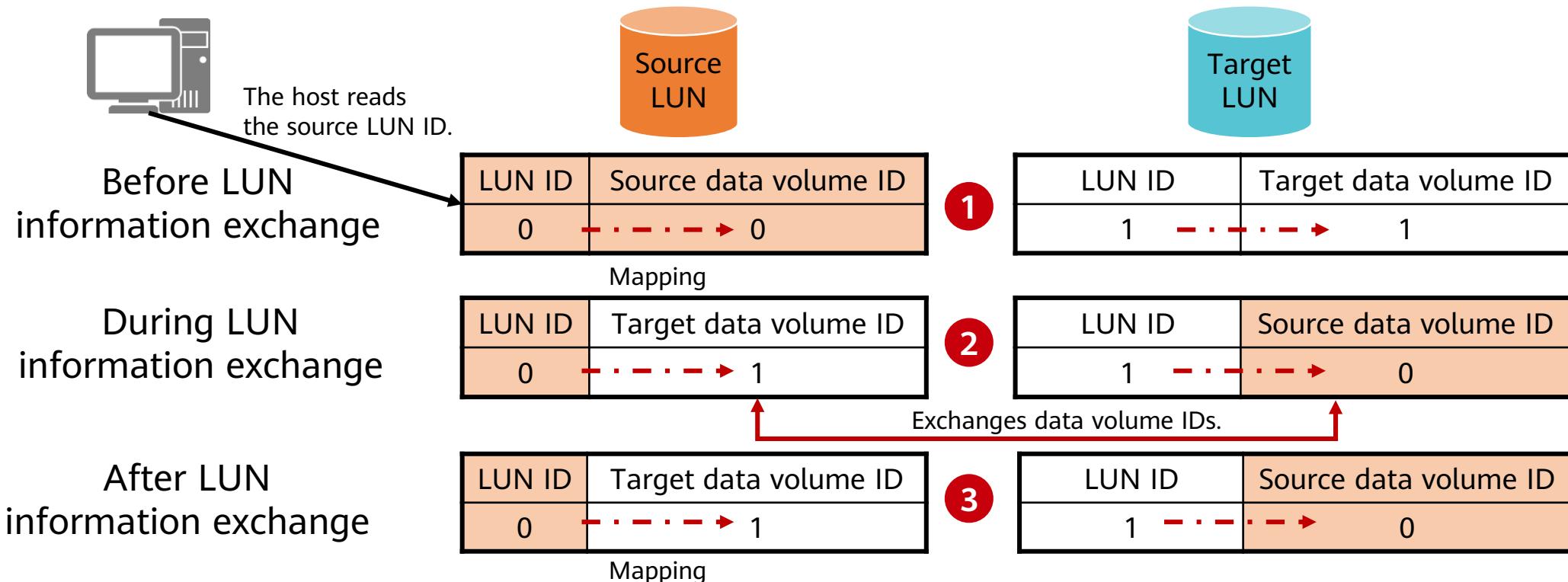
Initial synchronization



Data change synchronization

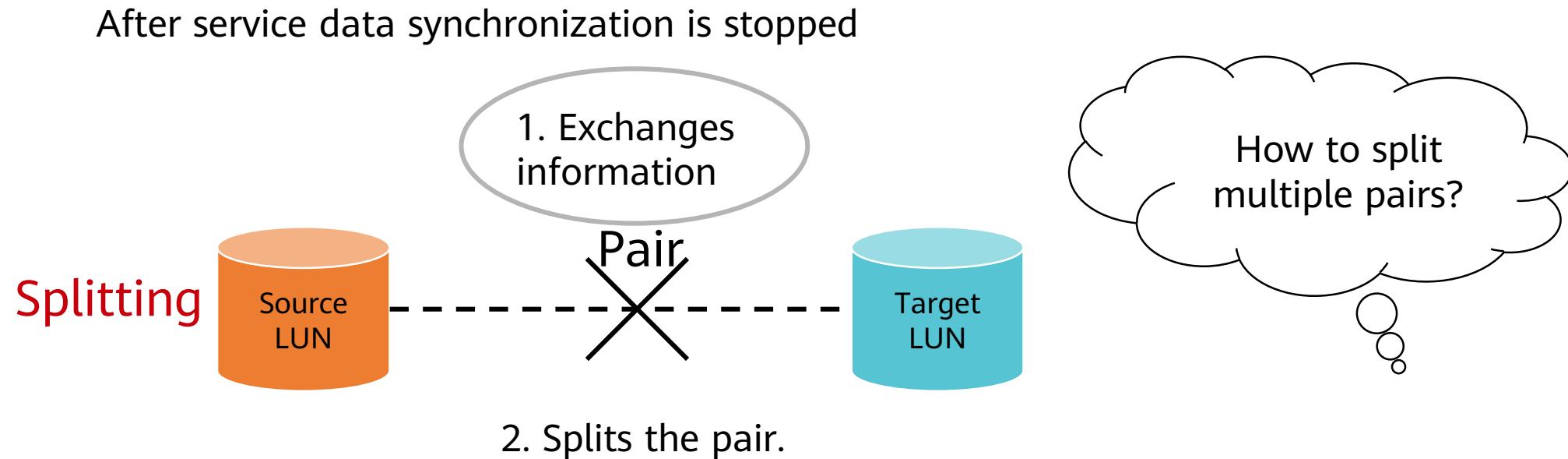
SmartMigration LUN Information Exchange

- LUN information exchange is used for mappings between LUNs and data volumes, namely, the exchange between both data volume IDs when the IDs of source LUNs and target LUNs remain unchanged.



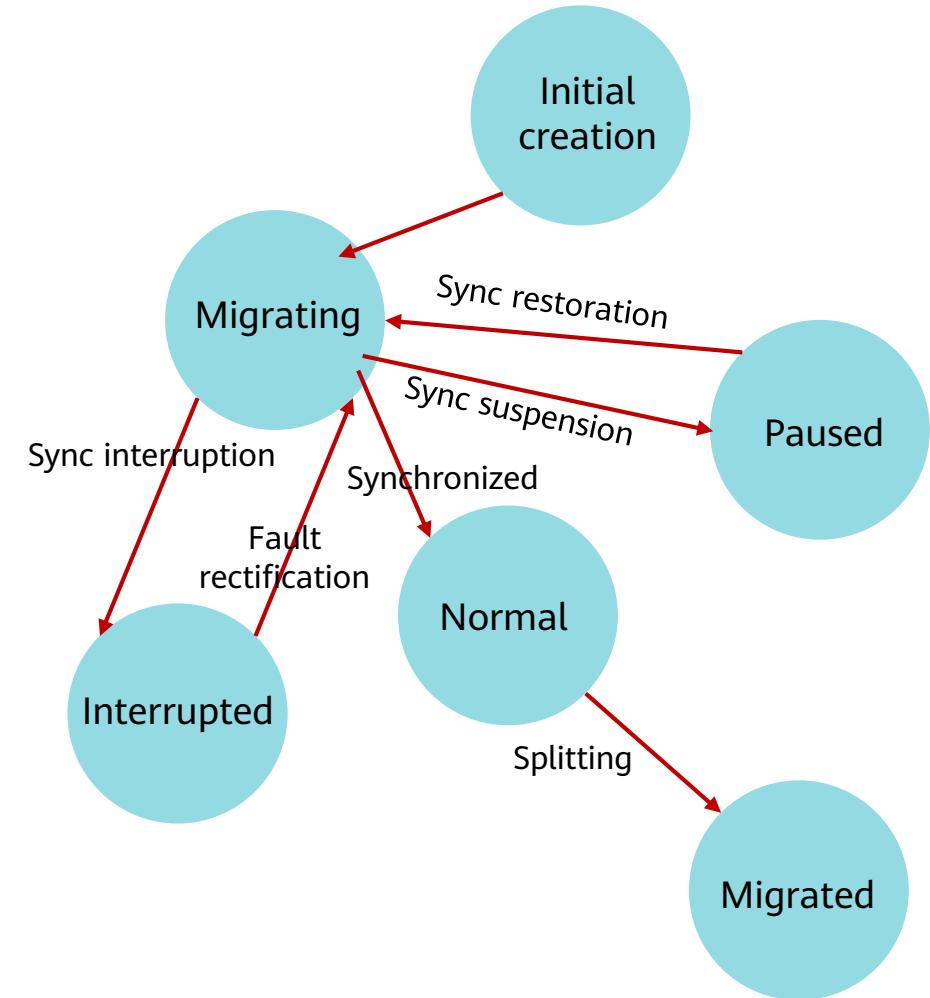
SmartMigration Pair Splitting

- Splitting is performed on a single pair. The splitting process includes stopping service data synchronization between the source LUN and target LUN in a pair to exchange LUN information, and removing the data migration relationship after the exchange.

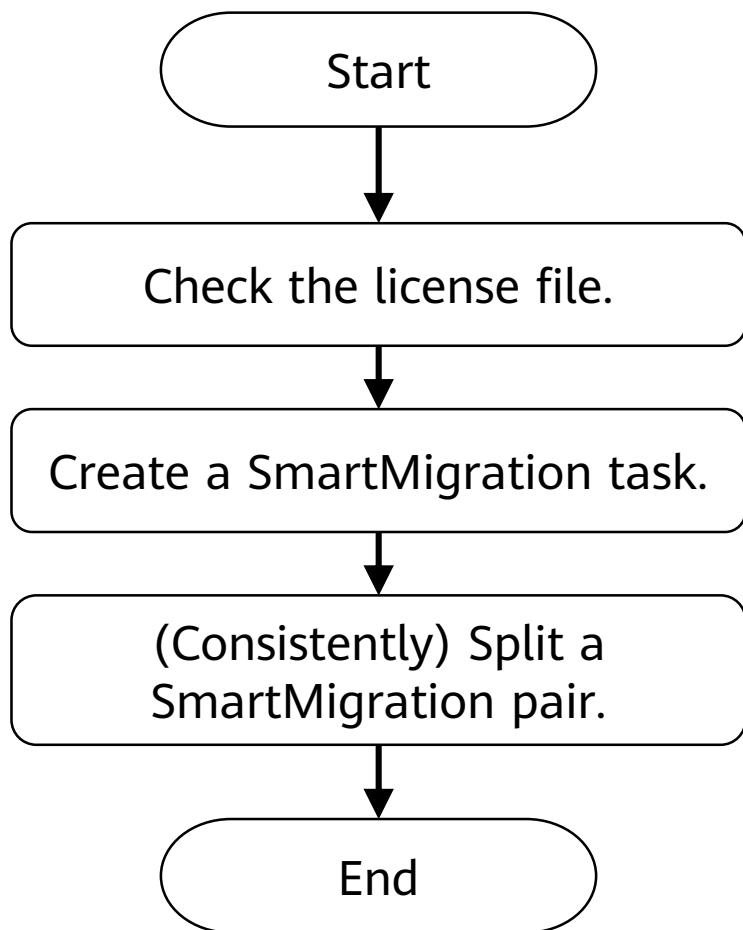


SmartMigration Status Transition

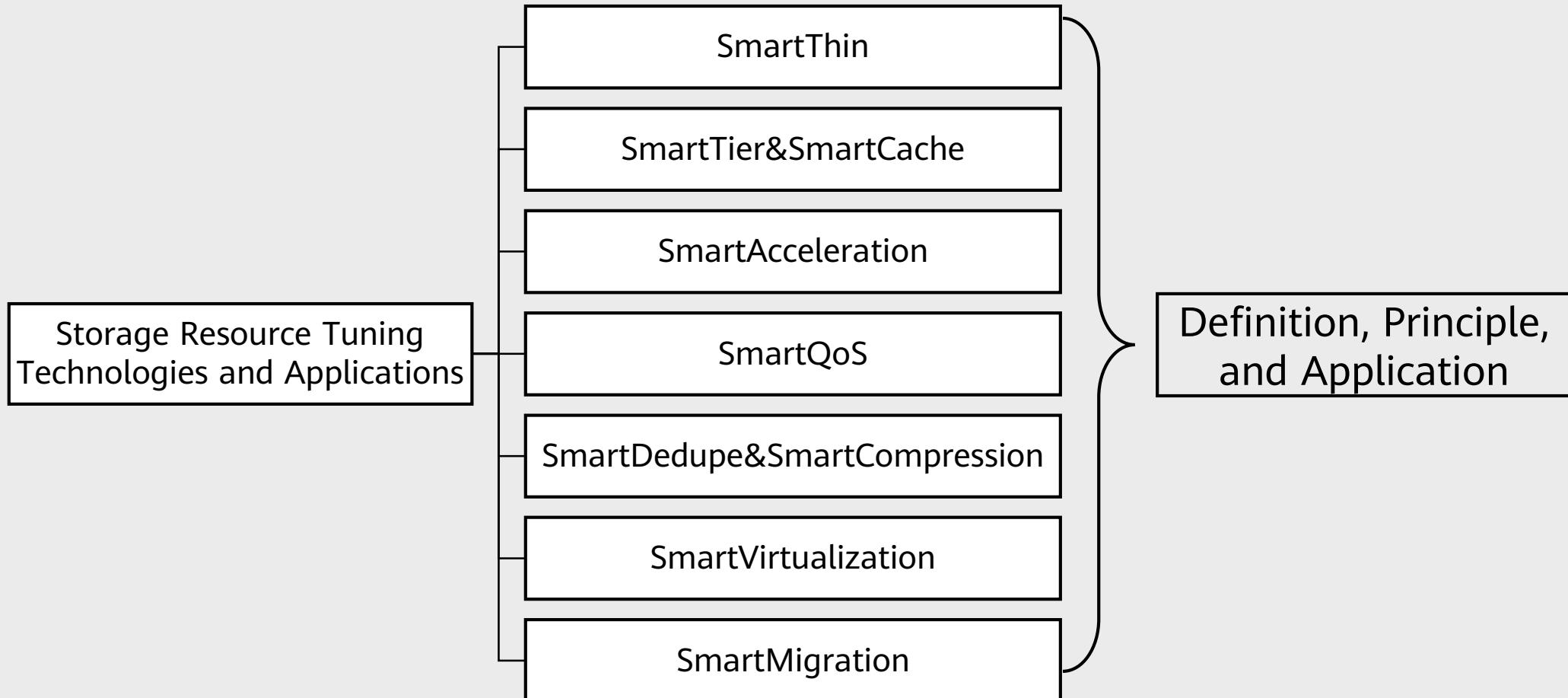
- **Migrating:** Data is being synchronized from the source LUN to the target LUN.
- **Normal:** Data is successfully synchronized between the source LUN and the target LUN.
- **Paused:** The pair is in the waiting queue.
- **Interrupted:** The replication relationship between the source LUN and target LUN is interrupted due to an I/O error in SmartMigration.
- **Migrated:** Data is successfully synchronized between the source LUN and the target LUN, and the splitting is complete.



Configuration Process



Summary



Quiz

- **(True or false)** SmartTier cannot be enabled for a storage pool whose member disks are of the same type. ()
- **(Multiple-choice)** Which of the following migration policies can be set for LUNs? ()
 - Automatic migration
 - Migration to the higher-performance tier
 - Migration to the lower-performance tier
 - No migration

Quiz

3. (Single-answer question) Which status must a pair be before consistency splitting during LUN migration? ()

- Migrating
- Paused
- Normal
- Migrated

Recommendations

- Huawei official websites
 - Enterprise business: <https://enterprise.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/>
 - Online learning: <https://learning.huawei.com/en/>
- Popular tools
 - HedEx Lite
 - Network Documentation Tool Center
 - Information Query Assistant

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.



Storage Data Protection Technologies and Applications



Foreword

- Traditional data protection solutions focus on periodic data backup. Therefore, problems such as no backup window, inconsistent data, and impact on the production system always occur.
- This course describes storage data protection technologies such as HyperSnap, HyperClone, HyperCDP, and HyperLock.

Objectives

On completion of this course, you will be able to understand the principles, configuration methods, and application scenarios of the following features:

- HyperSnap
- HyperClone
- HyperCDP
- HyperLock

Contents

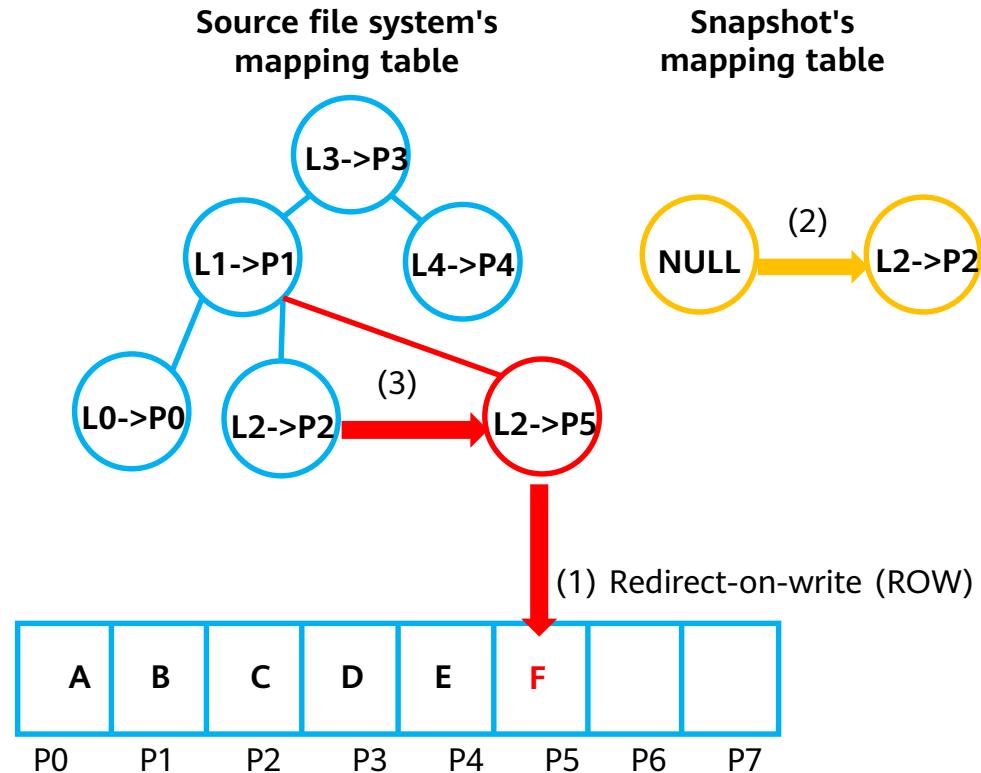
- 1. HyperSnap**
2. HyperClone
3. HyperCDP
4. HyperLock

Overview



- HyperSnap:
It is a snapshot feature. A snapshot generated by HyperSnap is a point-in-time, consistent, and fully usable copy of source data. It is a static image of the source data at the copy point in time.
- A snapshot can be implemented using the **copy-on-write (COW)** or **redirect-on-write (ROW)** technology.
 - COW enables data to be copied in the initial data write process. Data copy affects write performance of hosts.
 - ROW does not copy data. However, after data is overwritten frequently, data distribution on the source LUN will be damaged, adversely affecting sequential read performance of hosts.

ROW Principle



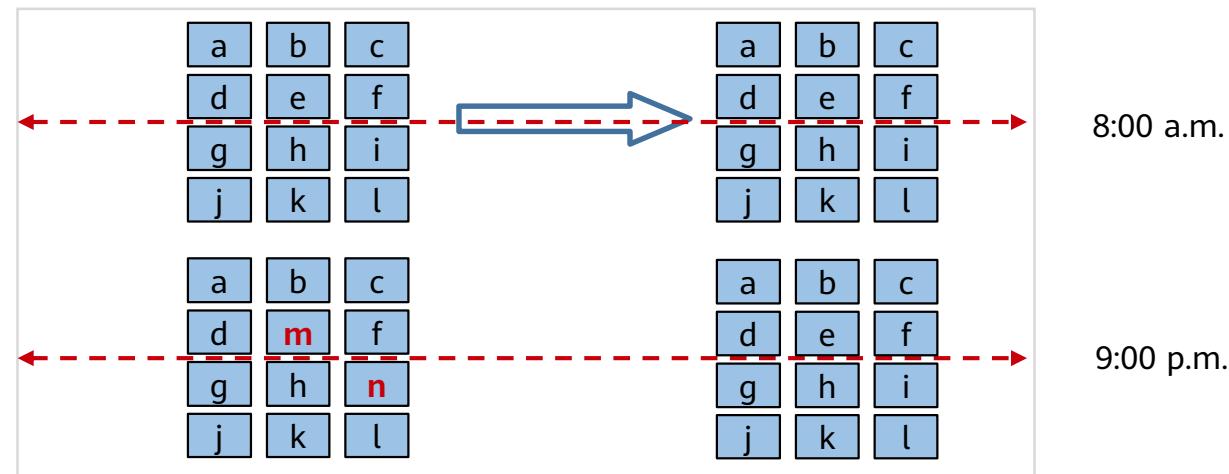
Working Principles of HyperSnap

- **Definition**

- A snapshot is a consistent copy of the source data at a certain point in time. After the snapshot is generated, it can be read by hosts and used as a data backup at a certain point in time.

- **Main features**

- Instant generation: A storage system can generate a snapshot within a few seconds to obtain the consistent copy of source data.
- Small storage space occupation: A snapshot is not a full physical data copy, which does not occupy large storage space. Therefore, a snapshot for a large amount of source data occupies only a small space.



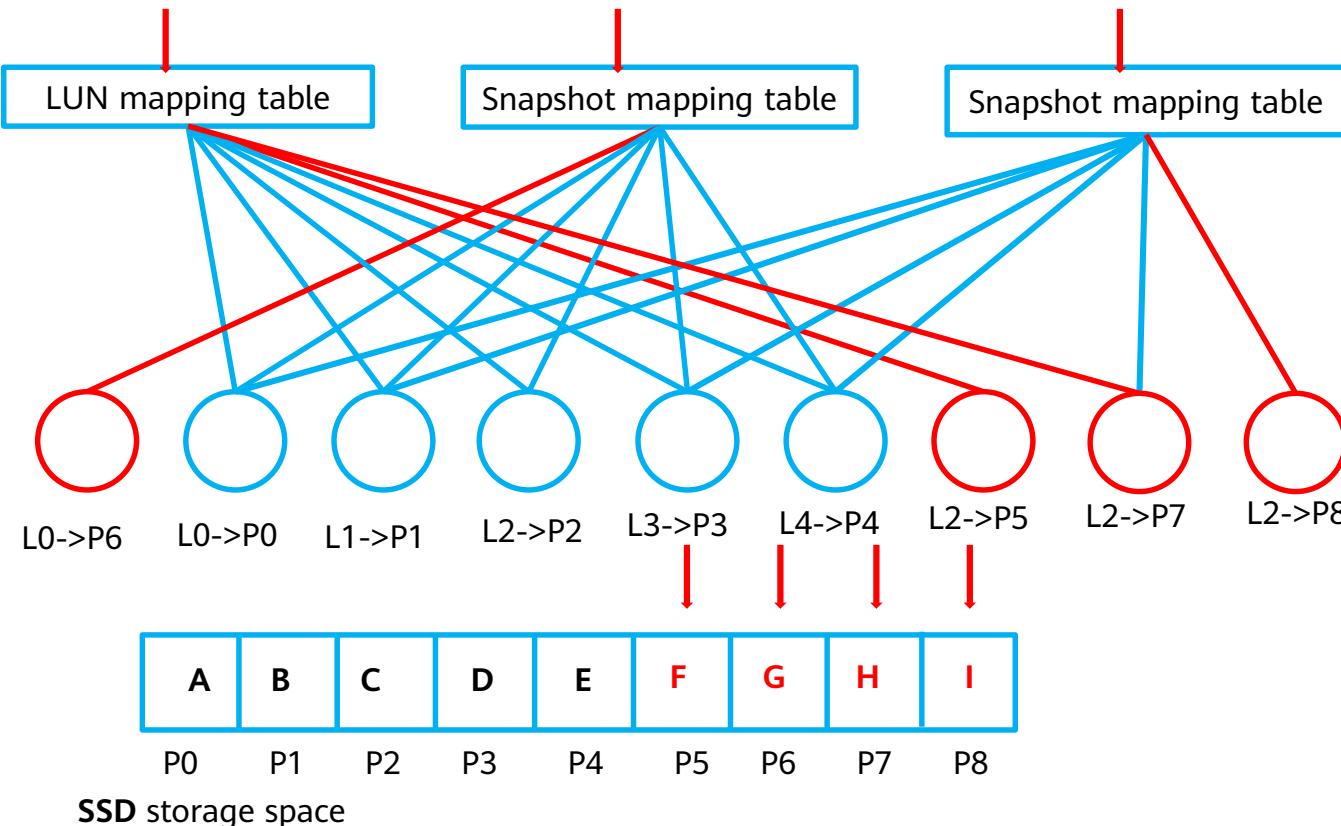
LUN Snapshot Principles: Zero Performance Loss

Data requested to be written to L2 of the source LUN is written to P5.

Data requested to be written to L2 of the source LUN is again written to P7.

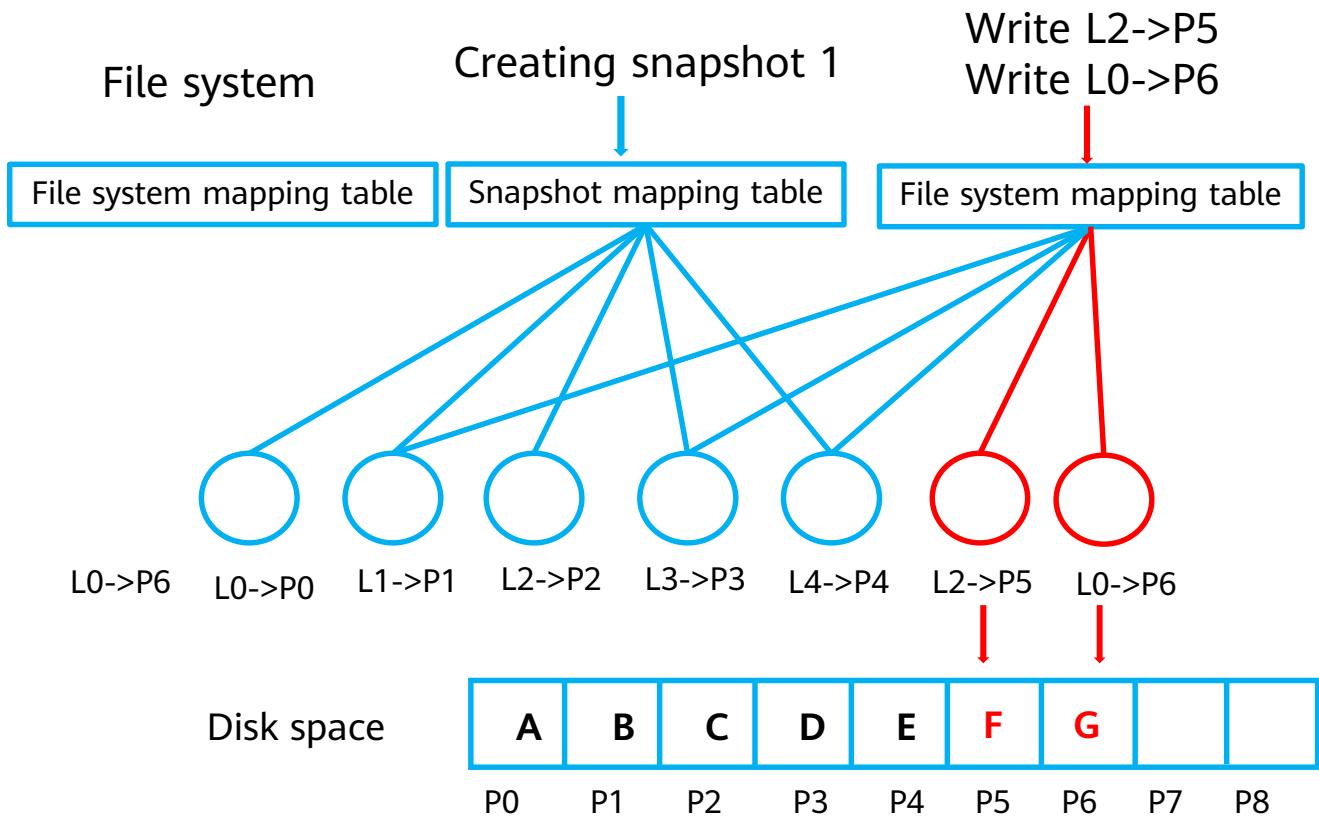
Data requested to be written to L0 of snapshot 1 is written to P6.

Data requested to be written to L2 of snapshot 2 is written to P8.



- Data requested to be written to L2 of the source LUN is written to a new space P5. The original space P2 is referenced by the snapshot.
- Data requested to be written to L0 of snapshot 1 is written to the new space P6, bringing no additional read and write overhead.
- When data is written to L2 of the source LUN again, the requested data is written to a new space P7. The original space P5 is released because it is not referenced by a snapshot.
- A new snapshot 2 is created and activated.

File System Snapshot Principles: Zero Performance Loss



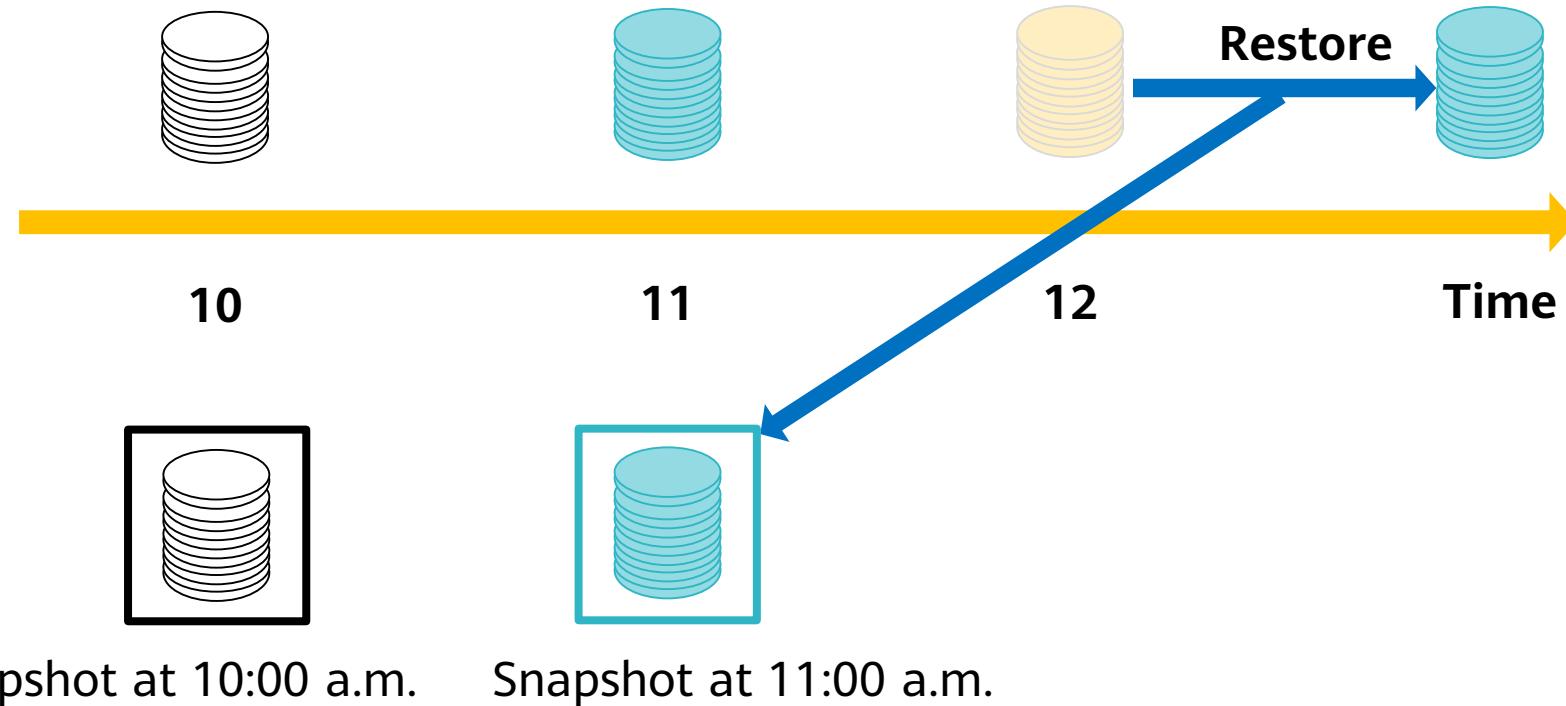
Snapshots do not affect read and write performance of source file systems.

Snapshots are read-only and have the same read performance as the source file systems.

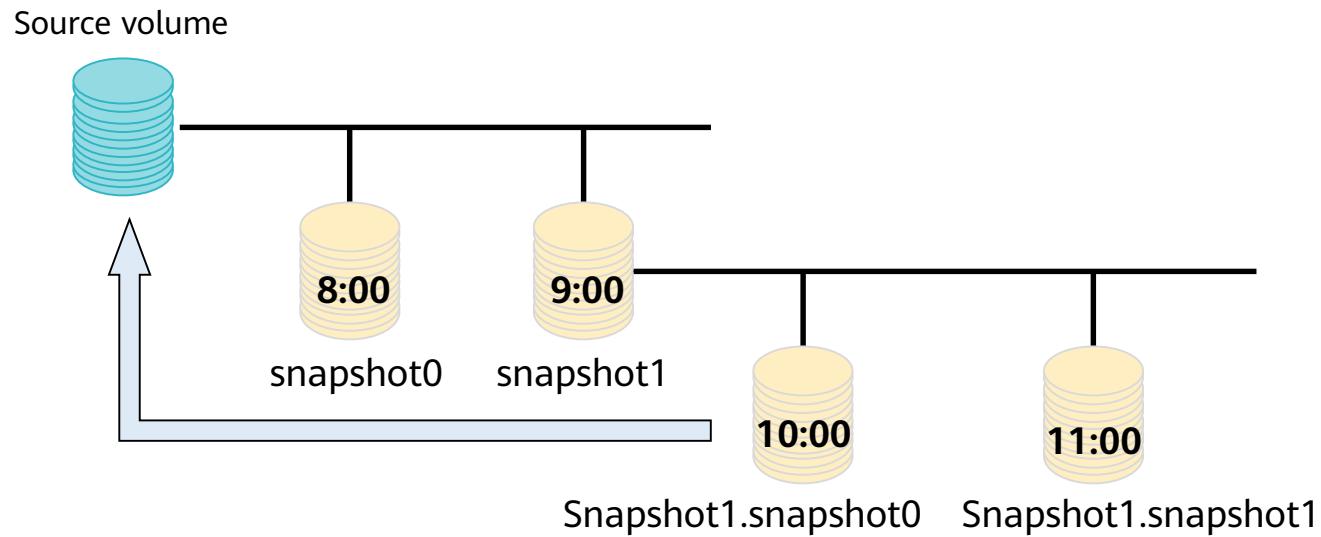
Data requested to be written to L0 and L2 of the source file system is written to new spaces P5 and P6. The original spaces P0 and P2 are referenced by the snapshot.

HyperSnap Principles: Rollback

Data at 10:00 a.m. Data at 11:00 a.m. Virus infection Data at 11:00 a.m.

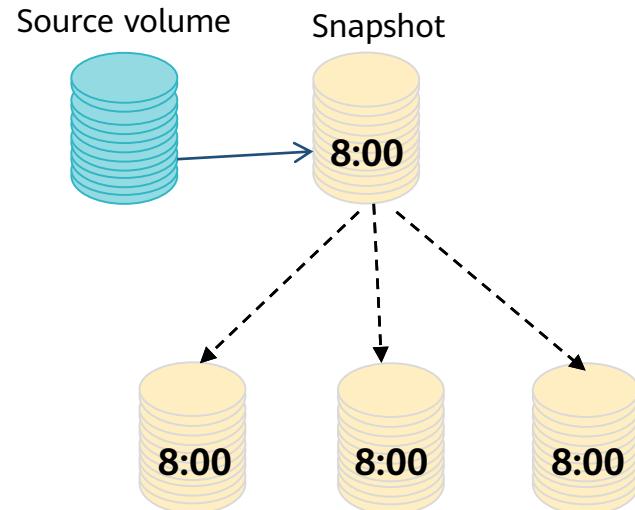


HyperSnap Principles: Snapshot Cascading and Cross-Level Rollback



- **Snapshot cascading:** It is a child snapshot of a parent snapshot. The difference between snapshot duplicates and snapshot cascading is that the latter includes the data of its parent snapshot. Other functions are the same as common snapshots.
- **Cross-level rollback:** Snapshots sharing the same source volume can roll back each other regardless of their cascading levels.

Key Technologies of HyperSnap: Duplicate



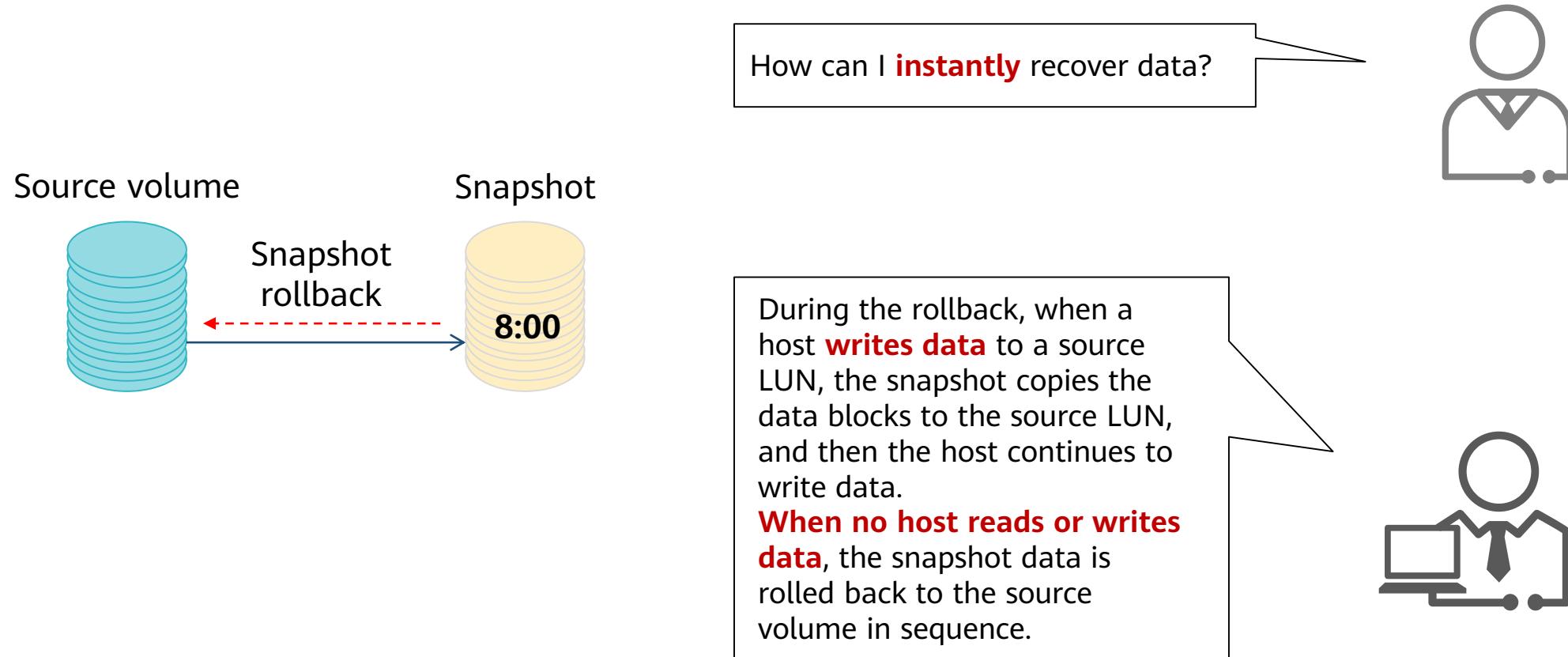
How can I obtain **multiple duplicates** of the same snapshot?



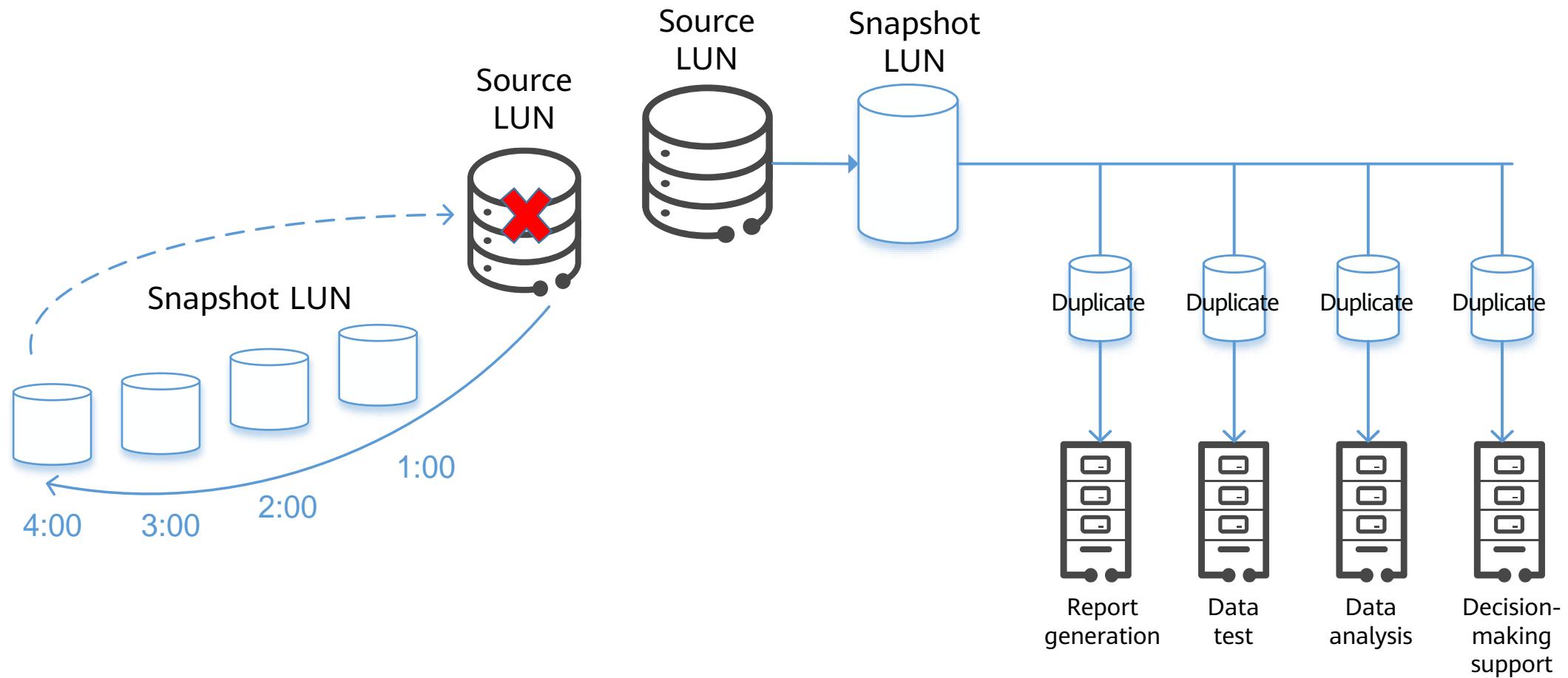
Snapshots are **virtual**, so they can be duplicated fast.



Key Technologies of HyperSnap: Rollback Before Write

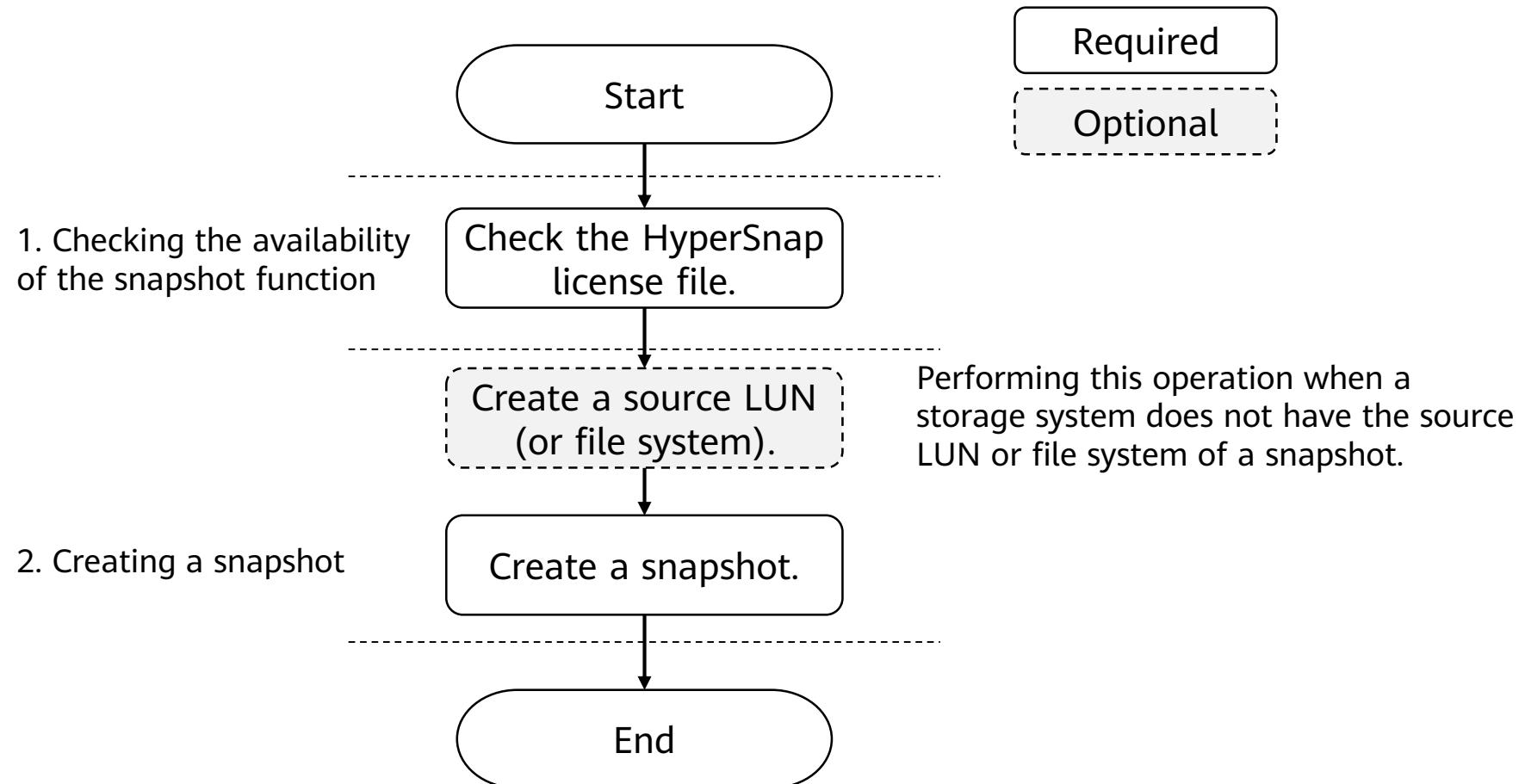


Application Scenario



- **Continuous data protection**
- **Data backup and restoration**

Configuration Process



Contents

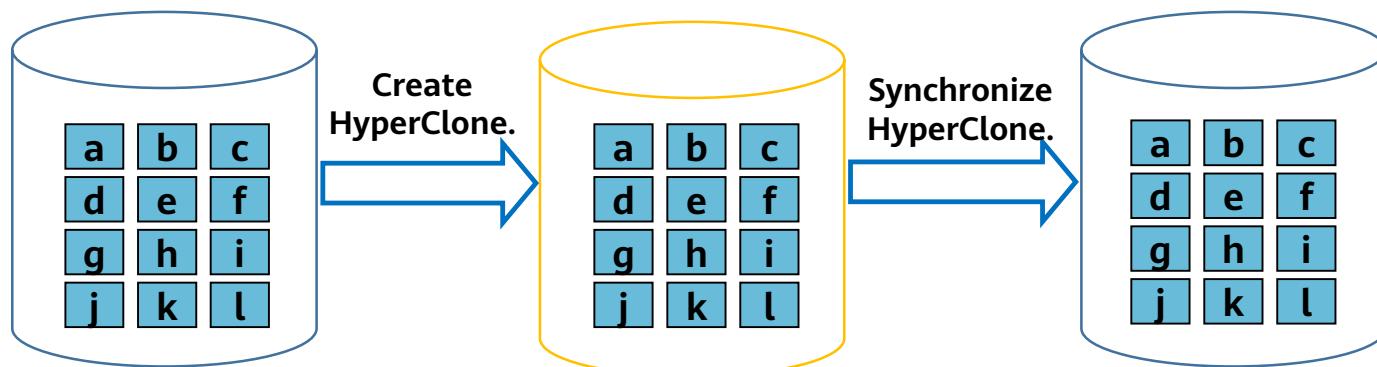
1. HyperSnap
- 2. HyperClone**
3. HyperCDP
4. HyperLock

Overview

- **Definition**
 - HyperClone creates a full data copy (a target LUN) of a source LUN at a specified point in time (synchronization start time).
 - HyperClone is to create a clone for a file system or a snapshot of the file system at a specific point in time. After a clone file system has been created, its data (including the dtree configuration and dtree data) is consistent with that of the parent file system at the corresponding point in time.
- **Features**
 - A target LUN can be read and written during synchronization.
 - Full synchronization and incremental synchronization are supported.
 - Forward synchronization and reverse synchronization are supported.
 - Consistency groups are supported.

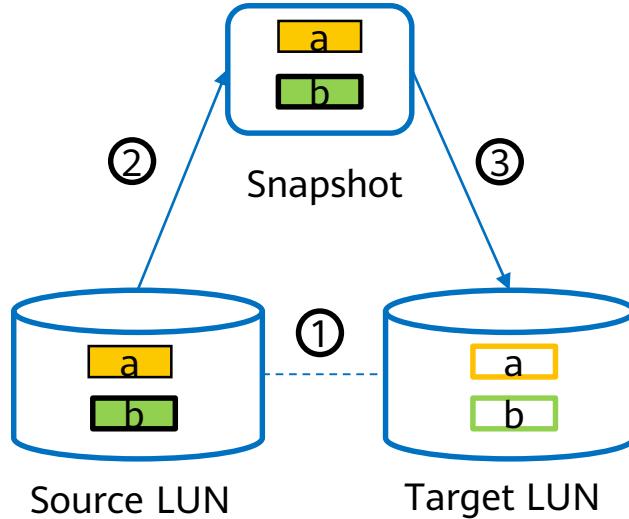
Working Principles of HyperClone

- **Definition:** Clone is a consistent data copy of a source data at a specific point in time. It functions as a complete data copy after data synchronization. It serves as a data backup and is accessible to hosts.
- **Main features**
 - **Quick clone generation:** A storage system can generate a clone within several seconds to obtain a consistency copy of a source data. The generated clone can be read and written immediately. Users can configure different deduplication and compression attributes for the generated clone.
 - **Online splitting:** A split can be performed to cancel the association between a source LUN and a clone LUN without interrupting services. The split read and write operation on the clone LUN will not affect the I/O process of the source LUN.



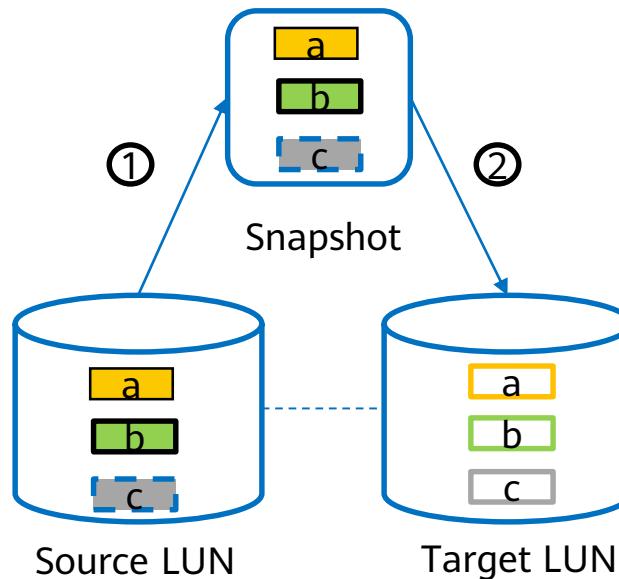
HyperClone Principles: Synchronization

Scenario 1: Initial synchronization and full copy are performed.

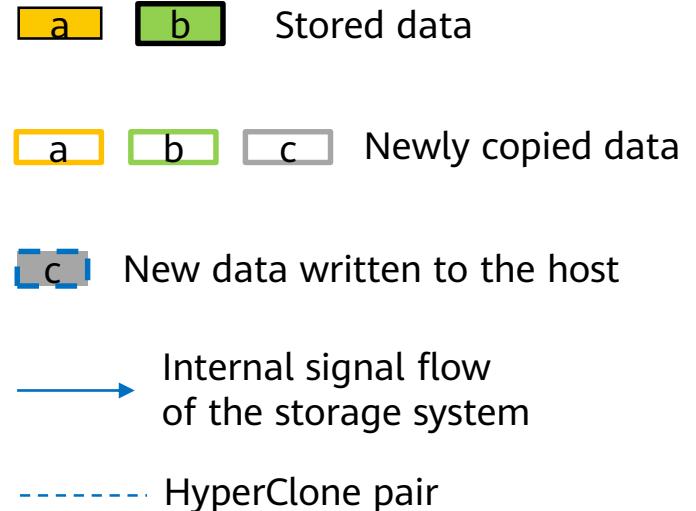


- ① Create a HyperClone pair.
- ② Create a snapshot for the source LUN after synchronization is started.
- ③ Copy all data a and b to the target LUN.

Scenario 2: Synchronization is performed again after the first synchronization, and differential copy is performed.

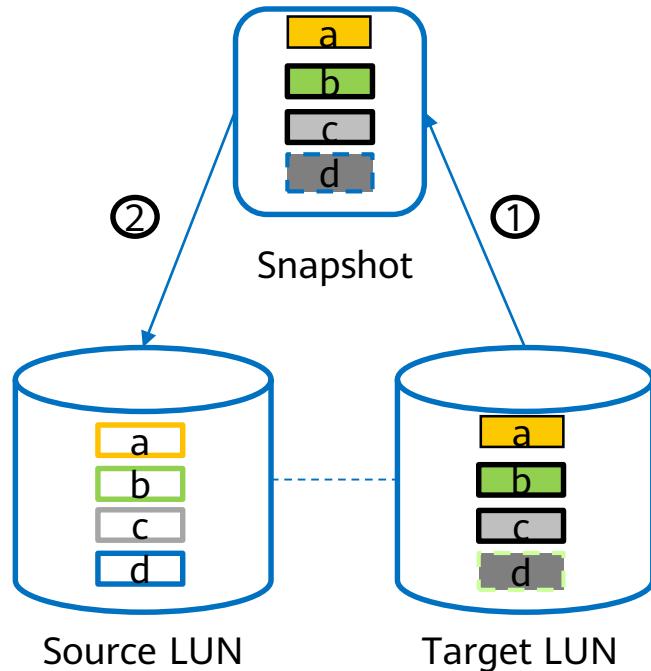


- ① Create a snapshot for the source LUN after a second synchronization.
- ② Copy incremental data c to the target LUN.



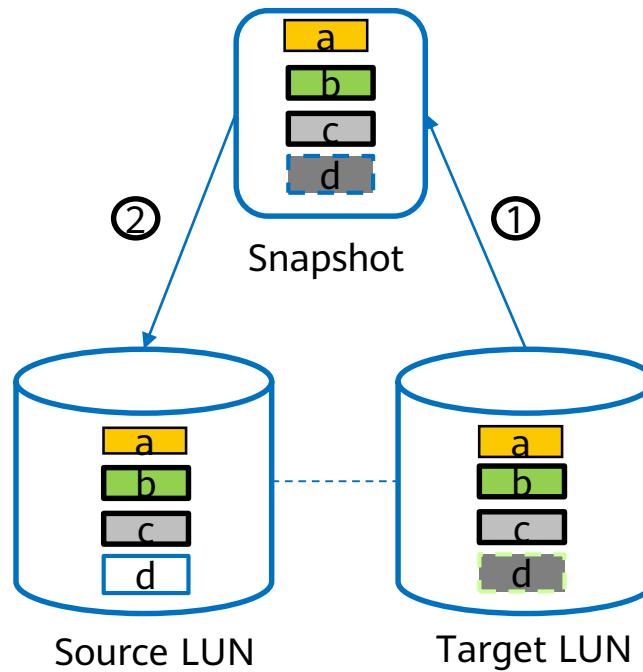
HyperClone Principles: Reverse Synchronization

Scenario 1: Full copy

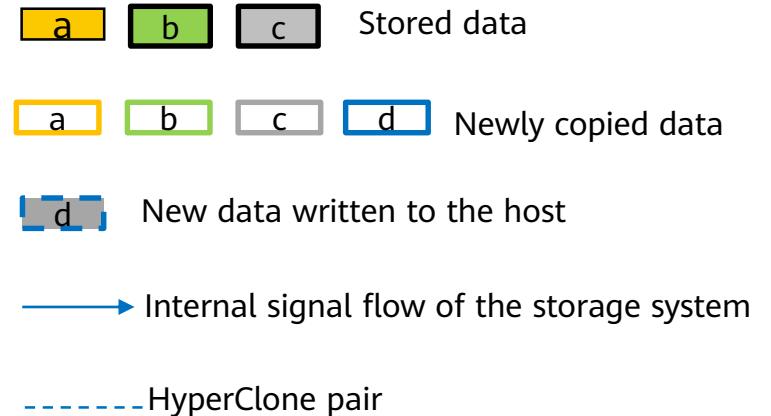


- ① Create a snapshot for the target LUN after the reverse synchronization is started.
- ② Copy all data a, b, c, and d to the source LUN.

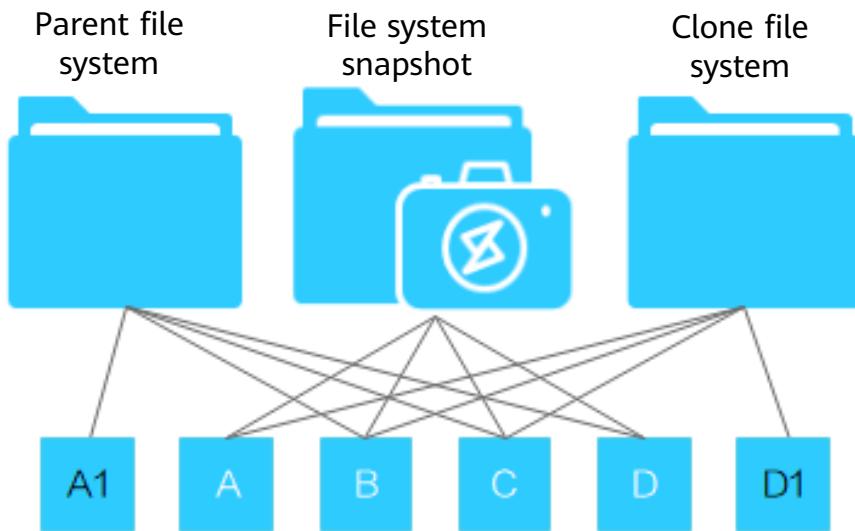
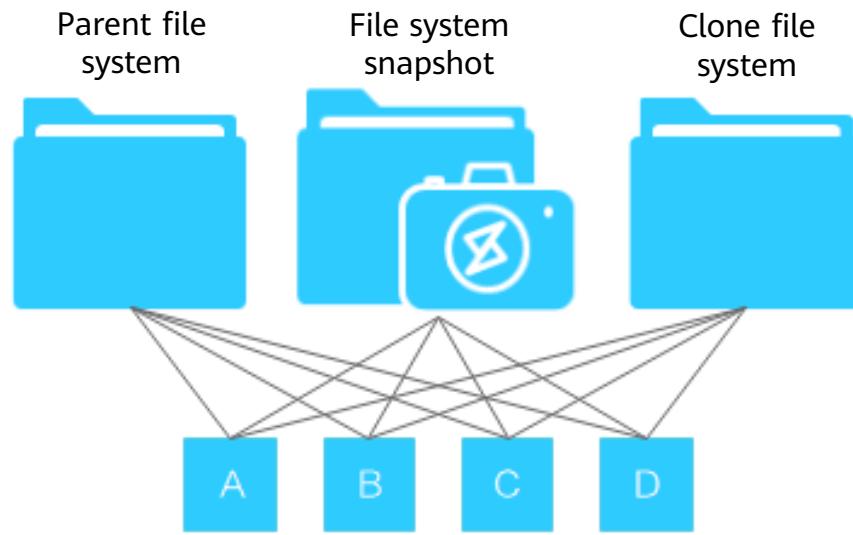
Scenario 2: Differential copy



- ① Create a snapshot for the target LUN after the reverse synchronization is started.
- ② Copy incremental data d to the source LUN.



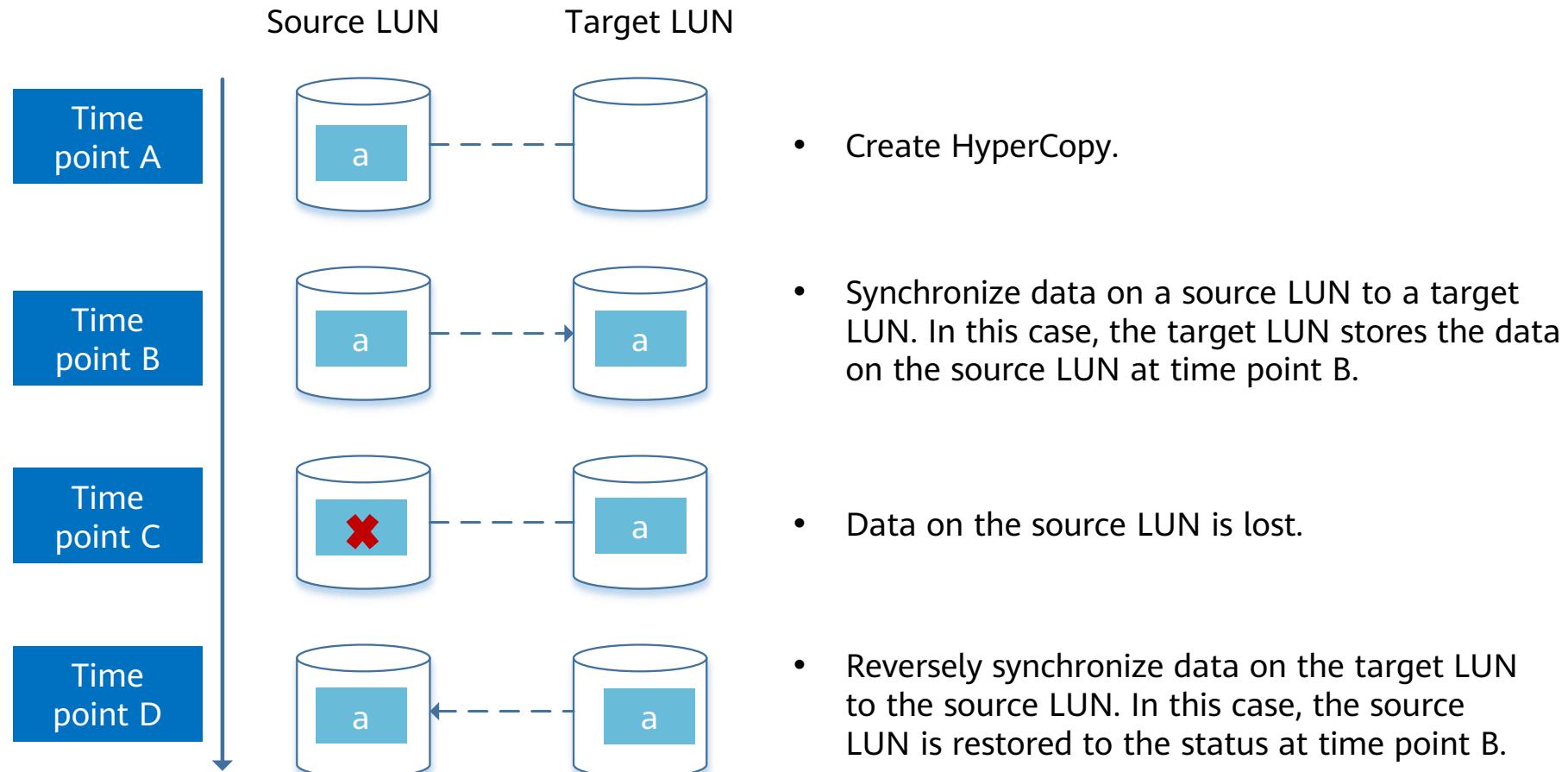
Read/Write Principles of a Clone File System



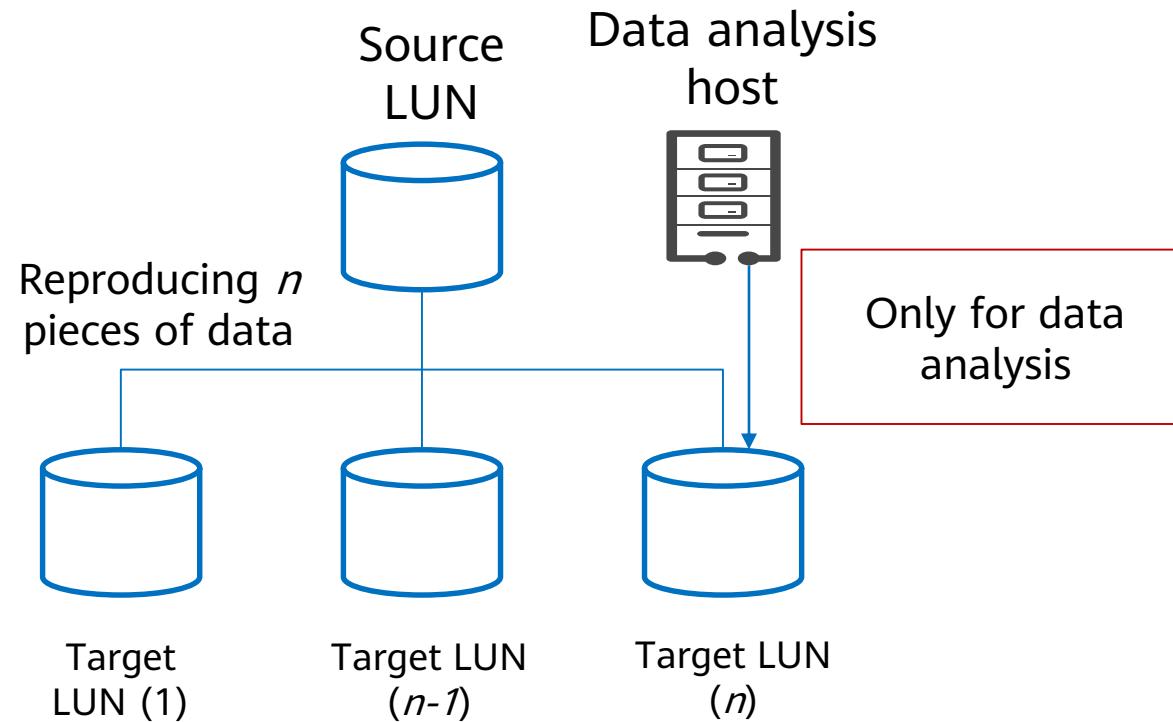
Data status in the clone file system before data change

Data status in the clone file system after data change

Application Scenarios: Data Backup and Restoration



Application Scenarios: Data Analysis and Reproduction



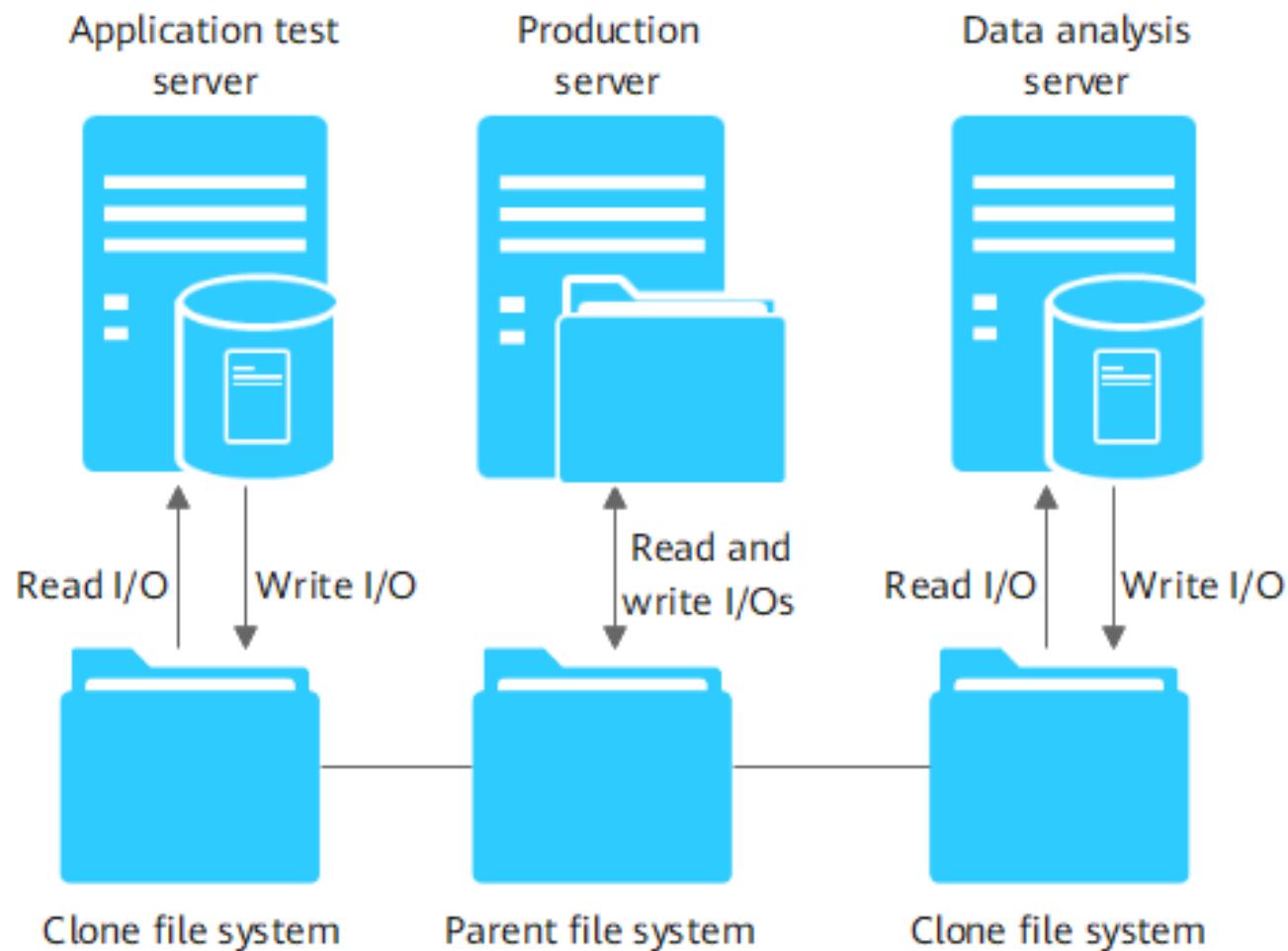
Data analysis

The data analysis service uses data on a target LUN to prevent the data analysis service and production service from contending for resources of a source LUN and affecting performance.

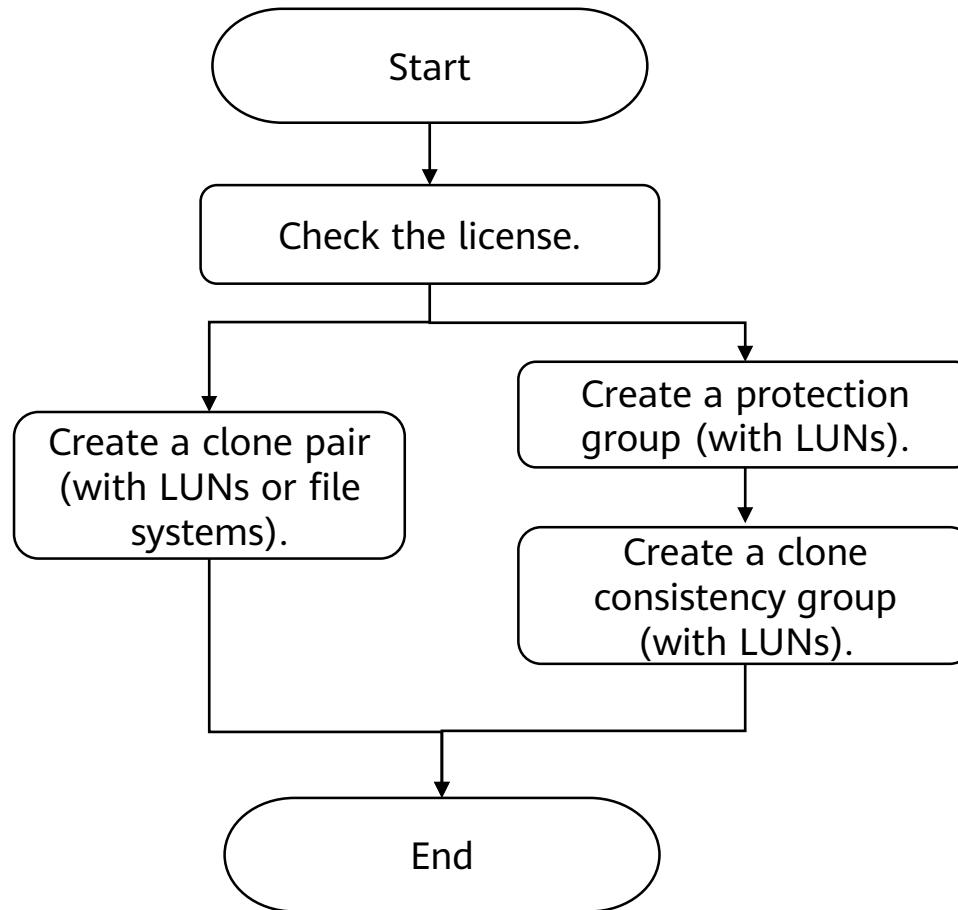
Data reproduction

HyperClone can create multiple copies of the same source LUN for multiple target LUNs.

Application Scenarios: Application Test and Data Analysis



Configuration Process



Contents

1. HyperSnap
2. HyperClone
- 3. HyperCDP**
 - **LUN HyperCDP**
 - File System HyperCDP

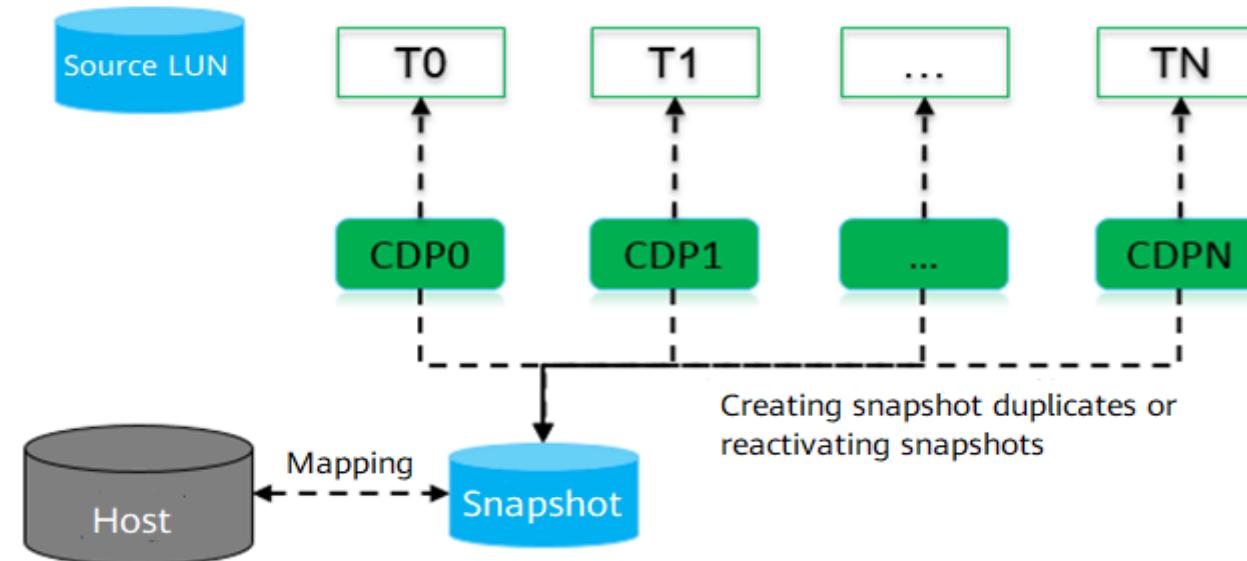
4. HyperLock

HyperCDP (for Block)

- HyperCDP creates high-density snapshots on a storage system to provide continuous data protection. A HyperCDP object is similar to a common writable snapshot, which is a point-in-time consistent copy of original data to which the user can roll back to, if and when it is needed. It contains a static image of the source data at the data copy time point.
- Main features
 - It provides intensive and persistent data protection. A single LUN supports 60,000 HyperCDP objects. The minimum interval is 3 seconds.
 - It provides data protection at an interval of seconds, with zero impact on performance and small space occupation.
 - It supports scheduled tasks. You can specify HyperCDP schedules by day, week, month, or specific interval.
 - It supports HyperCDP consistency groups.
 - A HyperCDP object cannot be directly mapped to a host for read and write operations but can be mapped to a host after being converted to a writable snapshot by creating a duplicate.

Working Principles of HyperCDP

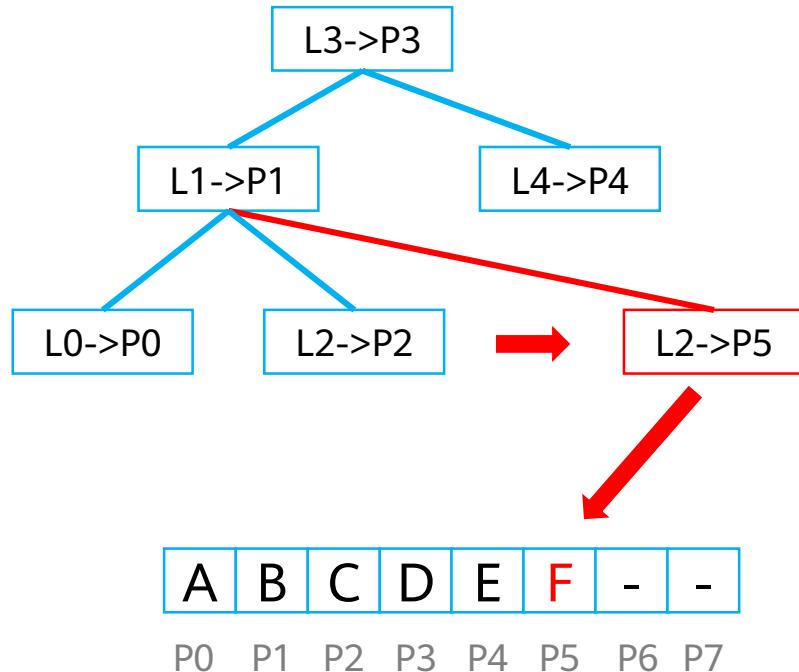
- Based on the lossless snapshot technology, HyperCDP has little impact on the performance of source LUNs. Compared with writable snapshots, HyperCDP does not need to build LUNs, greatly reducing memory overhead and providing stronger and continuous protection.
- HyperCDP objects cannot be mapped to hosts directly. To read data from a HyperCDP object, you can create a duplicate for it and map the duplicate to the host.



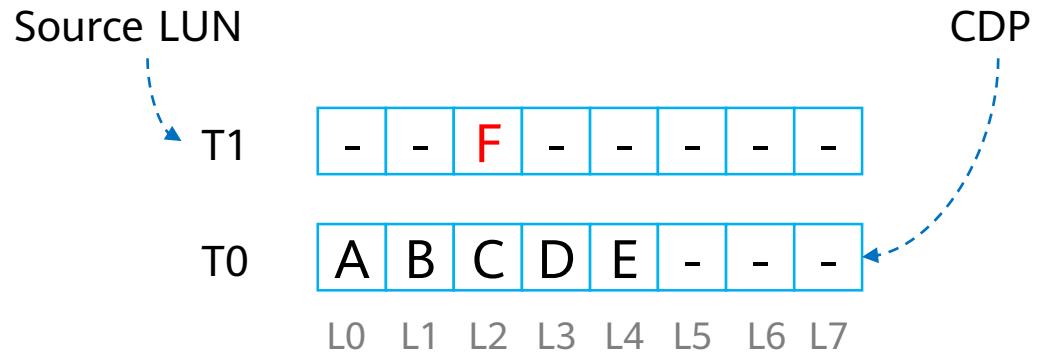
Creating a HyperCDP Object

- Creating a HyperCDP object is to save the data status of the source LUN at the activation time.

ROW principle



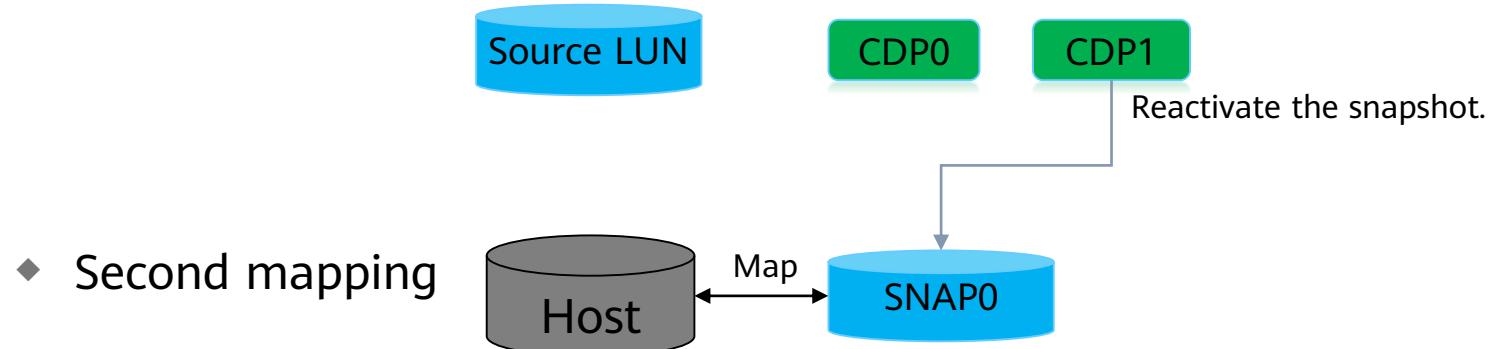
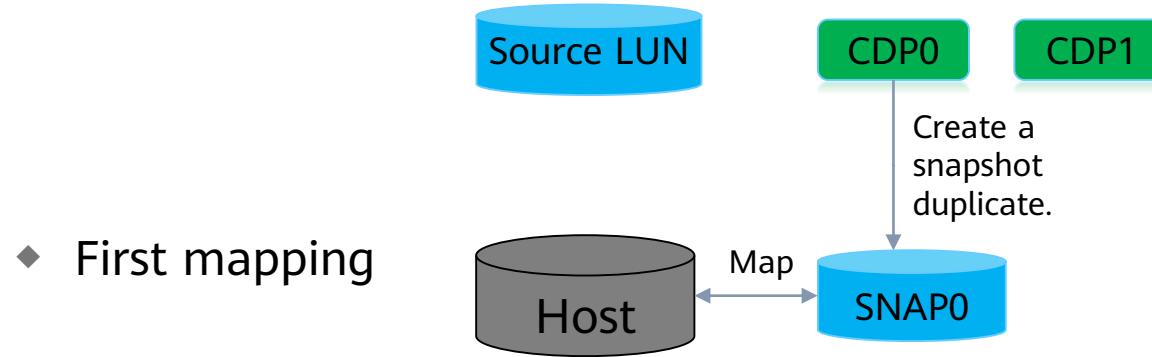
Host data access



- When a host accesses the source LUN data, the mapping table of the source LUN is accessed from the T1 layer. If the T1 layer does not exist, the data at the T0 layer is returned. For example, if the host accesses data of L2, F is returned. If the host accesses data of L0, A is returned.
- The HyperCDP mapping table is accessed from the T0 layer. If the host accesses data of L2, C is returned.

Reading and Writing a HyperCDP Object

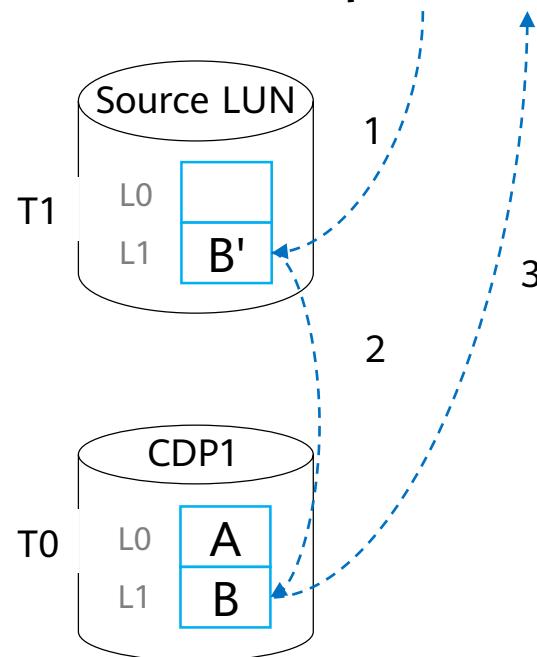
- A HyperCDP object cannot be directly mapped to a host for read and write operations but instead, should be mapped to a host after being converted to a writable snapshot.



Rolling Back a HyperCDP Object (1)

- HyperCDP rollback is a process of copying data from a HyperCDP object to the source LUN. The source LUN is available immediately after the rollback is started (data on the source LUN is the data on a HyperCDP object).

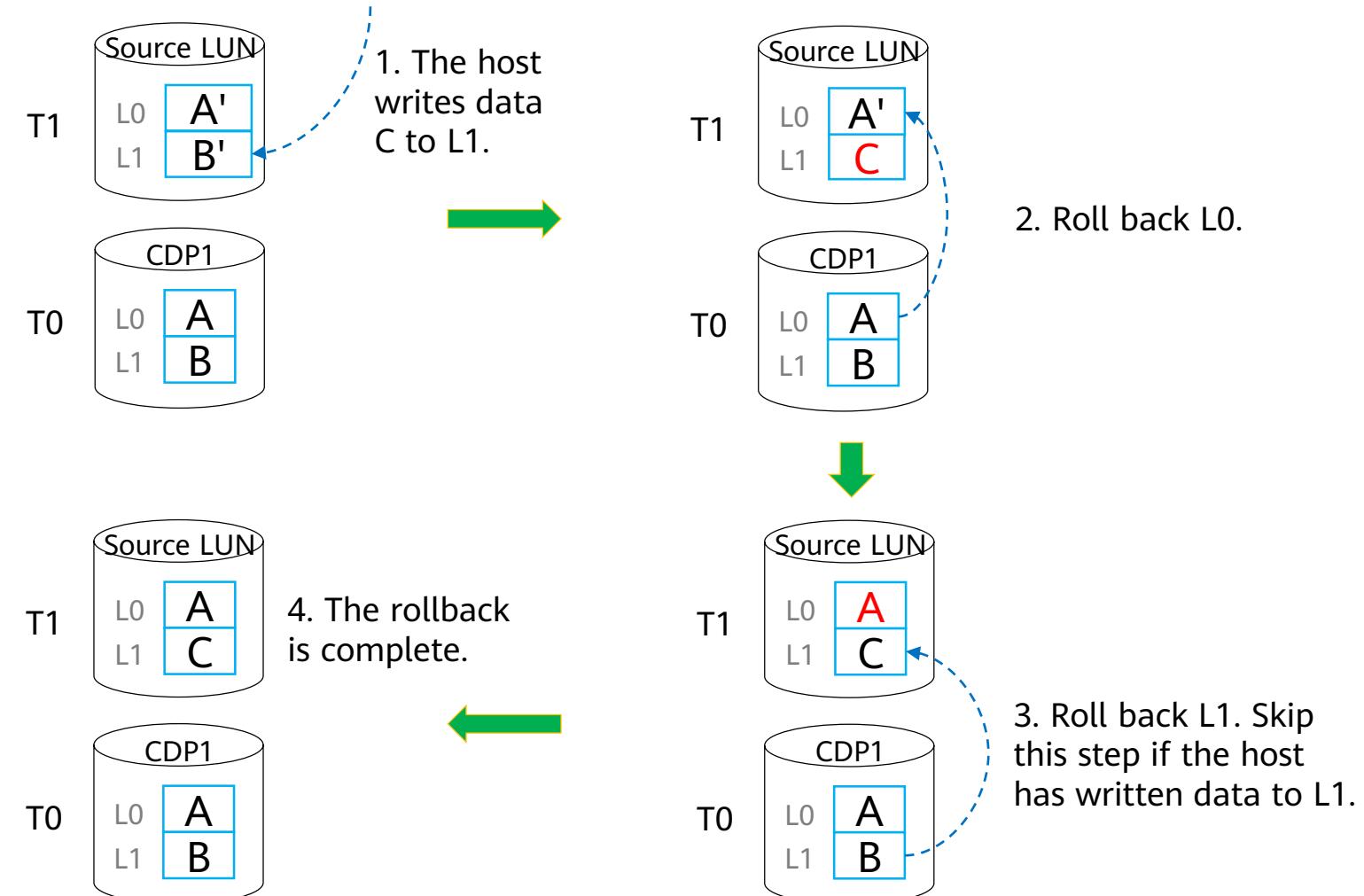
Read data from the source LUN during HyperCDP rollback and perform read redirection.



1. The host reads L1.
2. If L1 is not rolled back and the host has not written data to L1 after the rollback is started, the host reads data from CDP1 (T0) instead.
3. Data **B** is returned to the host.

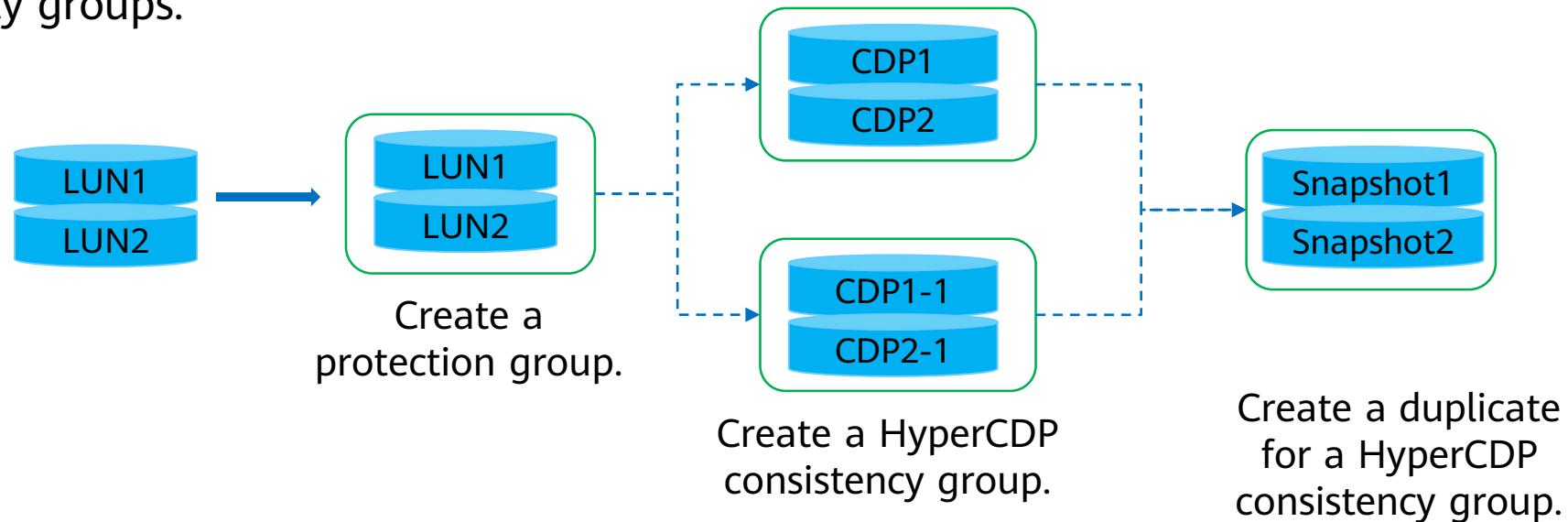
Rolling Back a HyperCDP Object (2)

Write the source LUN during HyperCDP rollback.



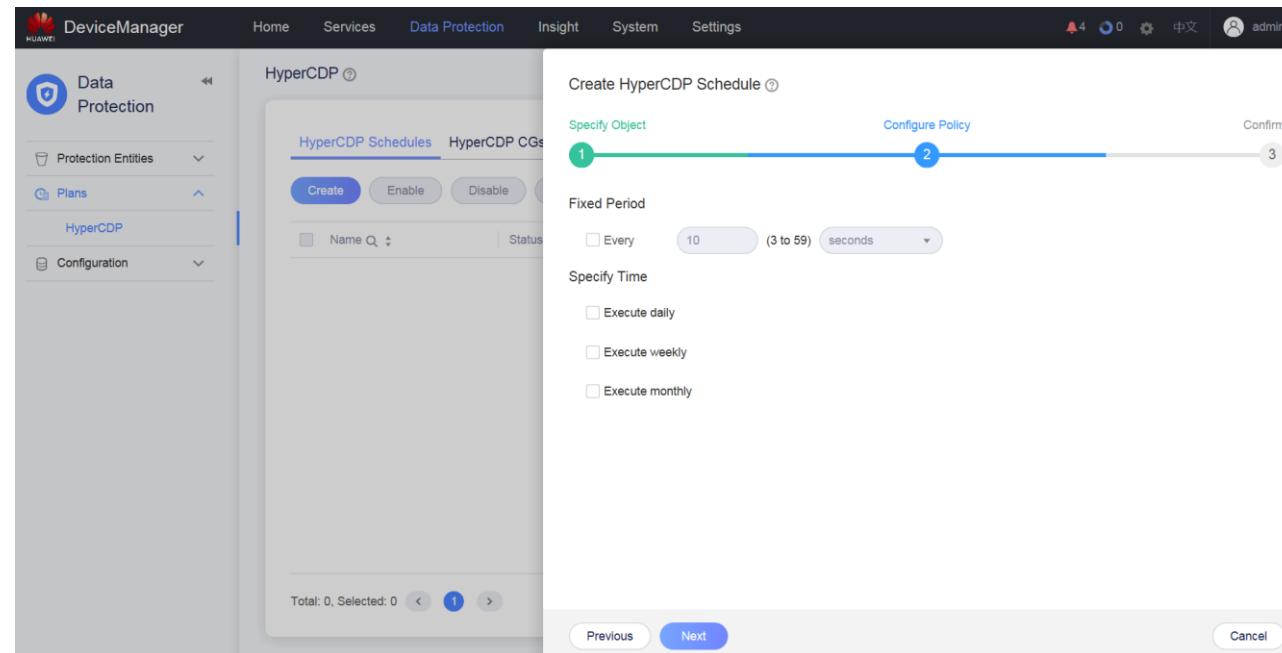
HyperCDP Consistency Group

- Medium- and large-size databases' data, logs, and modification information are stored on different LUNs. If data on one of these LUNs is unavailable, data on the other LUNs is also invalid. The HyperCDP consistency group ensures the consistency of application data during restoration.
- Similar to individual HyperCDP objects, you can create, delete, roll back, or stop rolling back a HyperCDP consistency group as required. You can also create or rebuild duplicates for HyperCDP consistency groups.



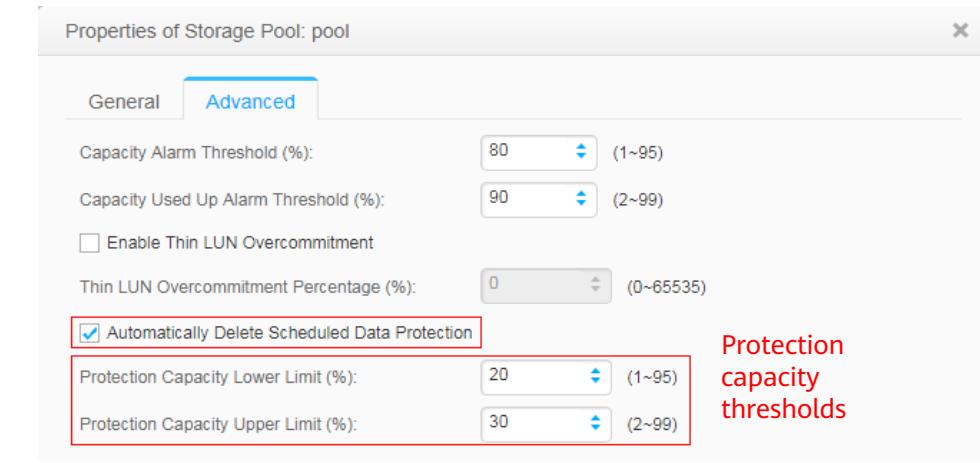
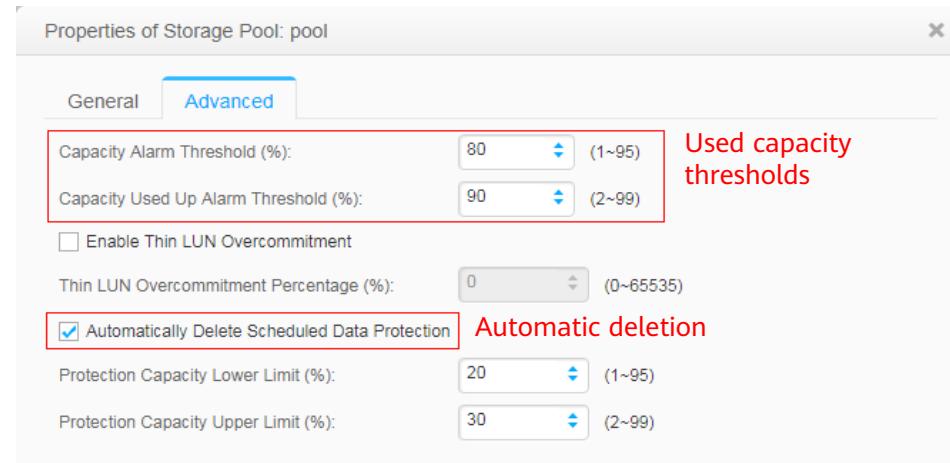
HyperCDP Schedule

- You can specify HyperCDP schedules by day, week, month, specific interval, or any combination of them. You can also specify the quantity of HyperCDP objects that can be retained for each schedule.
- You can add multiple LUNs and LUN consistency groups to a HyperCDP schedule, but you can add a LUN or a LUN consistency group to only one HyperCDP schedule.
- A HyperCDP schedule supports the minimum interval of 3 seconds and retention of up to 60,000 objects.



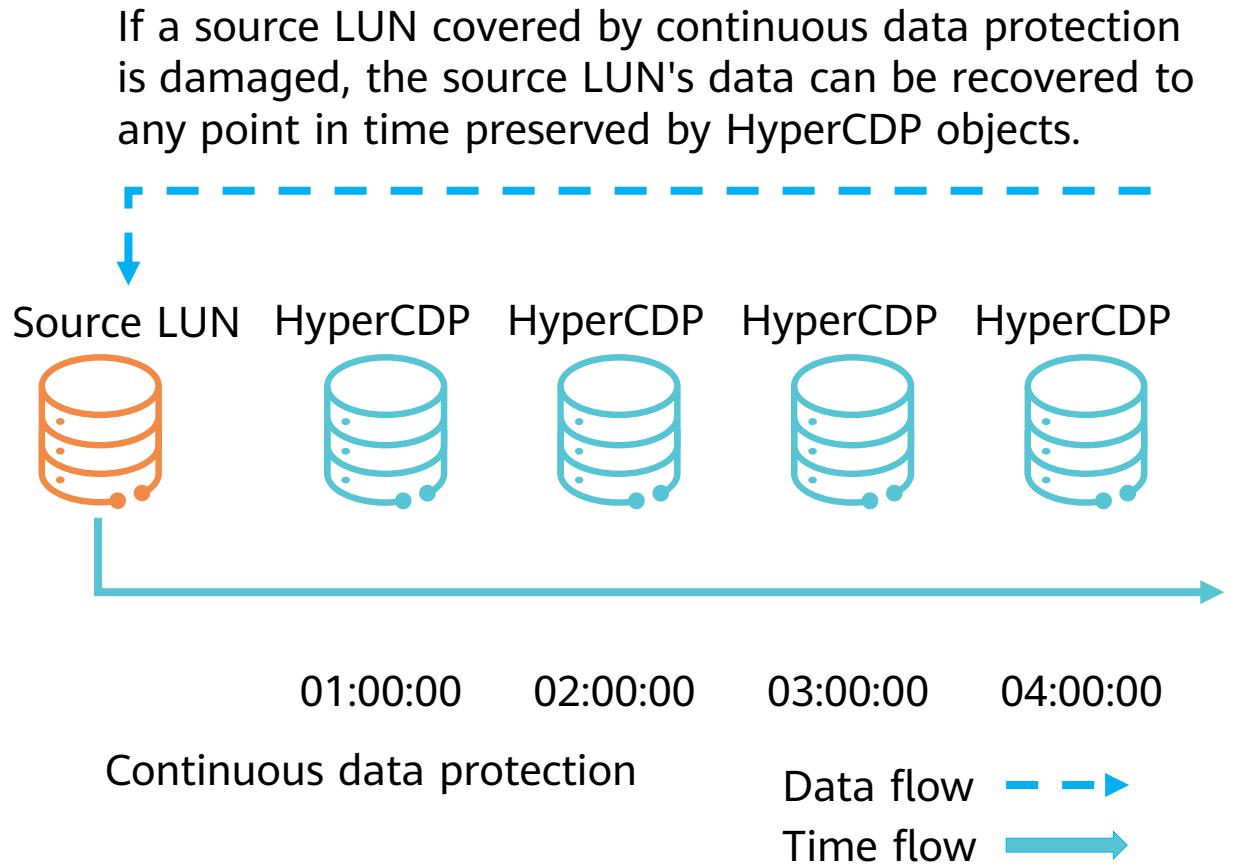
Storage Pool Capacity Threshold

- Because HyperCDP supports a minimum interval of 3 seconds, a large amount of data protection capacity may be required if new data is writing to the source LUN constantly. As a result, the space of the storage pool may be used up, affecting host services.
- You can set a threshold to the used capacity of the storage pool and the protection capacity, respectively. You can enable or disable it by using automatic deletion of protection data.

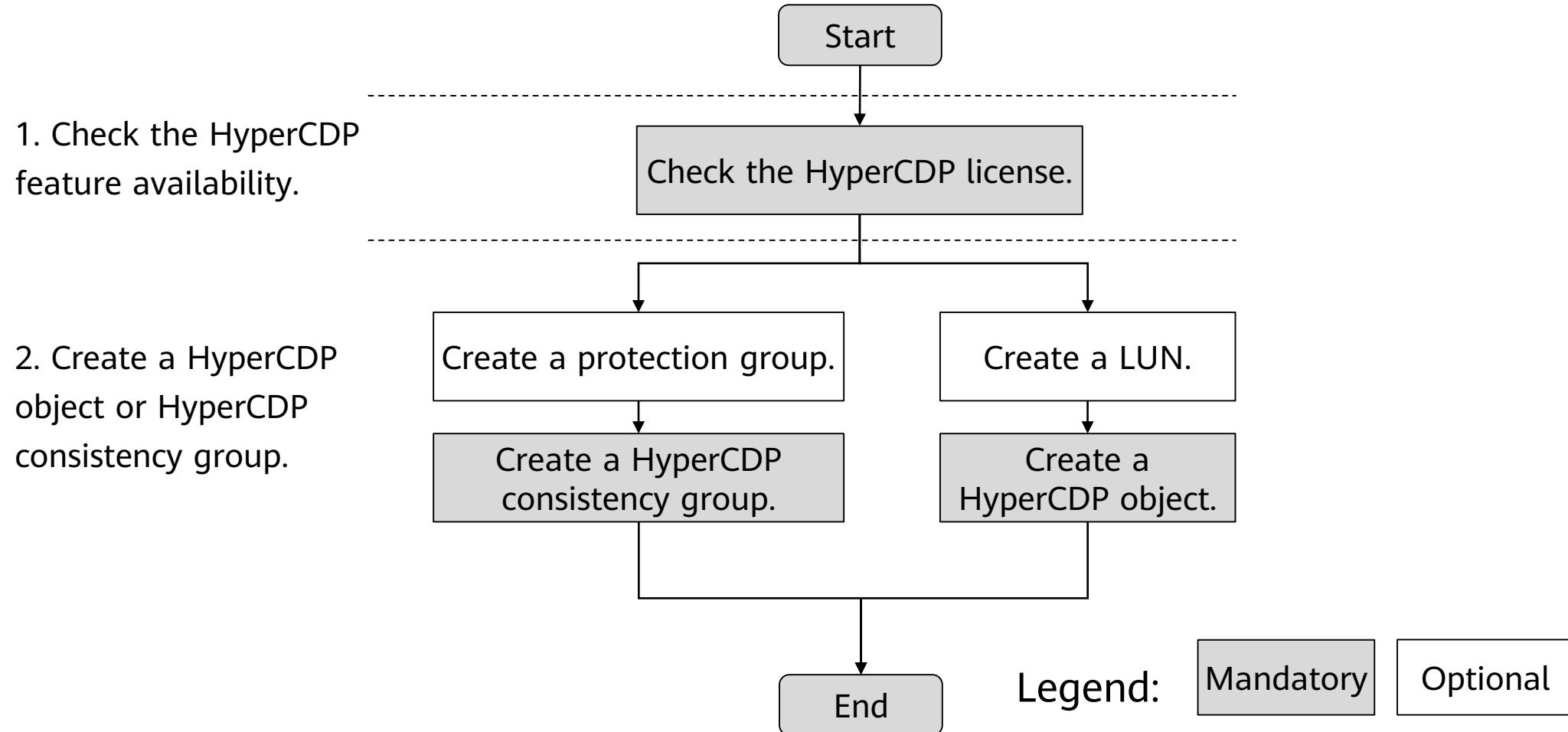


Application Scenario: Quick Data Backup and Restoration

- HyperCDP backup allows fast data restoration in the following scenarios:
 - Virus infection
 - Misoperations
 - Malicious tampering
 - Data corruption caused by system breakdown
 - Data corruption caused by application bugs
 - Data corruption caused by storage system bugs



Configuration Process



Contents

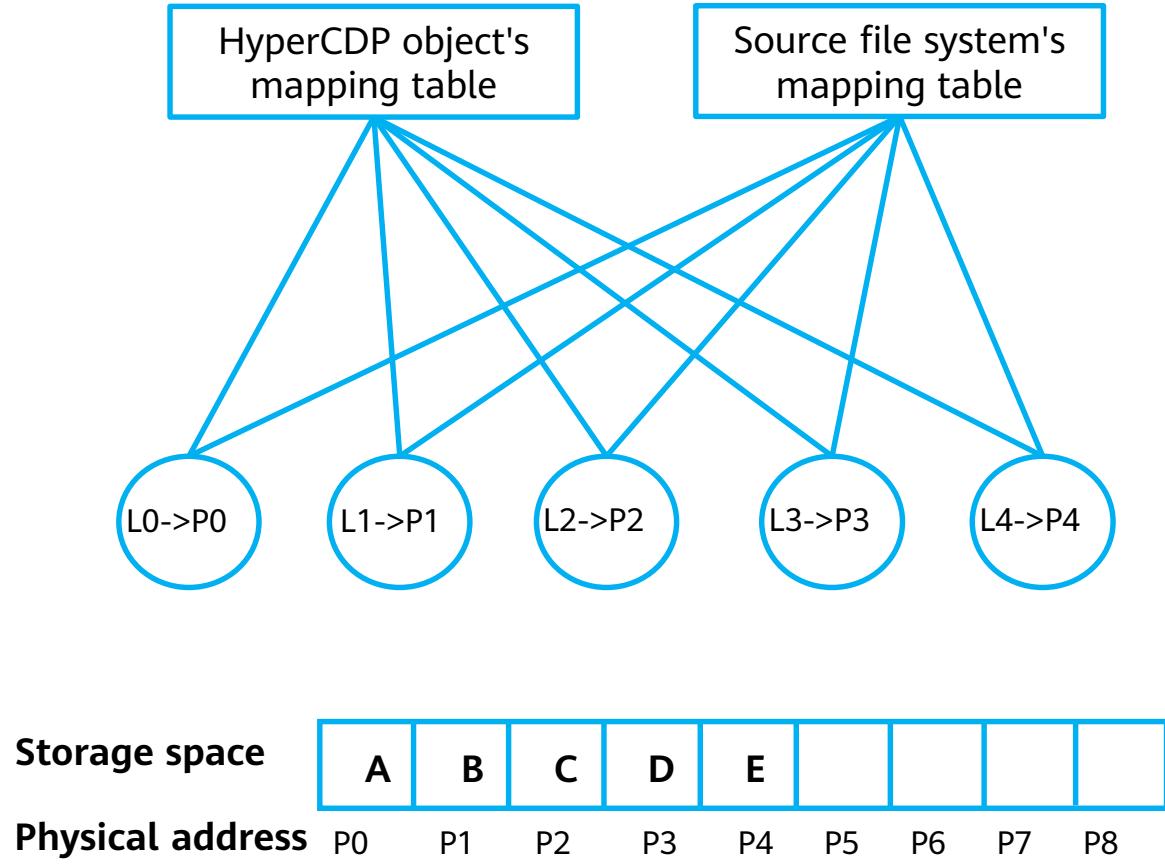
1. HyperSnap
2. HyperClone
- 3. HyperCDP**
 - LUN HyperCDP
 - **File System HyperCDP**

4. HyperLock

HyperCDP (for File)

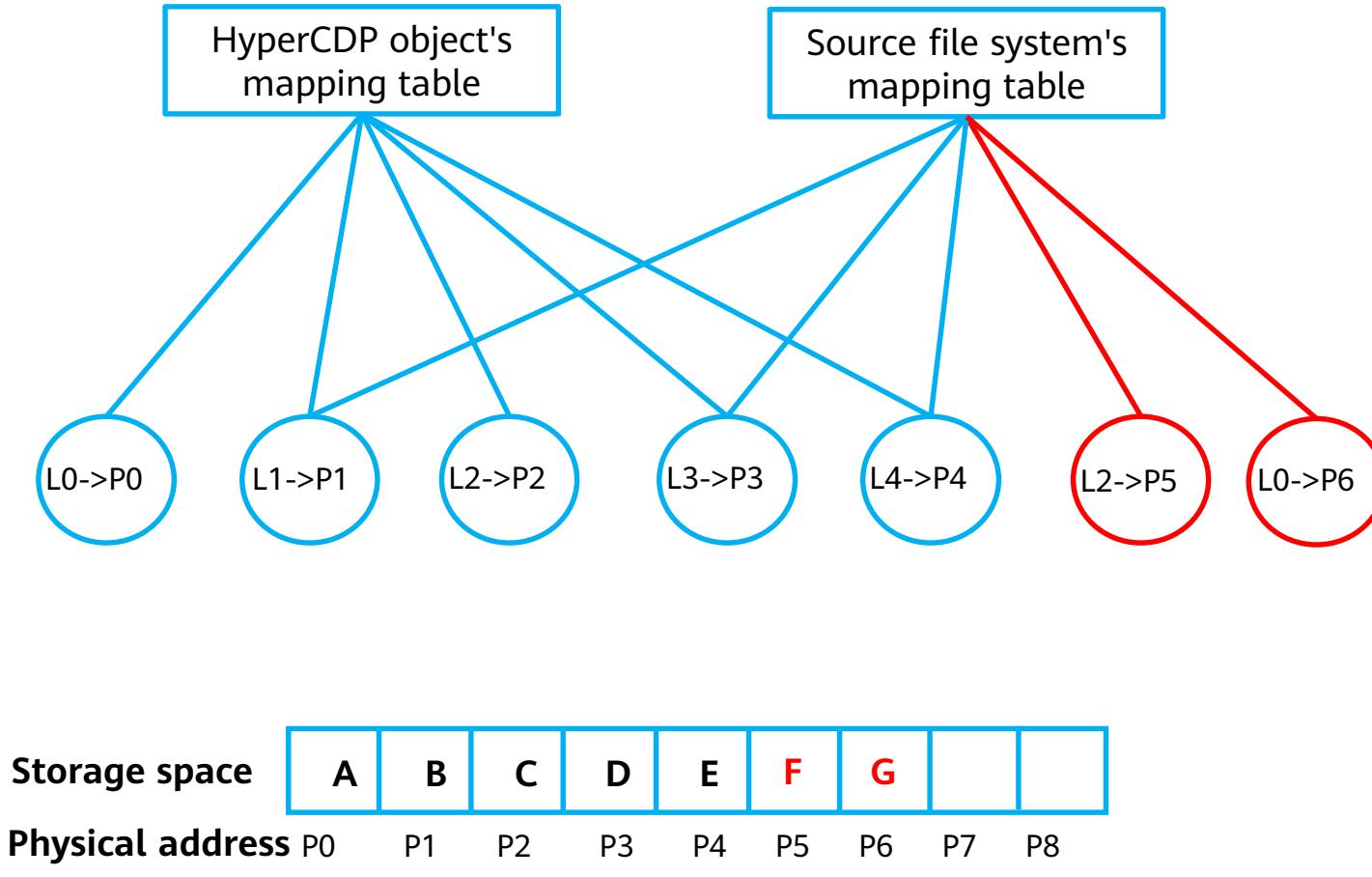
- HyperCDP creates high-density snapshots on a storage system to provide continuous data protection. Based on the lossless snapshot technology, HyperCDP has little impact on the performance of source file systems. Compared with common writable snapshots, HyperCDP does not need to build file system objects, which greatly reduces the memory overhead, delivering more intensive and continuous protection.
- OceanStor storage system supports HyperCDP schedules to meet customers' backup requirements. HyperCDP objects cannot be directly mapped to hosts. To read data from a HyperCDP object, you need to share the file system with the host so that the host can read the source file system data at the point in time when the HyperCDP object was created.

Creating HyperCDP Duplicates



- After HyperCDP is enabled, the system generates a HyperCDP object for the file system. After the HyperCDP object is created, it shares the mapping table of the source file system. In addition, the creation of the HyperCDP object does not affect the data read or write performance of the source file system. The HyperCDP object is read-only.
- The original data in the source file system is ABCDE. A HyperCDP object is created and shares the mapping table of the source file system. In the figure, L0 to L4 are logical addresses, P0 to P8 are physical addresses, and A to E are the data.

Data Writing to the Source File System



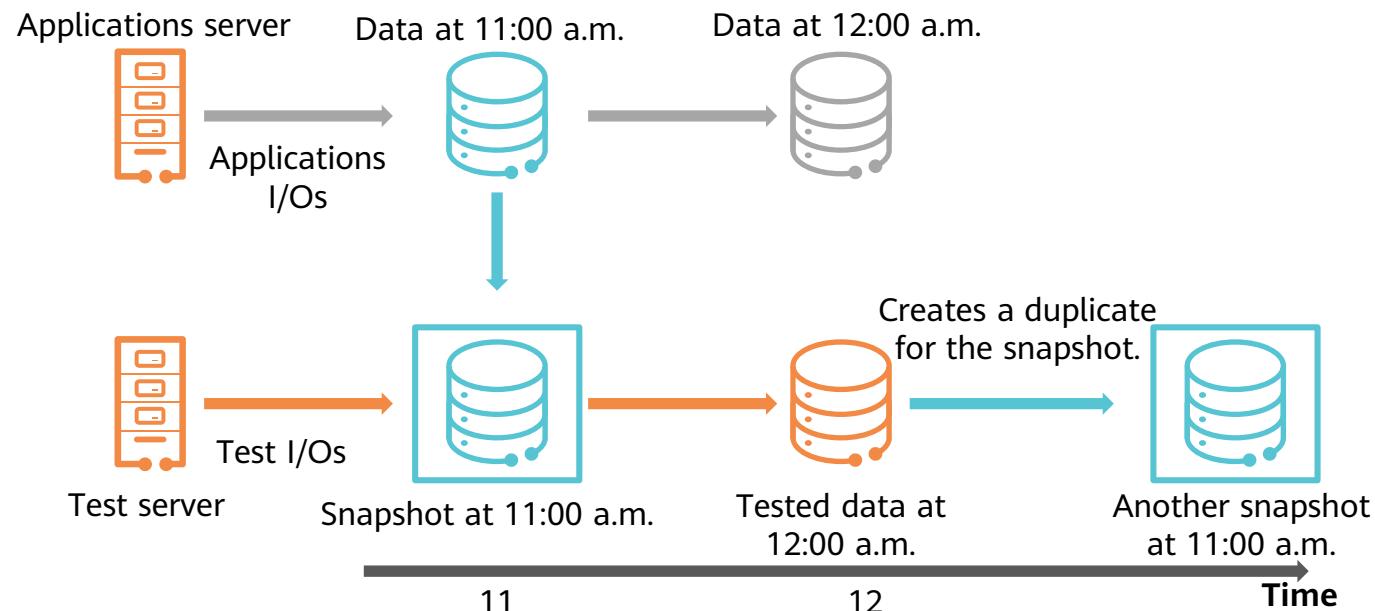
- After the HyperCDP object is created, the host writes data F and G to L0 and L2 of the source file system. Because the ROW mechanism is used, the host directly applies for new physical addresses P5 and P6 for data writing, and the original addresses P0 and P2 are referenced by the HyperCDP object.
- The HyperCDP object does not change the write process of the source file system or cause any extra overhead. This avoids any loss of performance.

Read Principle of HyperCDP

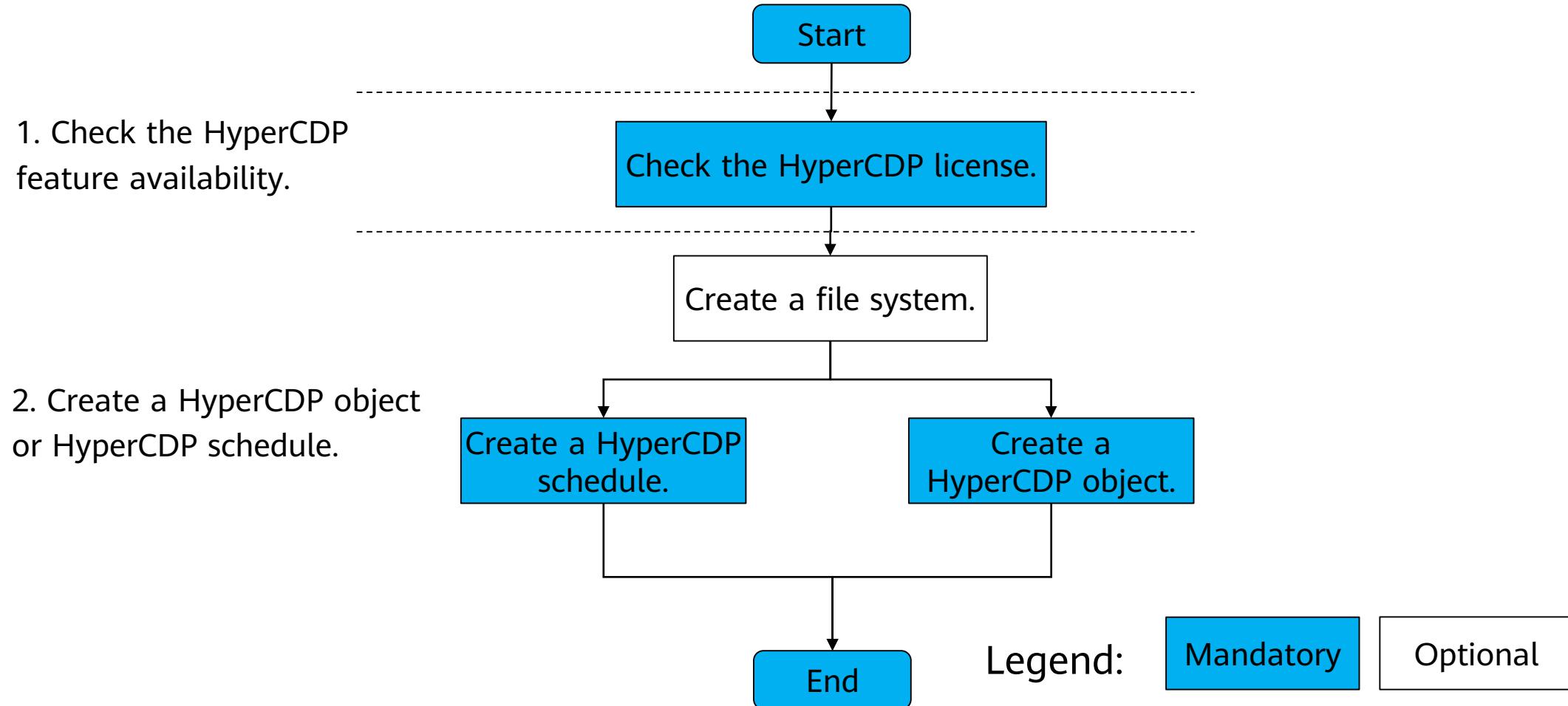
- You can configure the NFS or CIFS sharing service in a storage system to share file systems with clients. A HyperCDP object is a periodic read-only snapshot of the file system. After a HyperCDP object is created, client applications can access the specified share path (**.snapshot** directory) in the file system to read the source file system data at the point in time when the HyperCDP object was created.
- The scheduled snapshot can be accessed in the following three modes:
 - Enter the **.snapshot** directory through the shared root directory or dtree directory of the file system to access the data of any snapshot.
 - You can go to the **.snapshot** directory from any directory. All scheduled snapshots for the directory are displayed.
 - You can enter the specific path to go to the **.snapshot** directory from any directory and access the snapshot. Access to a scheduled snapshot that does not protect the previous directory will fail.

Application Scenario - Data Mining and Test Scenario

- Duplicates can be created for HyperCDP objects and used for data mining and testing, which will not affect service data.
 - A HyperCDP object is generated for the data to be tested at 11:00 a.m.
 - A duplicate is created for the HyperCDP object and is read and written by the test server. During the test, the source data and services accessing the source data are not affected.
 - 1 hour later, the source data and duplicate data are changed based on the data at 11:00.
 - After the test, users can create another duplicate for the HyperCDP object to obtain the data at 11:00 and use the duplicate for another test.



Configuration Process



Contents

1. HyperSnap
2. HyperClone
3. HyperCDP
- 4. HyperLock**

Overview

- Write Once Read Many (WORM), also called HyperLock, protects the integrity, confidentiality, and accessibility of data, meeting secure storage requirements.
- A file protected by WORM enters the read-only state immediately after data is written to it. In read-only state, the file can be read but cannot be deleted, modified, or renamed. The WORM feature can prevent data from being tampered with, meeting data security requirements of enterprises and organizations.
- A file system with the WORM feature (a WORM file system for short) can be configured only by the administrator. File systems with the WORM feature are classified into regulatory compliance WORM (WORM-C) and enterprise WORM (WORM-E) according to the administrator's permissions.

Mode	Application Scenario
WORM-C	This mode applies to archive scenarios where data protection mechanisms are implemented as required by laws and regulations.
WORM-E	This mode is mainly used by enterprises to implement internal control.

WORM Compliance Clock

- To prevent users from changing protection periods of files by changing the system time, storage systems maintain a WORM compliance clock. WORM compliance clocks include the global security compliance clock and WORM file system compliance clock.

Clock Type	Function	Description
Global security compliance clock	The storage system maintains a global security compliance clock that serves as the clock source for all WORM file systems.	When creating a WORM file system for the first time, the system administrator must initialize the global security compliance clock. The time of the global security compliance clock cannot be changed after initialization.
WORM file system compliance clock	Each WORM file system maintains a compliance clock. The protection periods of files are based on the compliance clock.	The system will automatically use the global security compliance clock to initialize the WORM file system compliance clock upon the creation of a WORM file system. You do not need to manually initialize the WORM file system compliance clock.

File Status

- There are four file states in a WORM file system, as described in the following table.

Status	Description
Initial	All newly created files are in the initial state. Files in the initial state can be read, written, and modified by all users.
Locked	Files in the locked state cannot be modified, deleted, or renamed by all users. These files can only be read and their properties can be viewed.
Expired	Files in the expired state can be deleted and read and their properties can be viewed. However, these files cannot be modified or renamed.
Appending	Data can be added to the end of files in the appending state and these files cannot be deleted, truncated, or renamed.

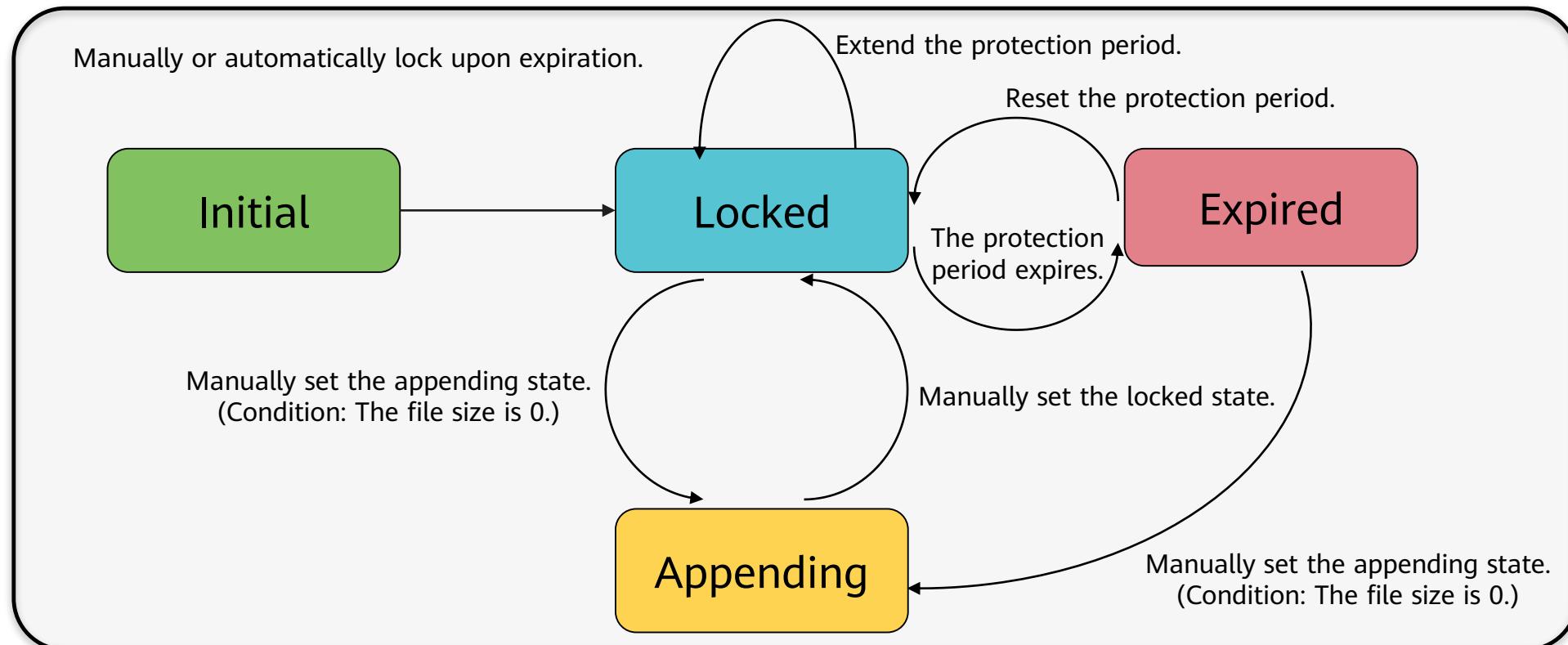
WORM Properties

- After the WORM feature is configured for a file system, the file system has the WORM properties. The WORM properties apply to files in the WORM file system. You can view the WORM properties to determine the lock time and expiration time of a file. The following table lists the WORM properties of a file system.

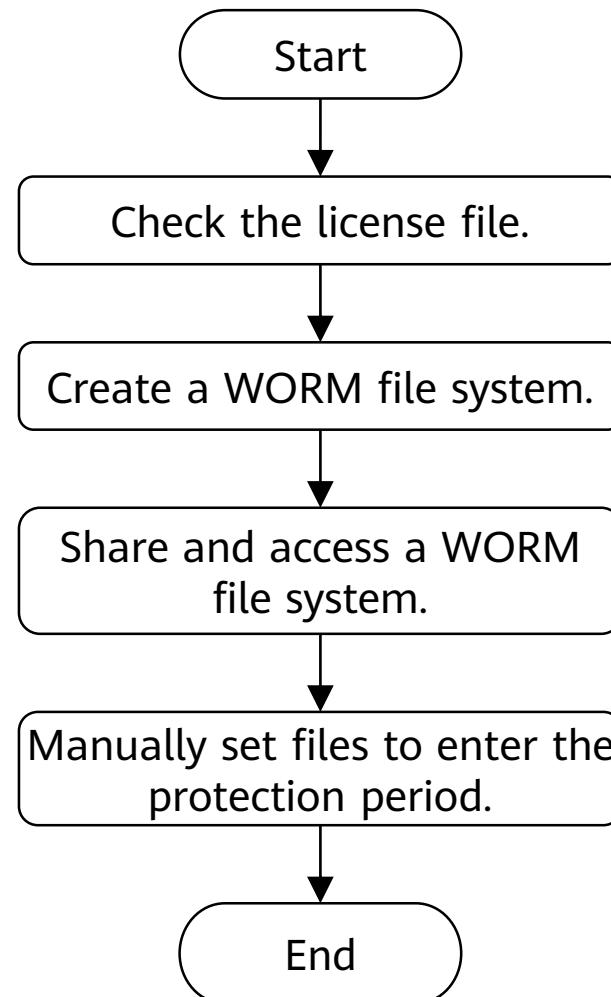
Property	Description
Mode	The system supports only the regulatory compliance mode.
Min. Protection Period	Minimum retention period supported by the WORM file system. The retention period of a file in the WORM file system cannot be smaller than the minimum retention period.
Max. Protection Period	Maximum retention period supported by the WORM file system. The retention period of a file in the WORM file system cannot be larger than the maximum retention period.
Default Protection Period	Default retention period supported by the WORM file system. The retention period of a file in the WORM file system is the default value of the parameter if you do not set a retention period for the file.
Automatic Lockout	After the automatic lockout function is enabled, files in the WORM file system automatically enter the locked state a specific period of time after data or metadata in the files is modified.
Lockout Wait Time	Default waiting time for modified files to automatically enter the protection state. This parameter is valid only when Automatic Lockout is enabled.
Automatic Deletion	After this function is enabled, the system automatically deletes files whose protection periods have expired. Note: Before enabling this function, ensure that files do not need protection and can be automatically deleted by the system after they expire.

Working Principles

- If a common file system is protected by the WORM feature, files in the file system can be read only within the protection period. After WORM file systems are created, you need to map them to application servers using the NFS or CIFS protocol.

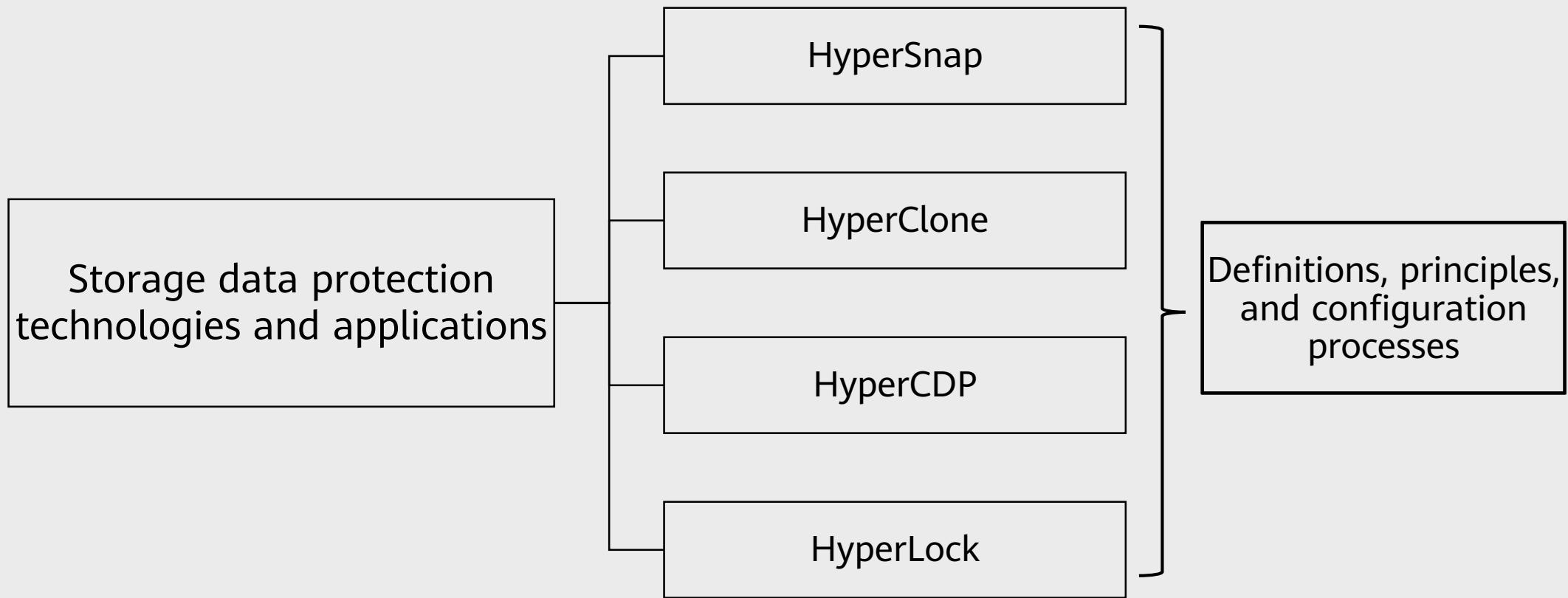


Configuration Process



- A license file grants the permission to use a specific value-added feature. Before configuring a value-added feature, ensure that the license file of the feature is valid.
- When creating a WORM file system for the first time, you need to initialize the global WORM compliance clock of the storage system.
- After a WORM file system is created, storage resources can be shared as file directories.
- After creating a WORM file system, share it with clients. You can store files that need to be protected in the WORM file system to prevent data tampering.
- After a WORM file system is created, if the automatic lockout function is not enabled, you may need to manually set files in the WORM file system to enter the locked state on the client. In addition, you can set the files to enter the appending state from the locked state to add content to the file.

Summary



Quiz

1. (True or false) A source LUN can form multiple HyperClone pairs with different target LUNs. A target LUN can be added to only one HyperClone pair.
2. (Short-answer question) When can a host read or write the source LUN after a rollback command is executed?

Quiz

3. (Single-answer question) Which of the following is not a state of a file in a WORM file system?
 - A. Initial
 - B. Locked
 - C. Deleted
 - D. Expired

Recommendations

- Huawei official websites
 - Enterprise business: <https://e.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/en/index.html>
 - Online learning: <https://www.huawei.com/en/learning>
- Popular tools
 - HedEx Lite
 - Network Documentation Tool Center
 - Information Query Assistant

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。
Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.



Storage System O&M Management



Foreword

- As the cost of storage devices decreases, large-capacity storage devices have been used by more and more enterprises to store data generated by enterprise service application systems and IT systems, such as emails, documents, service data, and data backup. Therefore, effective management of storage devices is critical to the continuity and stability of enterprise services.

Objectives

Upon completion of this course, you will understand:

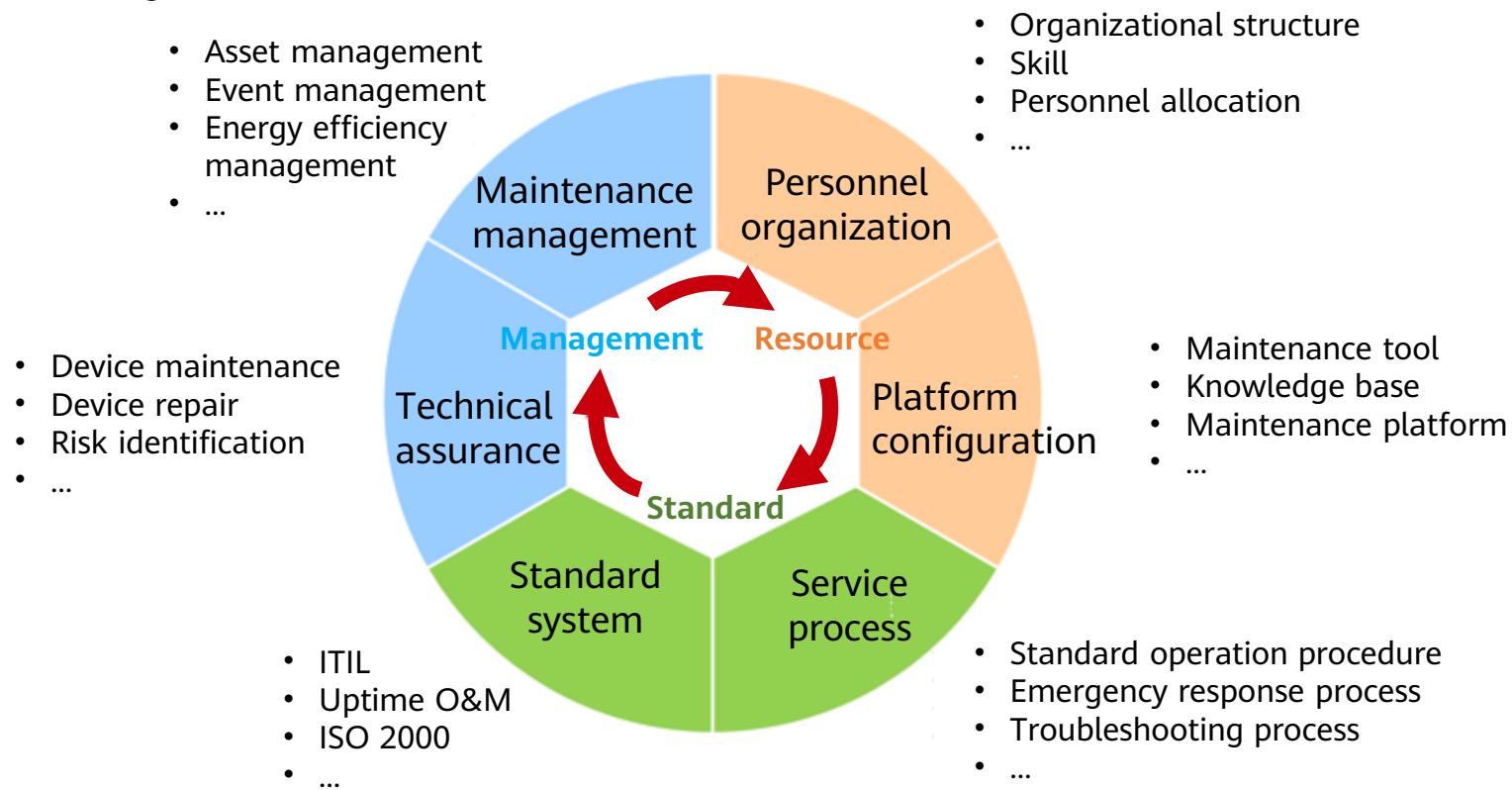
- General O&M management process
- Common storage system O&M management tools
- Process and methods of typical storage system O&M scenarios

Contents

- 1. O&M Overview**
2. O&M Tools
3. O&M Scenarios

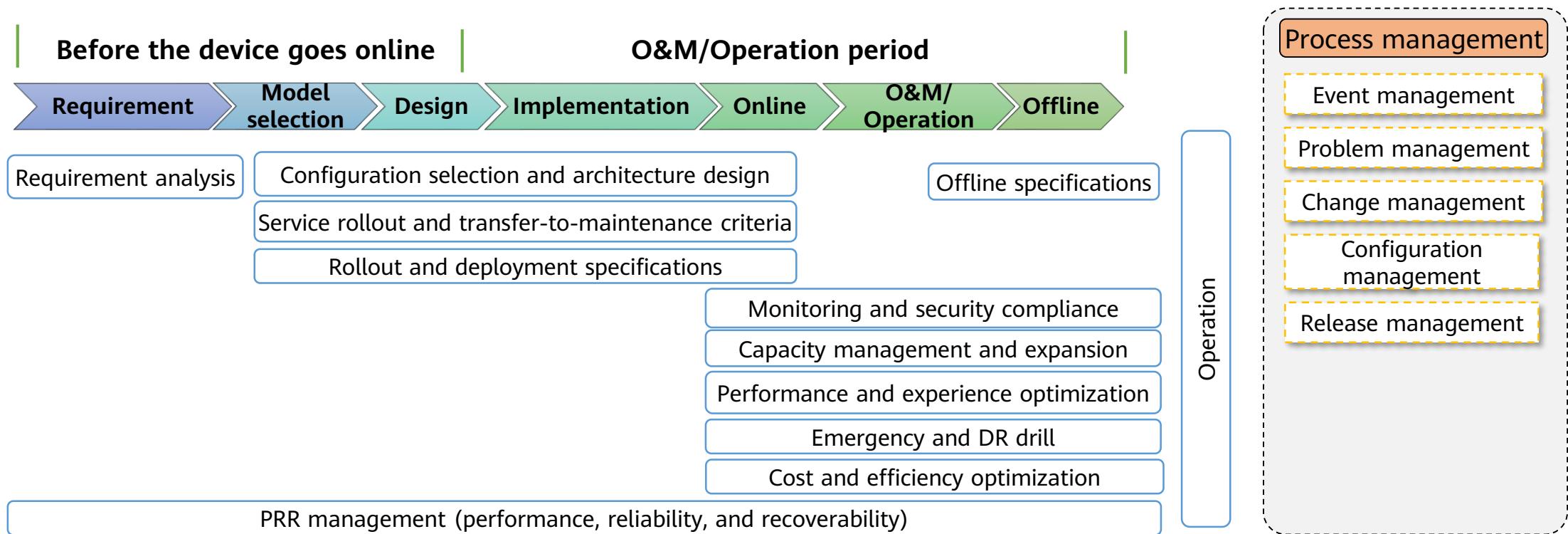
What Is O&M?

- O&M is essentially the operation and maintenance of networks, servers, and services in each phase of their life cycles to achieve a consistent and acceptable status in terms of cost, stability, and efficiency.



How to Perform O&M

- Technical layer: Streamline the O&M lifecycle of each product and identify the key measures of each task.
- Process layer (ITIL process management framework): Change, event, and problem management.



Event Management



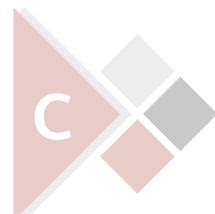
Objective

- ❑ Restore services as soon as possible.
- ❑ Minimize the impact of emergencies on service running.
- ❑ Ensure that the service quality and availability meet the SLA requirements.



Definition

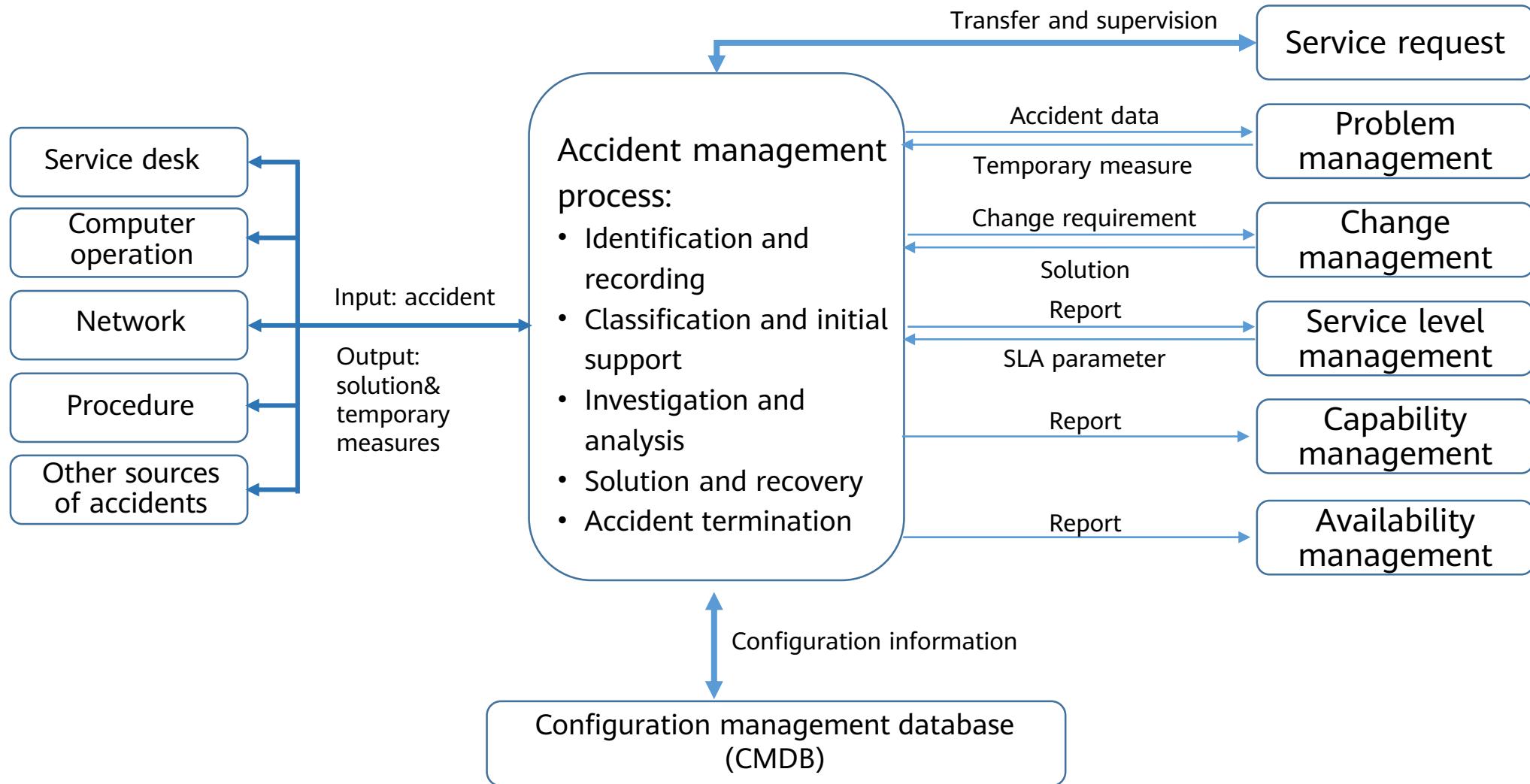
- ❑ Emergency
 - Any event that causes or may cause service interruption or service quality deterioration
 - Hardware faults, software faults, and service request interruption



Task

- ❑ Detection and recording
- ❑ Classification and online support
- ❑ Priority determining based on the impact and urgency
- ❑ Investigation and diagnosis
- ❑ Solution and recovery
- ❑ End
- ❑ Responsibilities, monitoring, tracking, and communication

Event Management Process



Problem Management



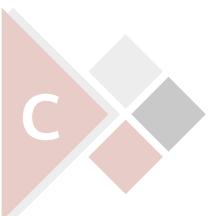
Objective

- Locate the root cause of the problem and take measures to eliminate known errors.
- Minimize the number of emergencies caused by IT infrastructure errors and minimize the negative impact of problems. Prevent the recurrence of emergencies related to errors.



Definition

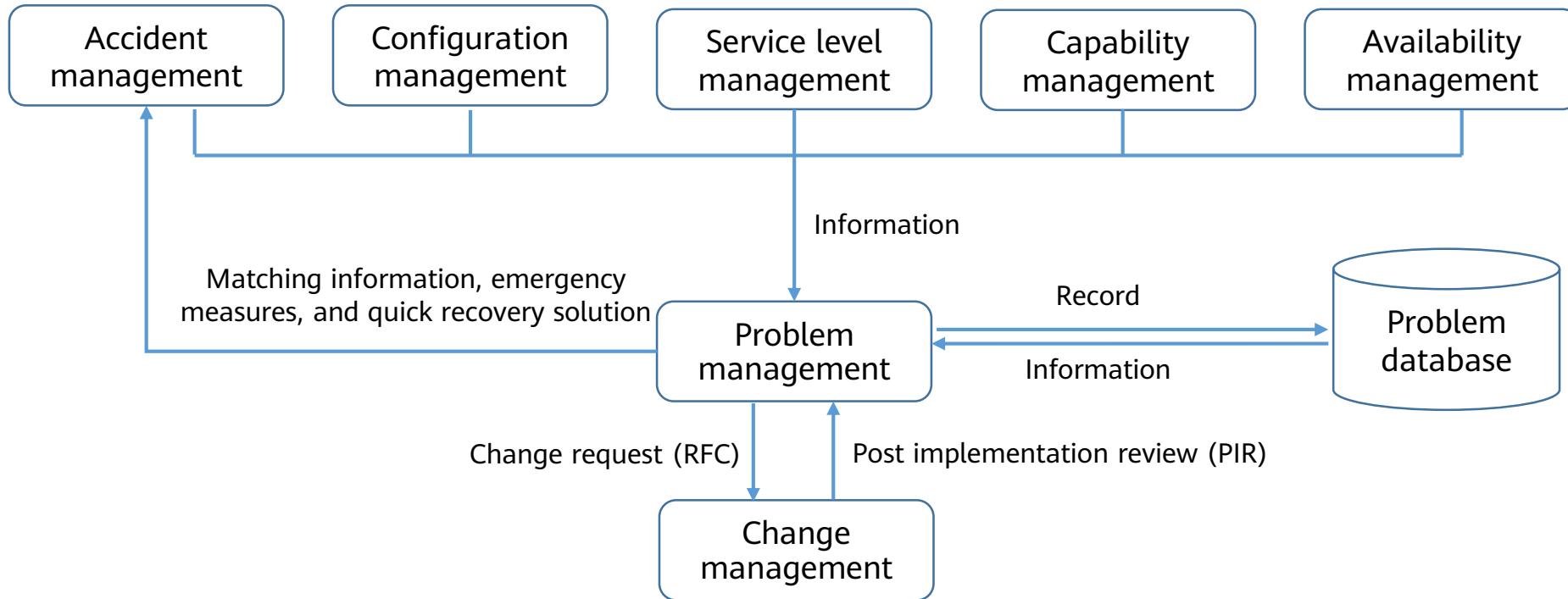
- Problem: obtained from multiple emergencies with the same symptom or a major incident and indicates that an error with unknown causes exists.
- Known errors: The root cause of a problem has been successfully located and a solution has been found.



Task

- Problem control
- Known error control
- Proactive problem management
 - Trend analysis
 - Review of major issues

Problem Management Process

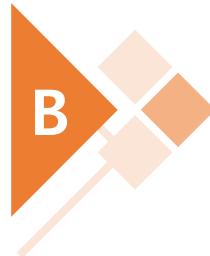


Change Management



Objective

- ❑ Ensure that all changes are effectively controlled and handled through standardized means and processes, and that approved changes are implemented with minimum risks, high efficiency, and high cost-effectiveness.



Definition

- ❑ Change: An action that causes the status of one or more IT infrastructure CIs to change.
- ❑ Standard change (approved in advance)
- ❑ Request for Change (RFC)
- ❑ Forward Schedule of Changes (FSC)
- ❑ Change Advisory Board (CAB)



Task

- ❑ Receive, record, approve, plan, test, implement, and review change requests.
- ❑ Provide the IT infrastructure change report.
- ❑ Propel CMDB modification.

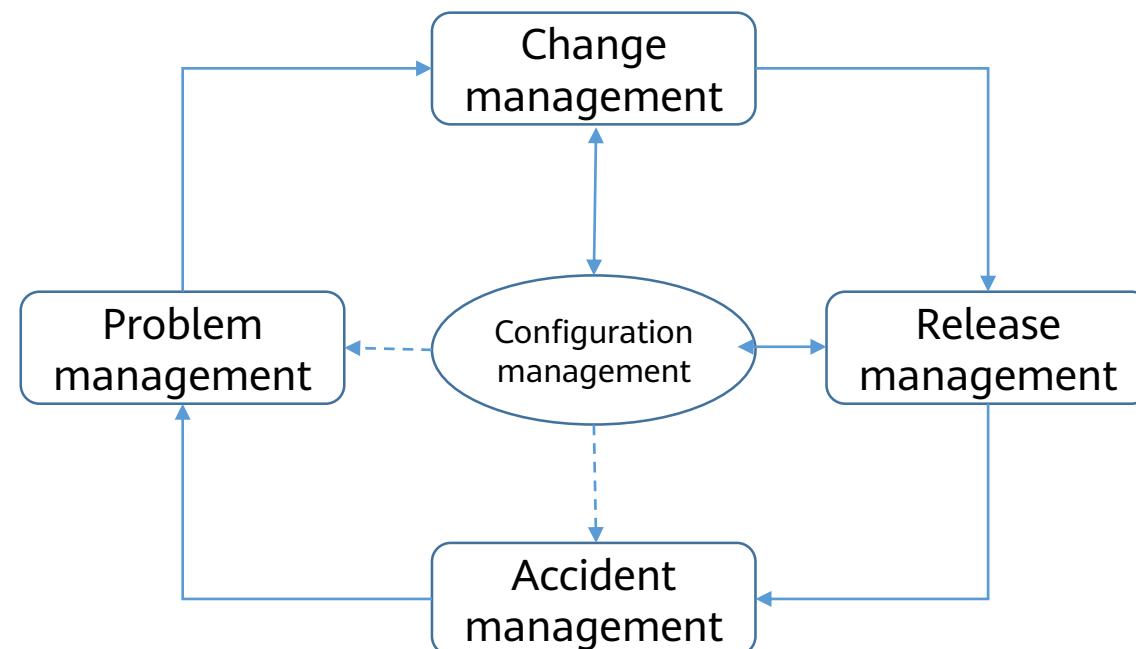
Change Management Process

The input information includes:

- ❑ Change request
- ❑ Data information provided by the CMDB, especially information about the impact of changes
- ❑ Change implementation schedule
- ❑ Capability database provided by capability management and budget information provided by the financial management process

The output information includes:

- ❑ Updated change implementation schedule
- ❑ Signals that trigger the start of configuration management and release management
- ❑ Agenda, minutes, and action items of CAB
- ❑ Change management report



Configuration Management



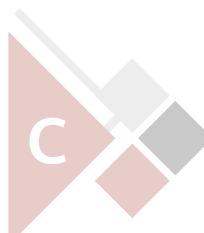
Objective

- Measure the value of all IT assets and configuration items used in organizations and services.
- Provide accurate information about IT infrastructure configuration for other service management processes.
- Support the operation of accident management, problem management, change management and release management.
- Verify the correctness of the configuration records related to the IT infrastructure and correct the detected errors.



Definition

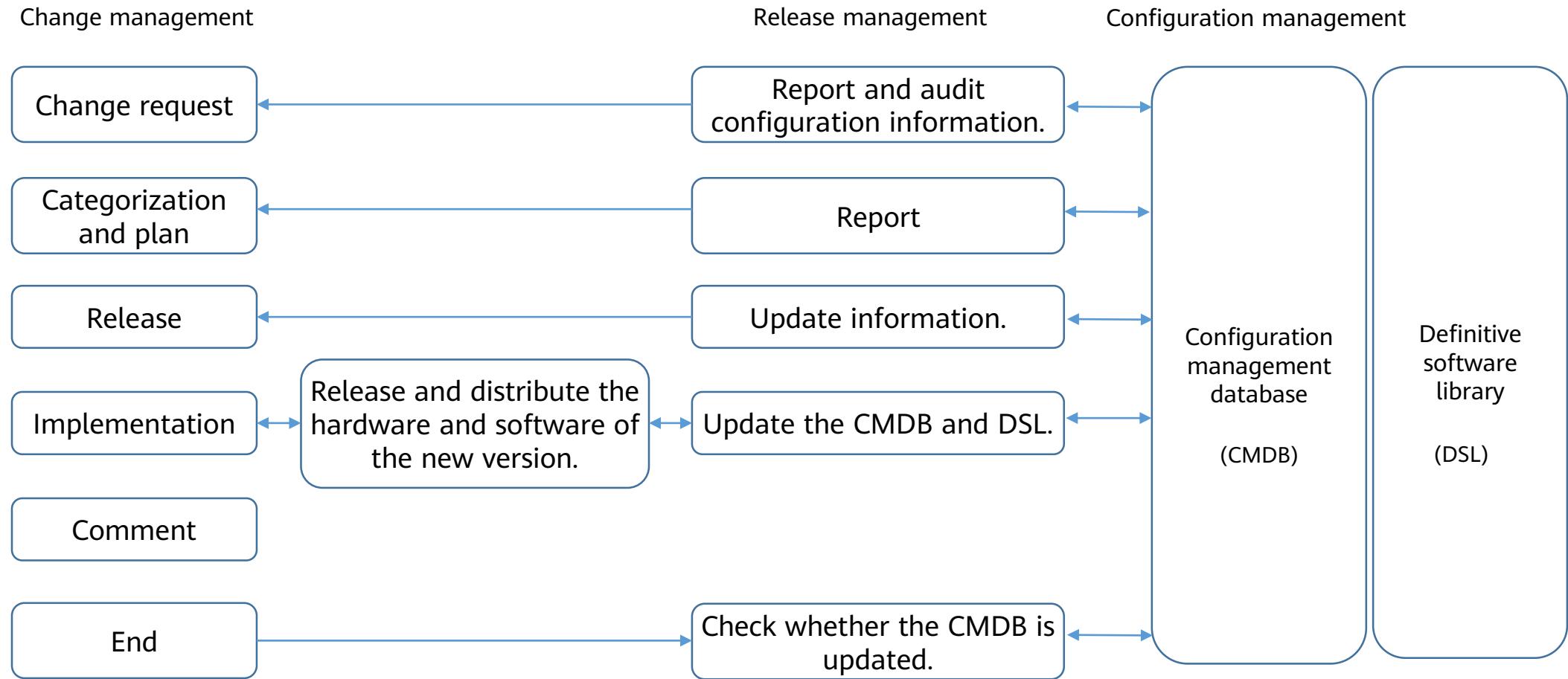
- Identify and define configuration items.
- Plan, define, and manage the CMDB.
- Periodically verify the accuracy and integrity of the CMDB.
- Detailed report of IT assets



Task

- Receive, record, approve, plan, test, implement, and review change requests.
- Provide the IT infrastructure change report.
- Propel CMDB modification.

Configuration Management Process



Publication Management



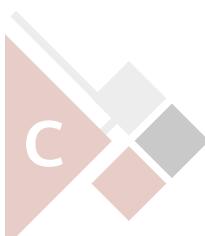
Objective

- Comprehensively assess changes to IT services and ensure that all aspects (including technical and non-technical factors) of a release are considered.



Definition

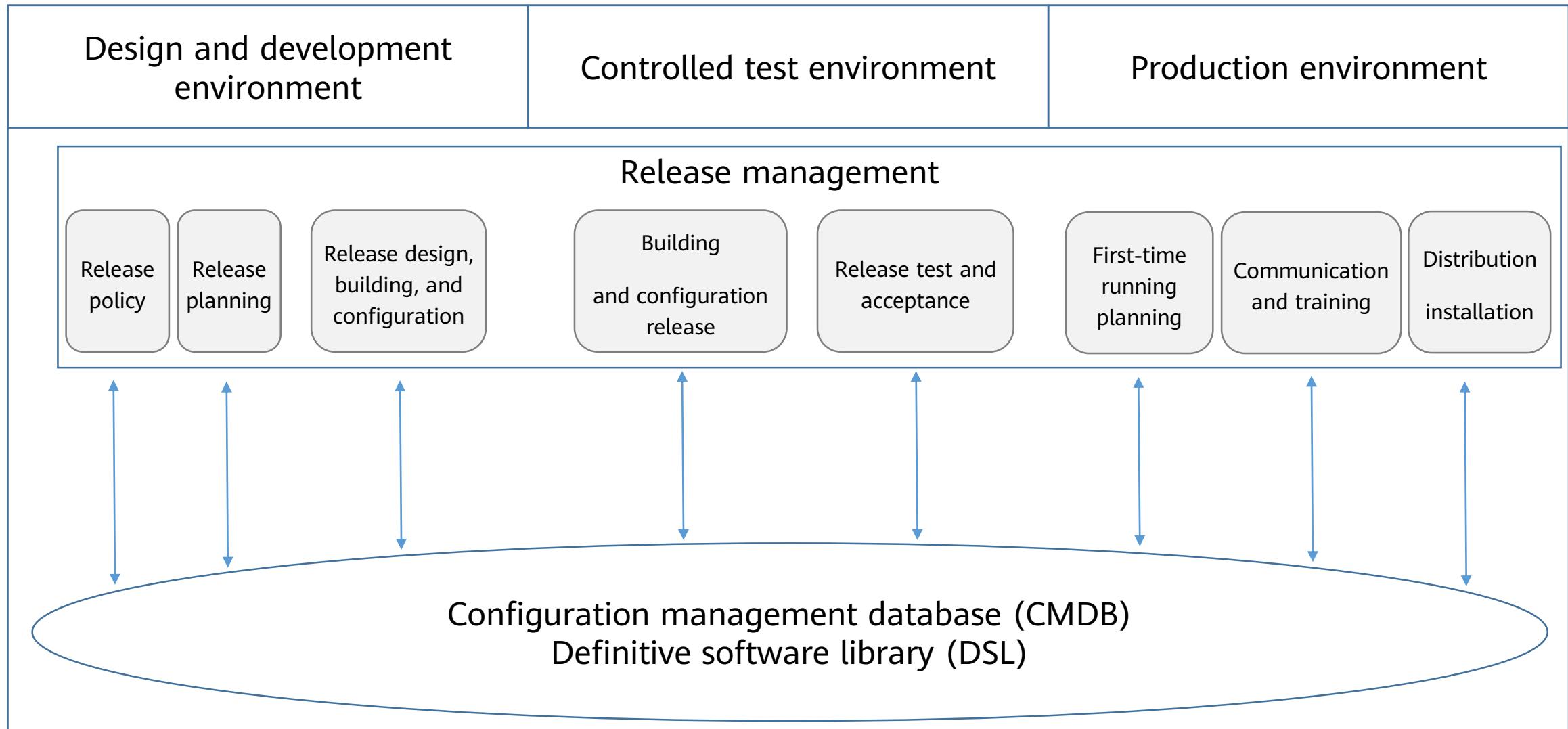
- Release
 - Delta release
 - Full release
 - Package release
- Emergency release
- Release policy



Task

- Release planning
- Design, development, and configuration release
- Release review
- Rollout plan
- Communication, preparation, and training
- Distribution and installation

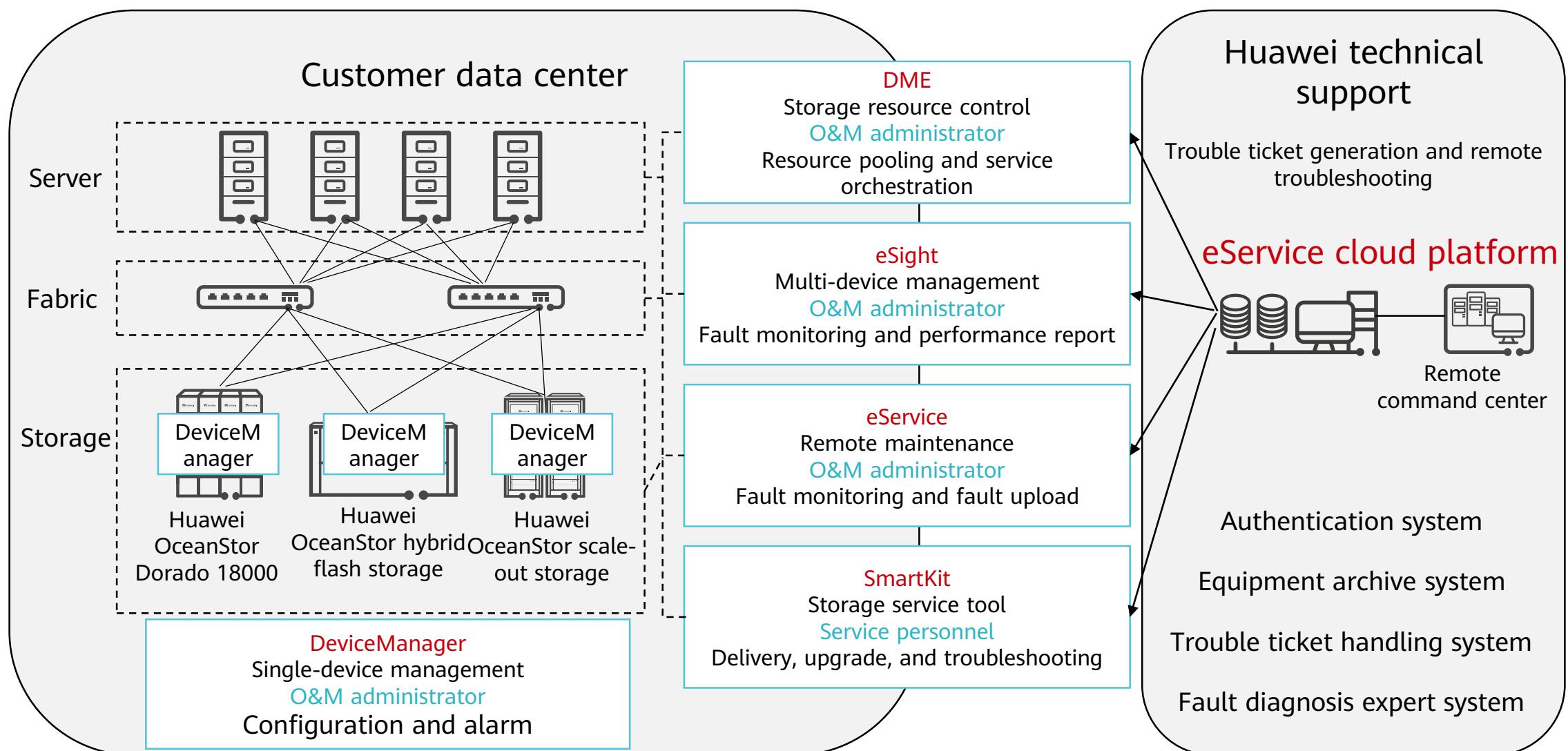
Release Management Process



Contents

1. O&M Overview
- 2. O&M Tools**
3. O&M Scenarios

Components of Huawei Enterprise Storage O&M Systems



Introduction to DeviceManager

- DeviceManager is the single-device management software designed by Huawei for easy configuration, management, and maintenance of storage devices.
- Main software functions include storage resource allocation, user management, data protection feature management, device performance monitoring, and alarm management.

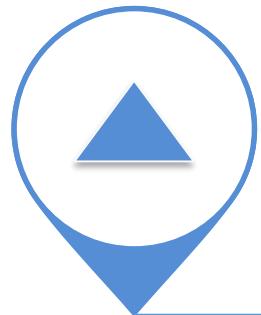


DeviceManager GUI

The screenshot displays the DeviceManager GUI dashboard. At the top, there is a navigation bar with links for Home, Services, Data Protection, Insight, System, and Settings. On the far right, there are icons for notifications (17), a blue circular progress bar (0), a gear (Settings), and a user profile (admin). The main content area includes:

- Alarms:** A summary section showing a server icon, the number 12.136, a red heart icon with the number 60, and device details: Model: Dorado 6000 V3, Version: 6.0.0, ESN: 210235982510G2000016.
- Common Operations:** Buttons for Create LUN Group, Create Host, Create PG, and Create Snapshot CG.
- Effective Capacity:** A summary showing Total capacity of 200.000 TB, with 0.000 MB used and 200.000 TB unused.
- Effective Capacity Trend:** A chart area showing a single data point labeled "No data".

Introduction to SmartKit



1. Unified platform

The desktop tool management platform integrates O&M tools for storage systems, servers, and cloud computing.



2. Scenario-based guidance

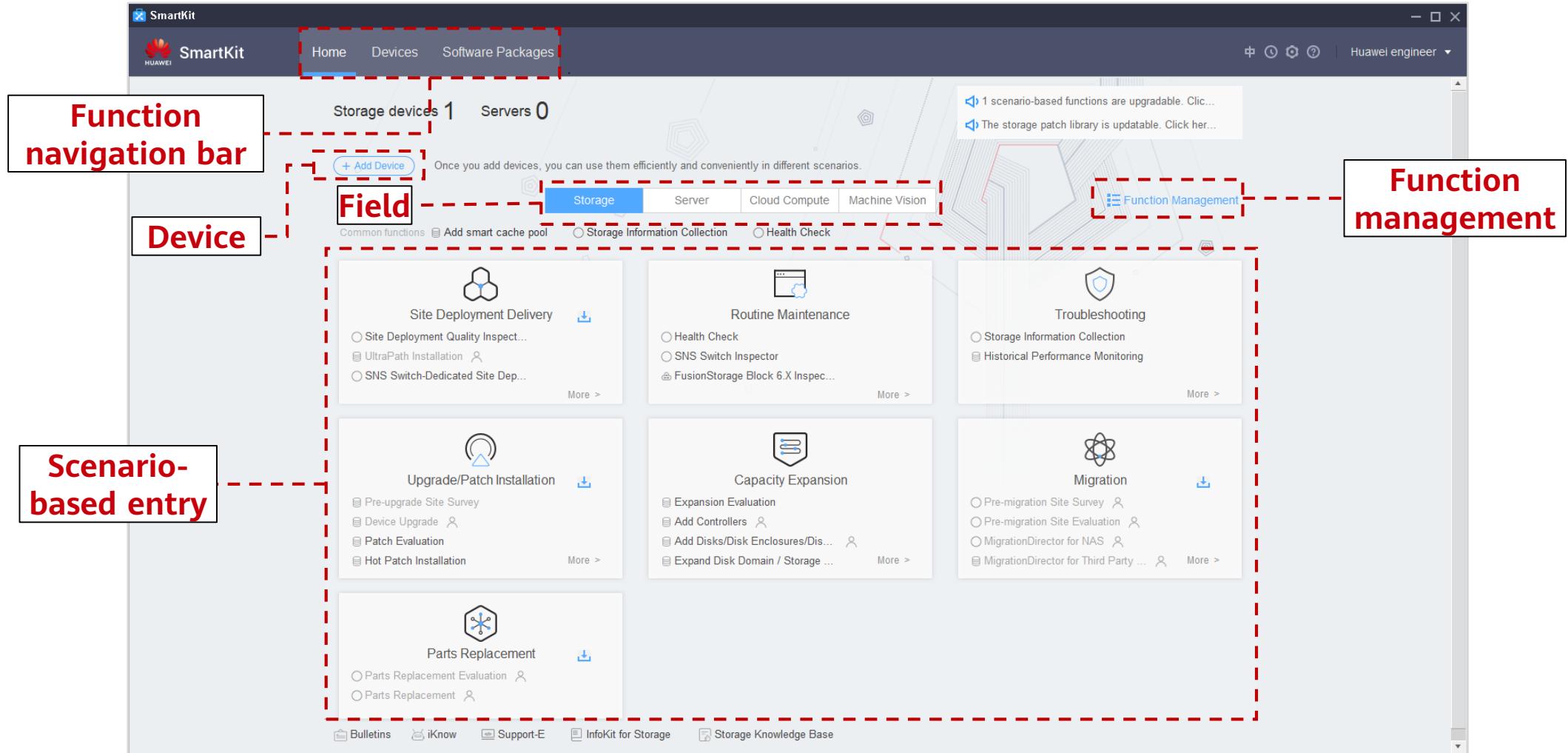
Tools specific to each O&M scenario can be downloaded on demand.



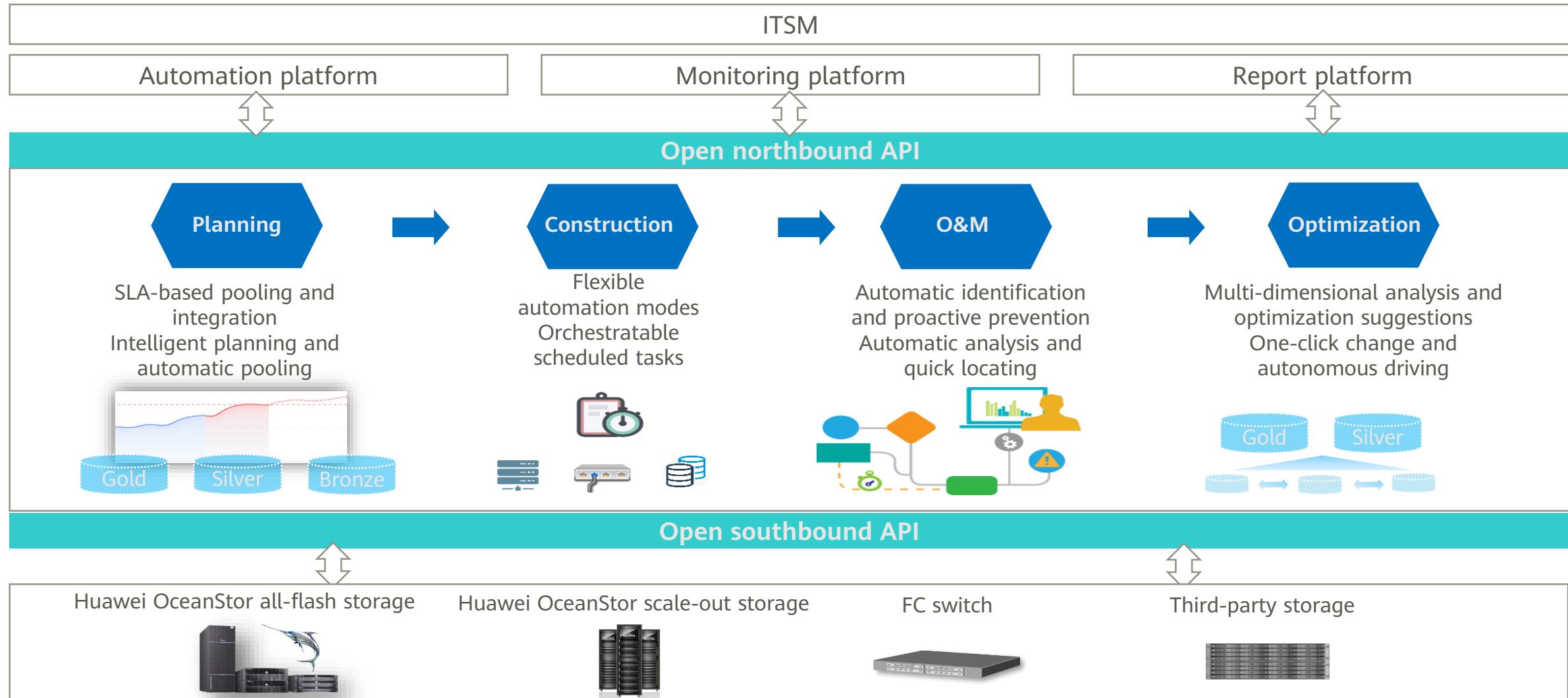
Standardized operations
The wizard guides you through operations based on scenarios in an easy and intelligent manner.

SmartKit GUI

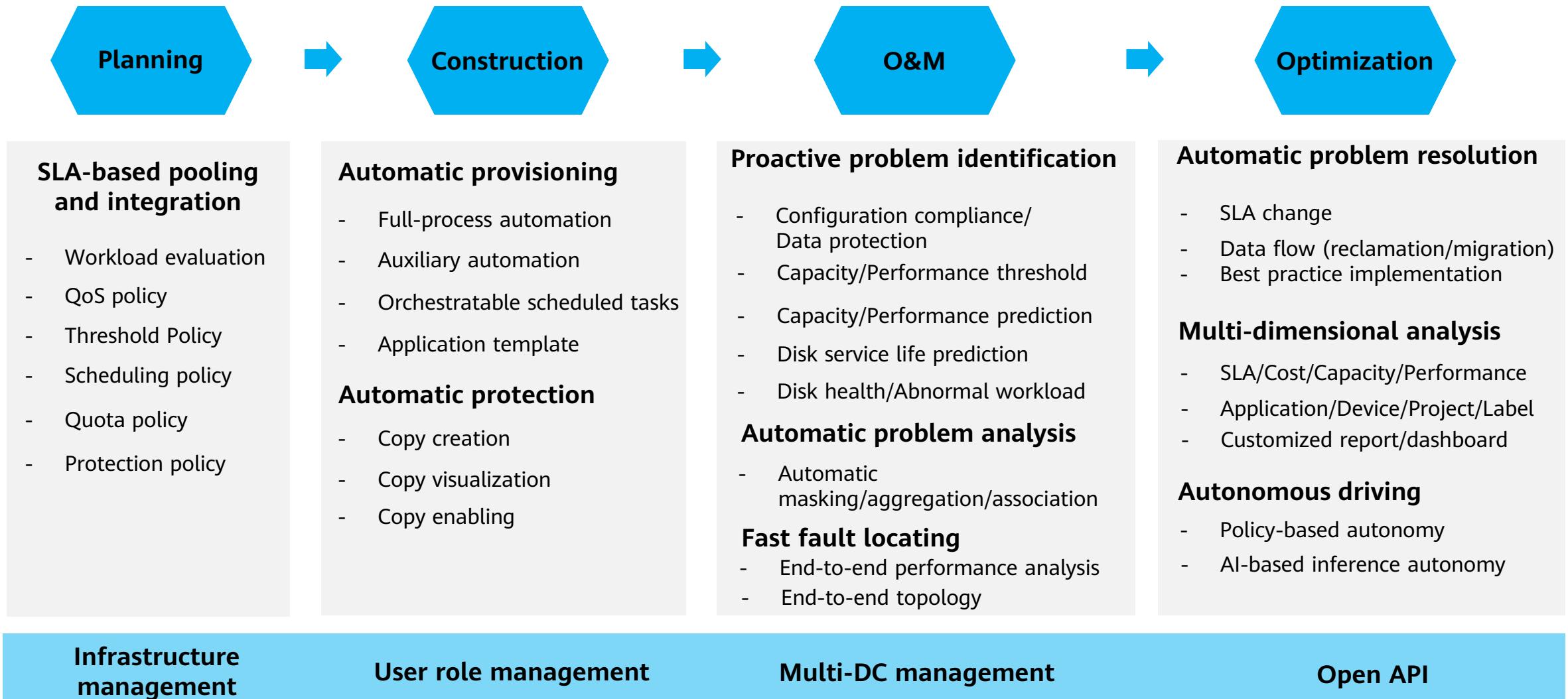
- Standardized and process-based operations in various service scenarios, improving operation efficiency



Introduction to DME Storage

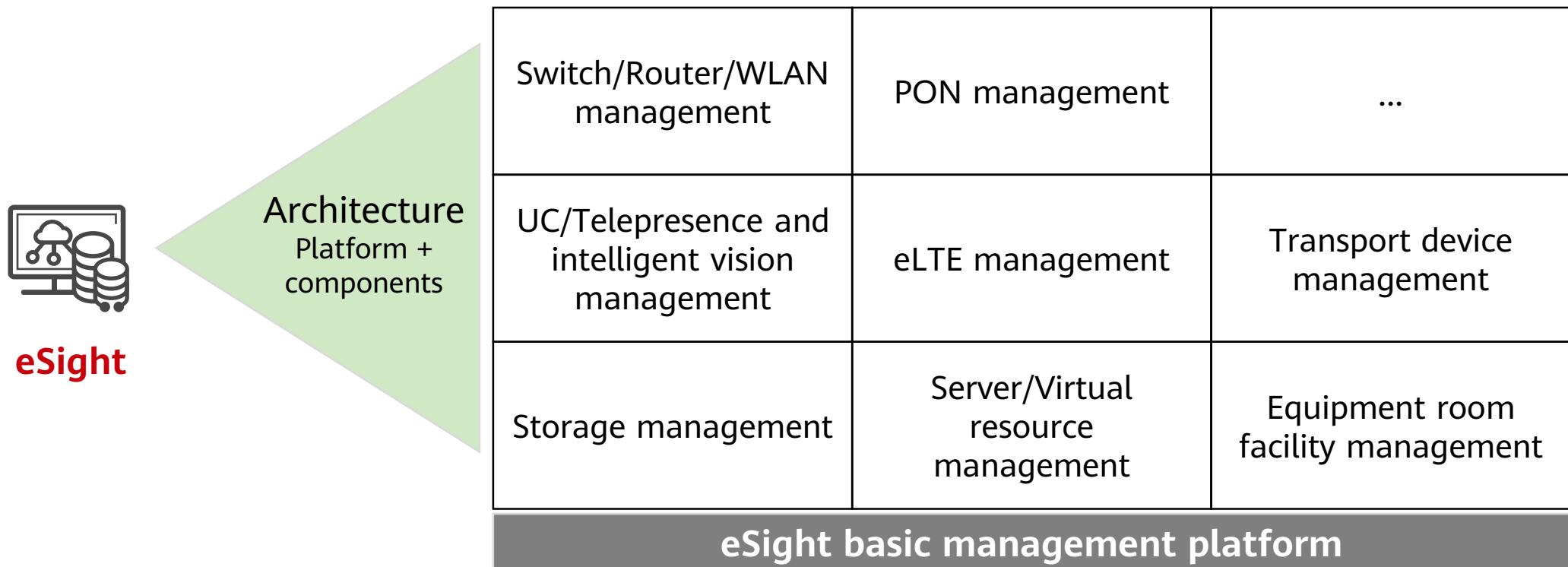


DME Storage Functions and Features

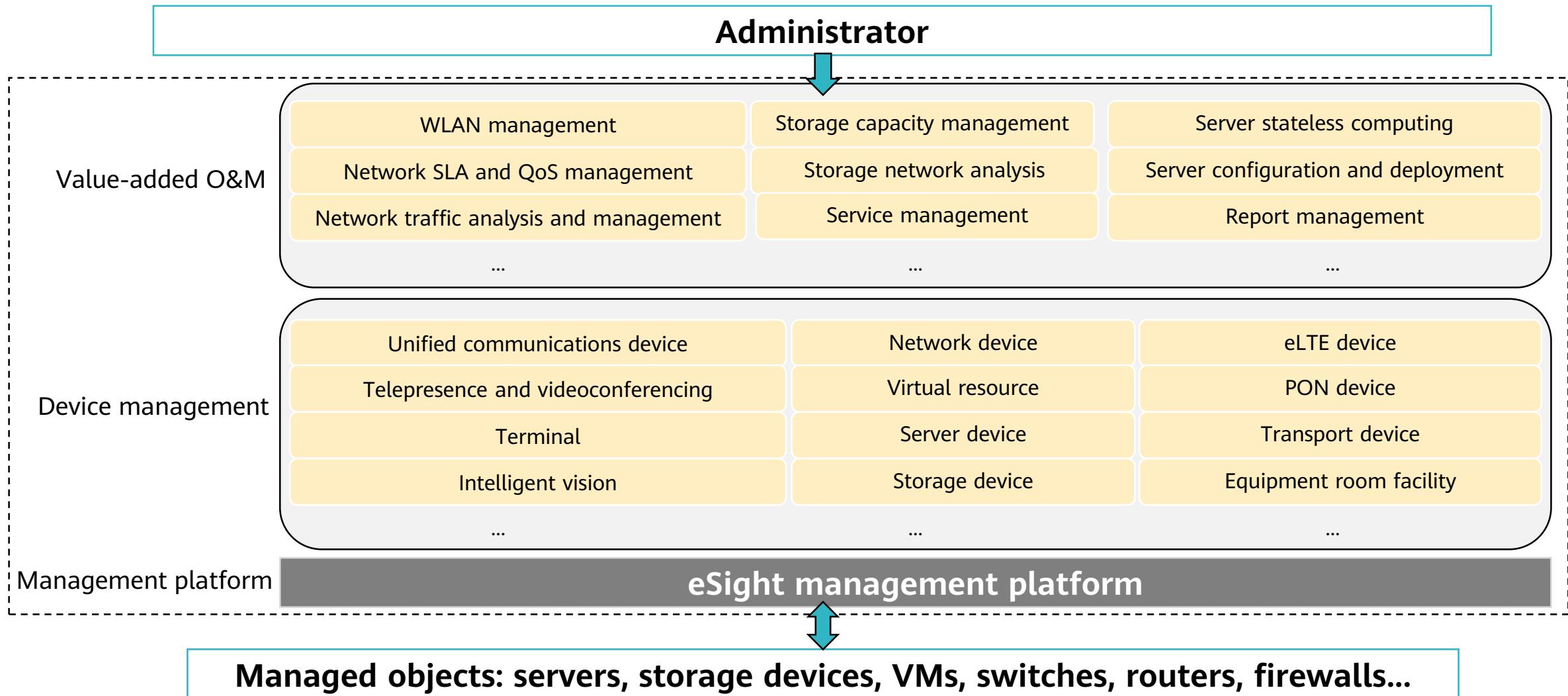


Introduction to eSight

- eSight provides multi-vendor device adaptation for unified network-wide device management, component-based architecture for on-demand construction of enterprise O&M platforms, and lightweight design and web client for lower system maintenance and upgrade costs.

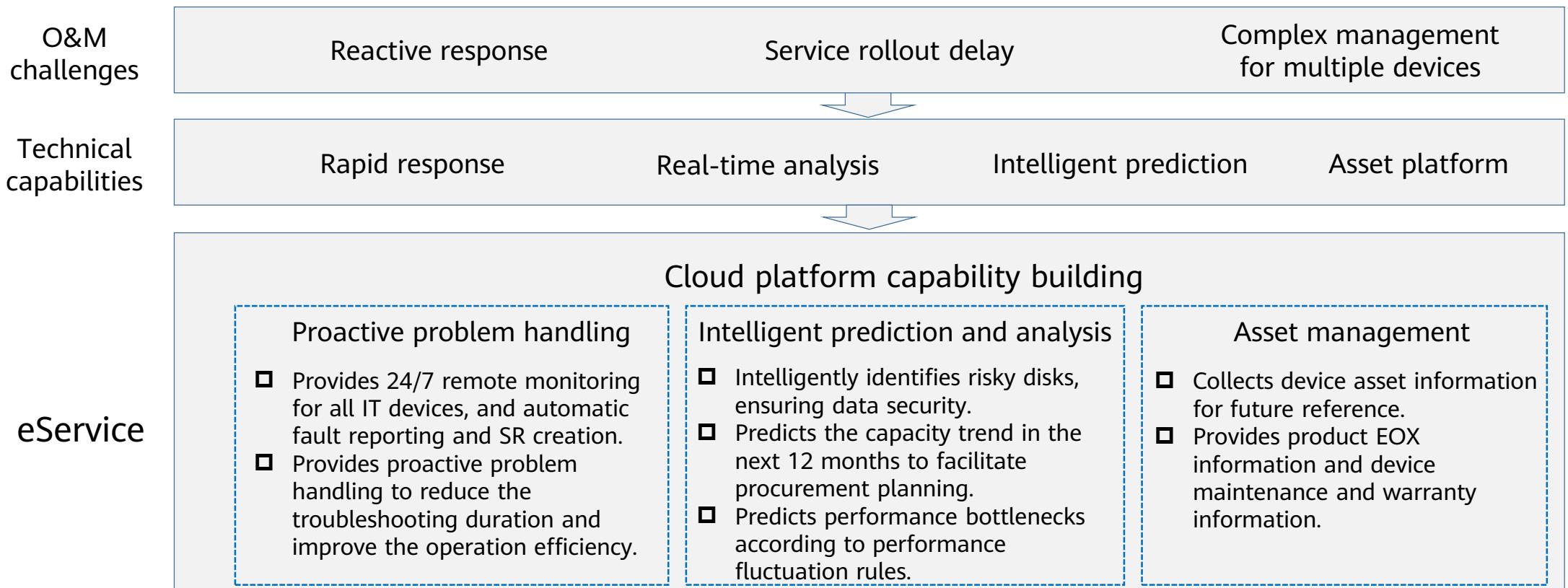


Logical Architecture of eSight

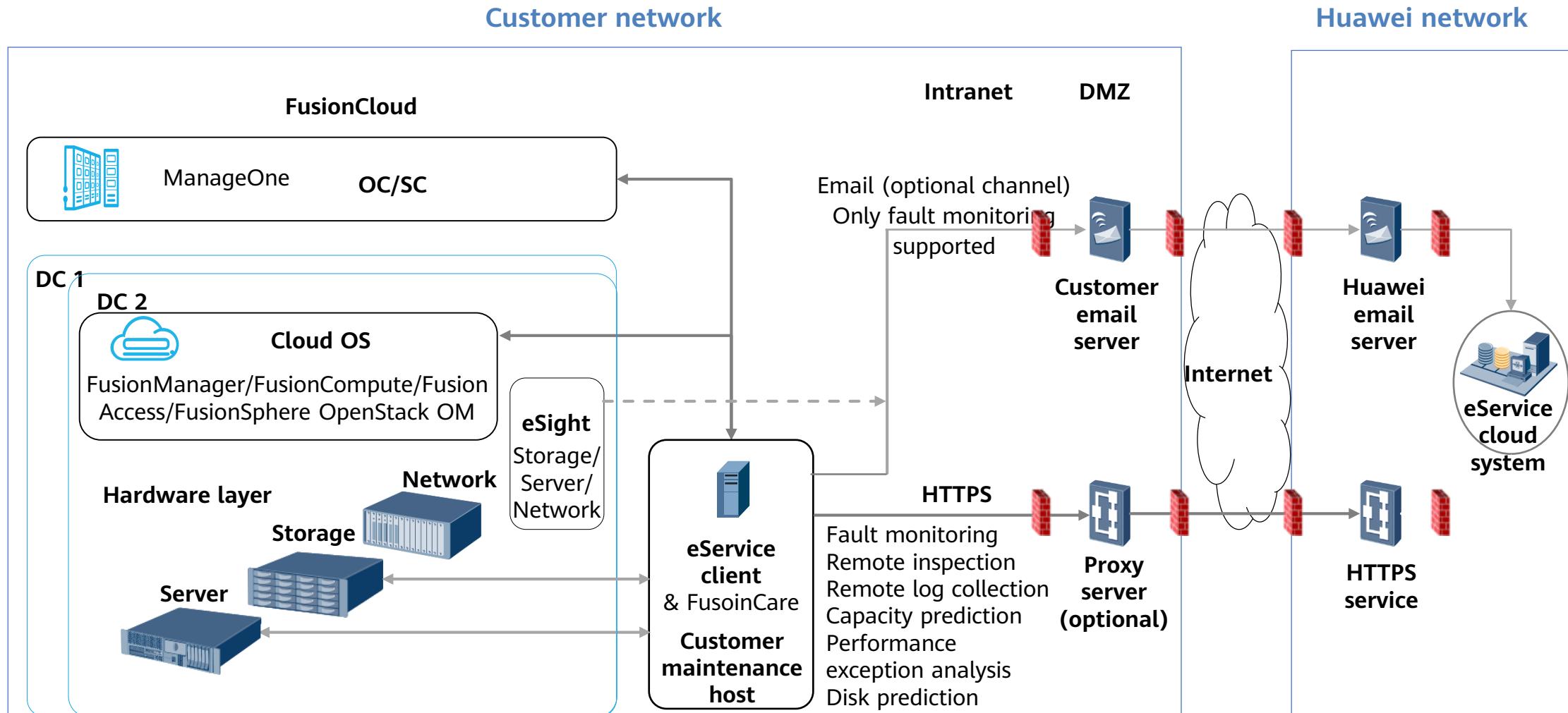


Introduction to eService

- Based on Cloud-Native, Huawei eService cloud intelligent management platform uses big data analysis and AI technologies to provide services such as automatic fault reporting, capacity and performance prediction, and disk risk prediction, preventing potential risks and providing a basis for capacity planning.



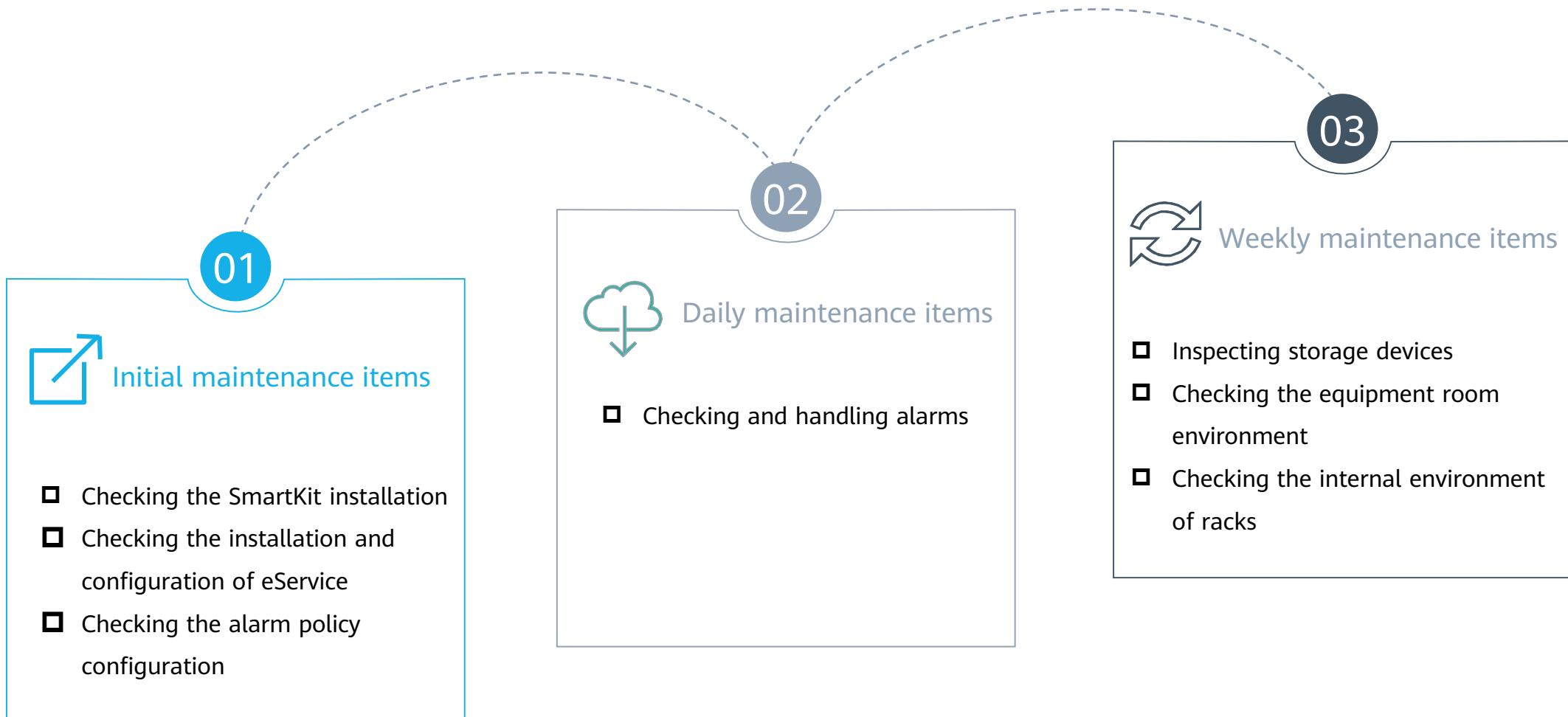
System Architecture of eService



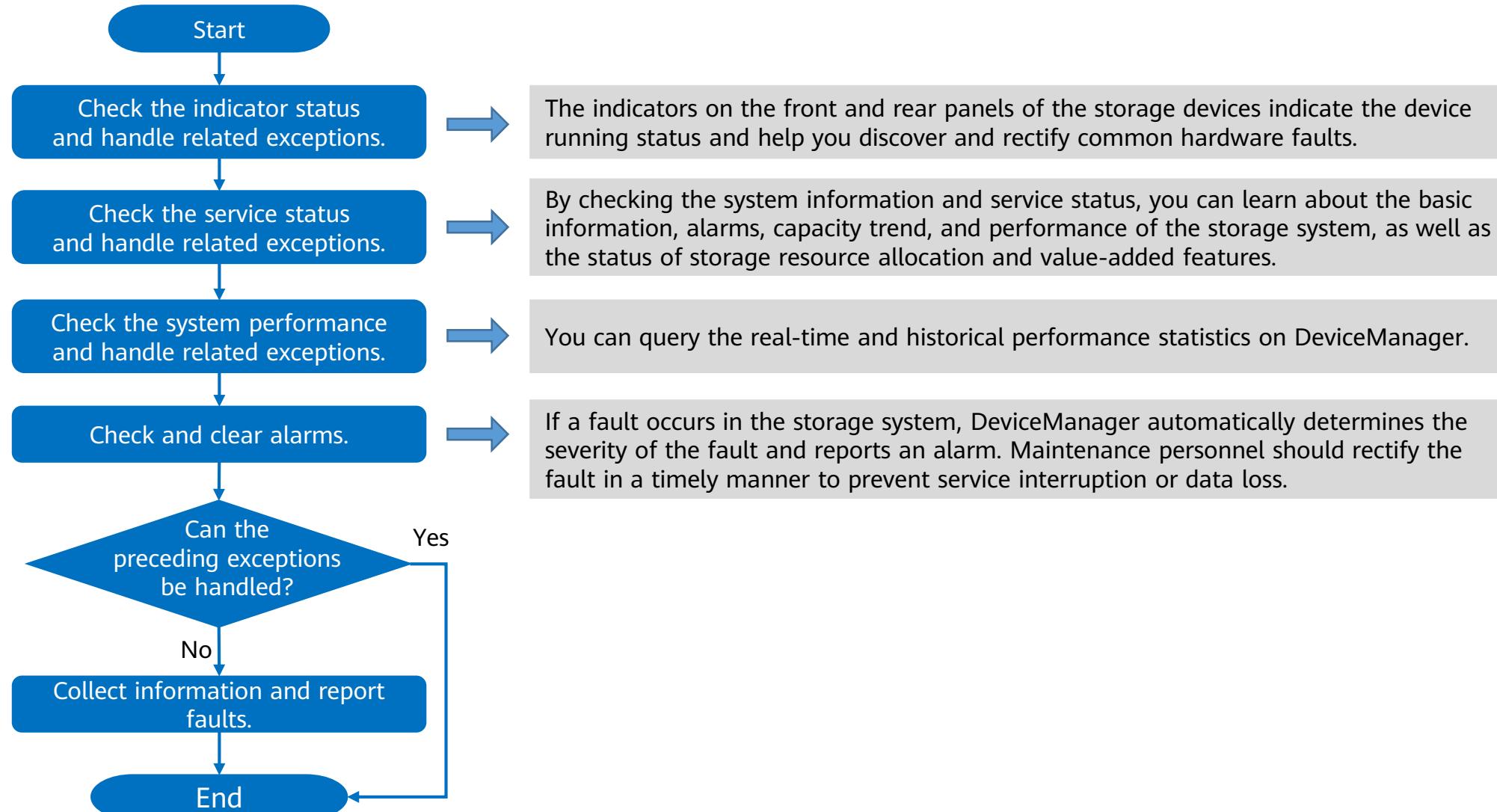
Contents

1. O&M Overview
2. O&M Tools
- 3. O&M Scenarios**

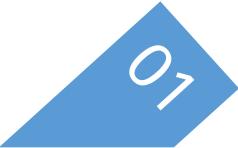
Maintenance Item Overview



Quick Maintenance Process



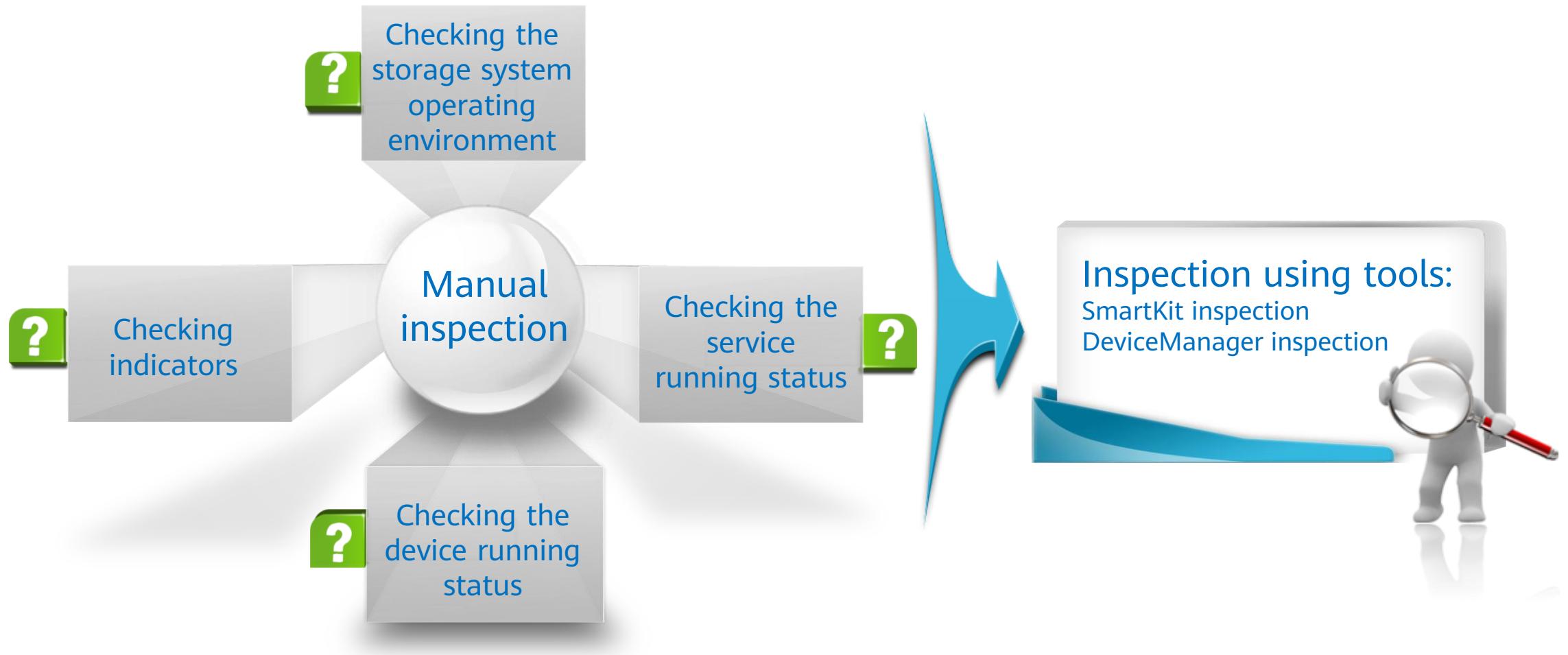
O&M Scenario 1: Inspection



Background

- After the storage devices purchased by company E are deployed, services are deployed and running properly. To ensure the storage security of core devices in the service system, engineer A in the IT department is responsible for the inspection of storage devices. Help engineer A make an inspection plan.

Inspection Method



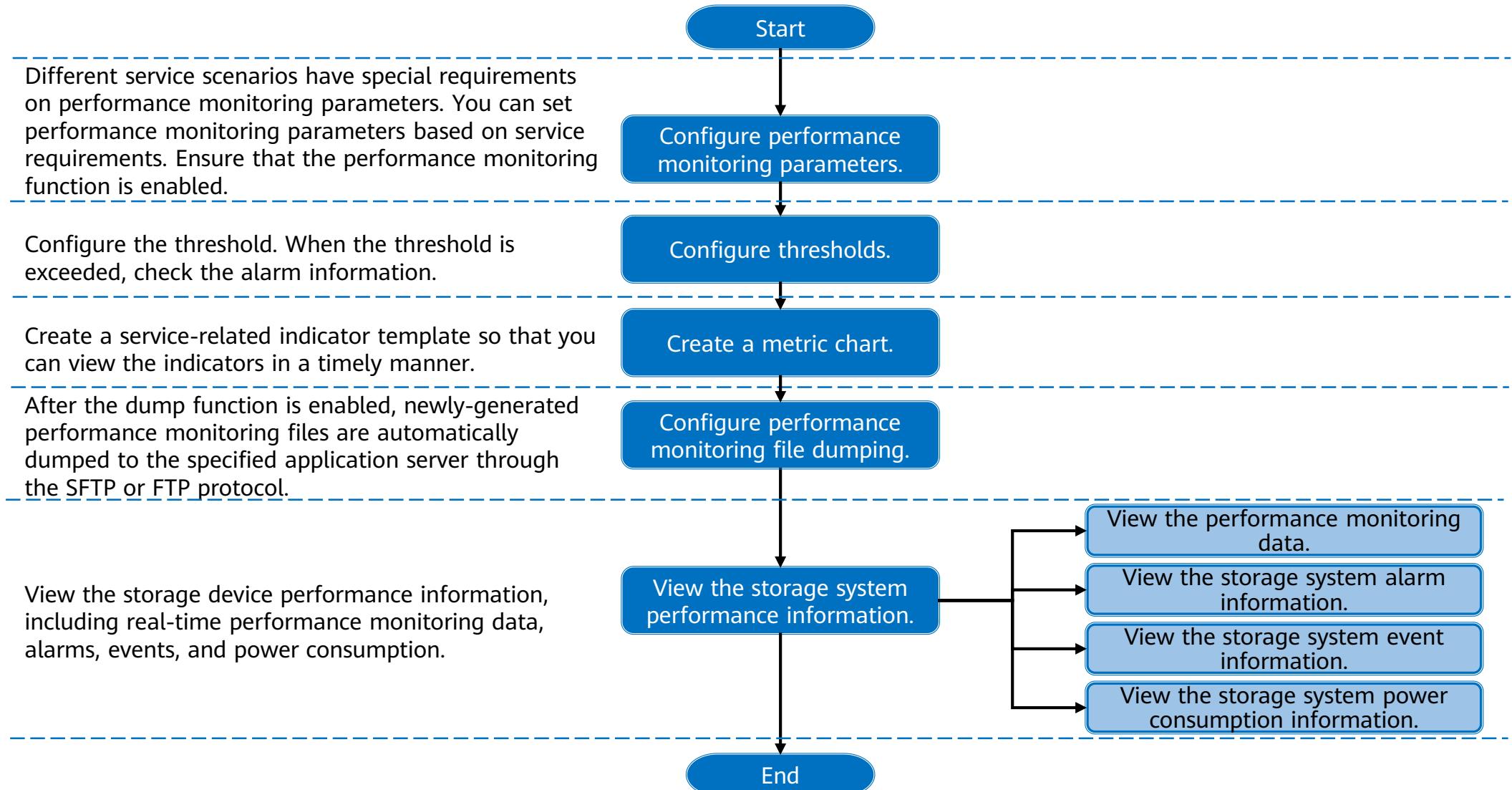
O&M Scenario 2: Performance Monitoring

02

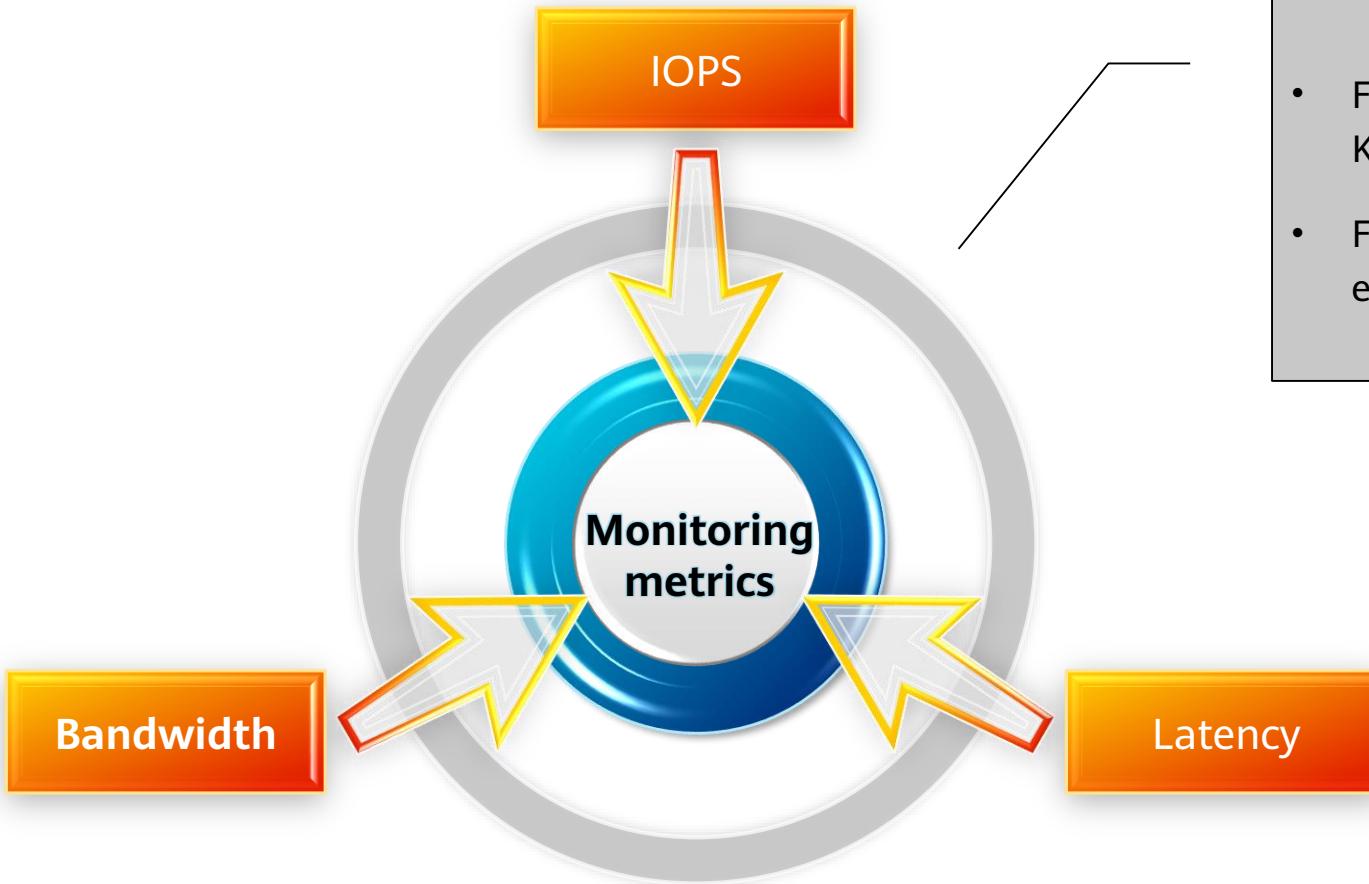
Background

- Company E's business has gone online. To learn about the performance usage of storage devices, engineer B in the IT department is responsible for monitoring the performance of the storage devices. Help engineer B monitor the performance of the storage devices.

Performance Monitoring Process



Performance Monitoring Metrics

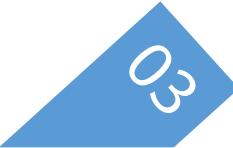


- For applications with an I/O size smaller than 64 KB, mainly focus on the IOPS.
- For applications with an I/O size greater than or equal to 64 KB, mainly focus on the bandwidth.

Performance Metrics

1	Snapshot	1	Logical port	1	Heterogeneous iSCSI link
2	Front-end Ethernet port	2	Host	2	Heterogeneous FC link
3	LUN priority	3	Controller	3	Remote replication consistency group
4	Back-end SAS port	4	LUN	4	FC replication link
5	Front-end FC port	5	Storage pool	5	Remote replication
6	Front-end bond port	6	SmartQoS policy	6	System
7	Disk	7	LUN group	7	Host group

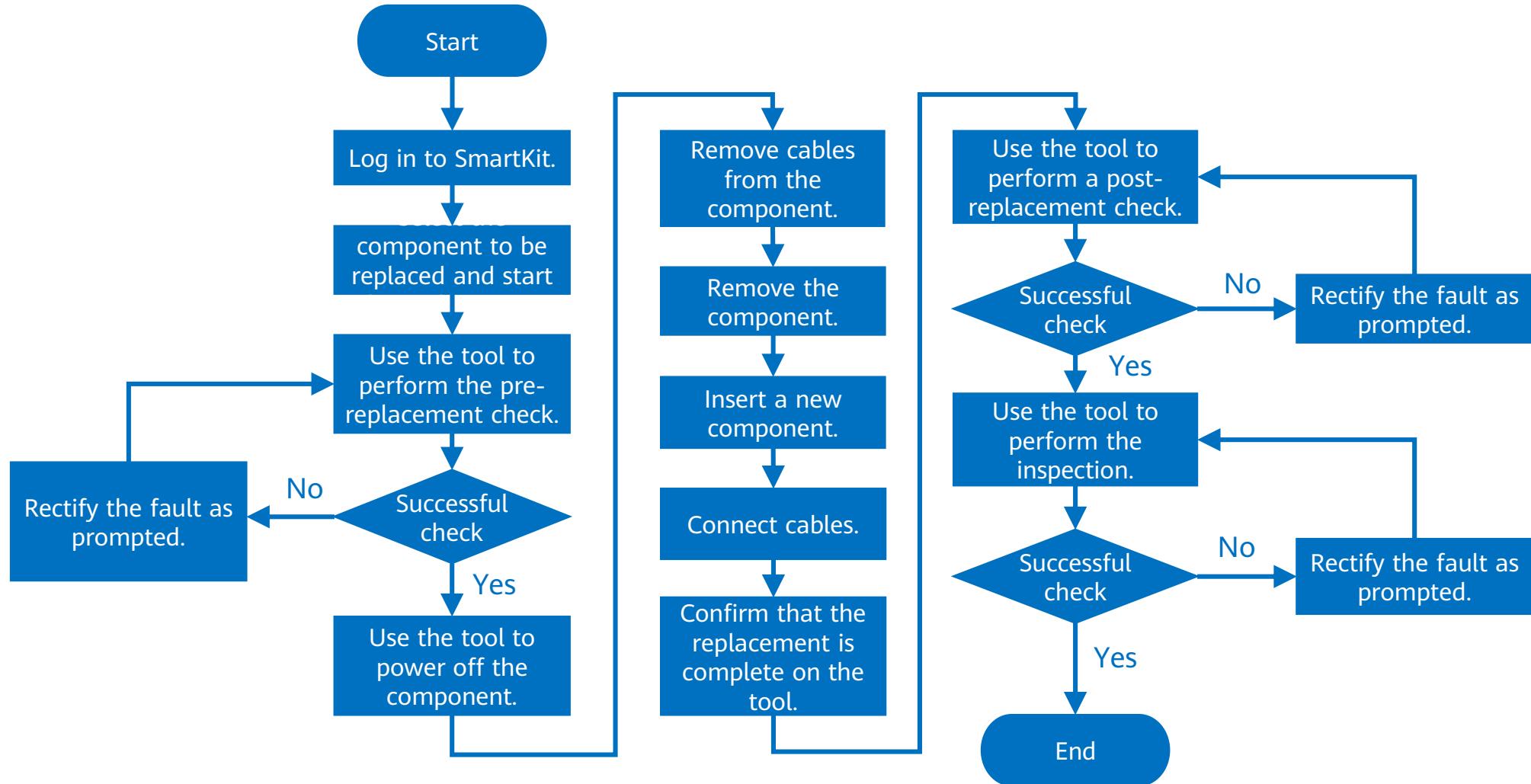
O&M Scenario 3: Parts Replacement



Background

- After the storage system of enterprise E has been running for a period, it reports a disk failure and disk replacement is needed. After receiving the disk replacement request, Huawei technical support engineer C is going to the customer site to perform the replacement.

Parts Replacement Process



Replaceable Parts

FRU	CRU
Controller	Power module
Interface module	BBU module
System subrack	Fan module
Management module	Disk module
Cable	Expansion module
Assistant cooling module	Optical module
Quorum server	-
Data switch	-

Spare Parts Query

- Huawei Spare Parts Query allows users to export the spare part manual and quickly obtain spare part information based on the storage product model, part number, and SN.
- You can log in to <https://support.huawei.com/enterprise/en/index.html> to use the Spare Parts Query tool.

RMA Delivery Status Inquiry

Enter your SR Number, SP Number or Faulty SN

Verification Code

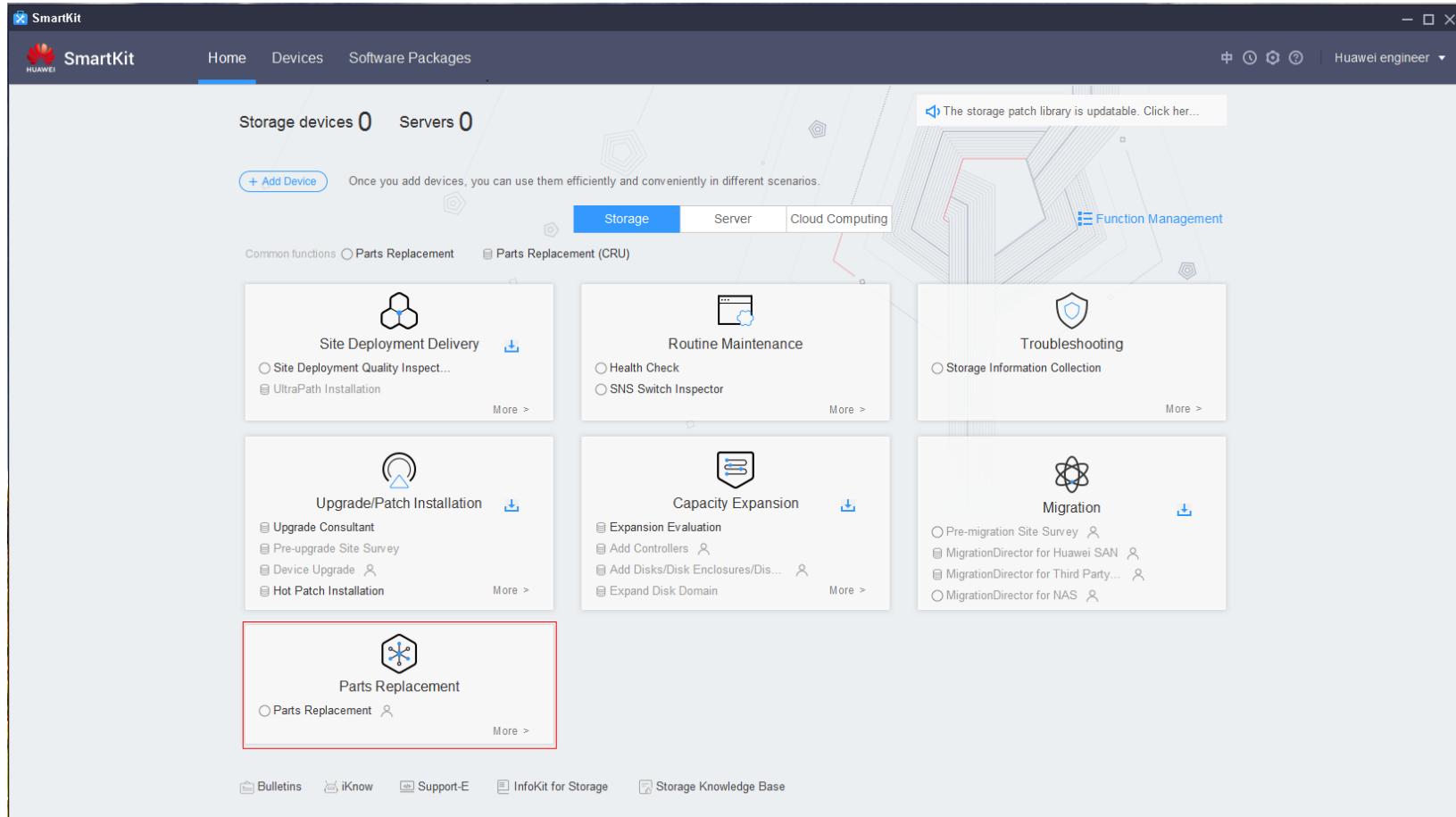
Submit

Key Points for Disk Replacement

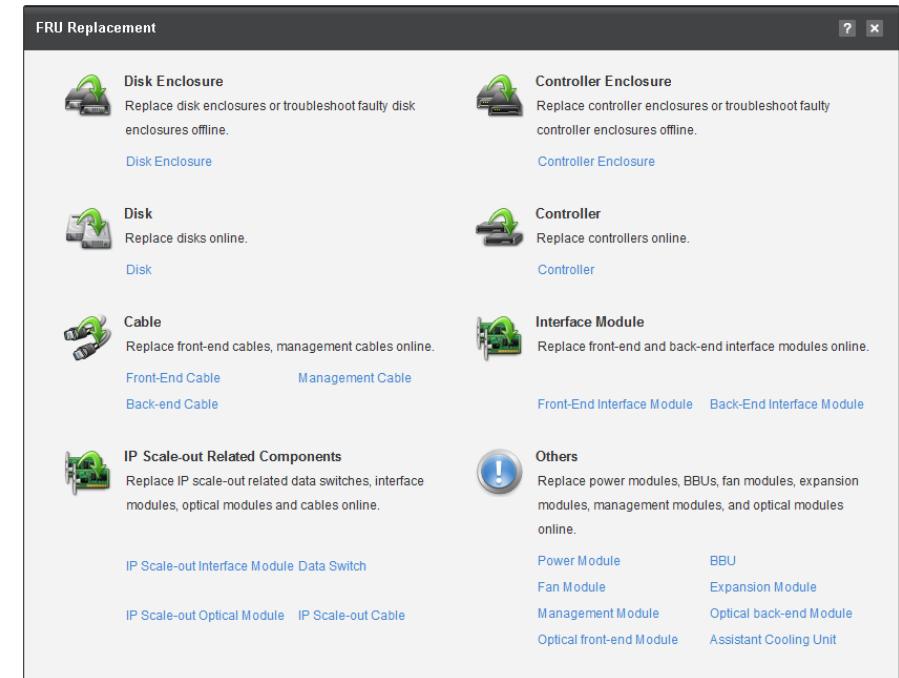
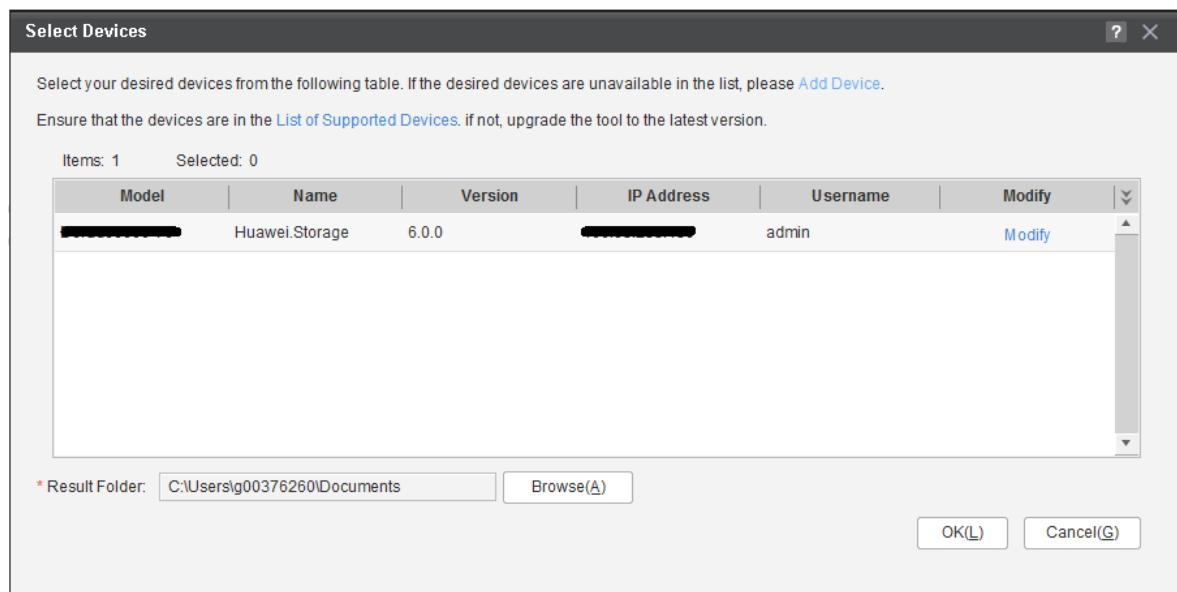
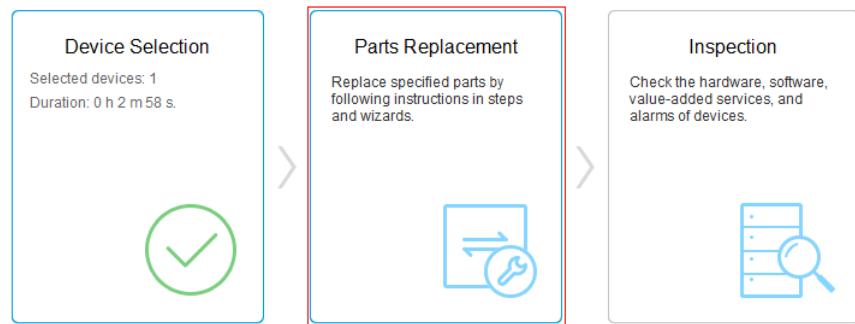
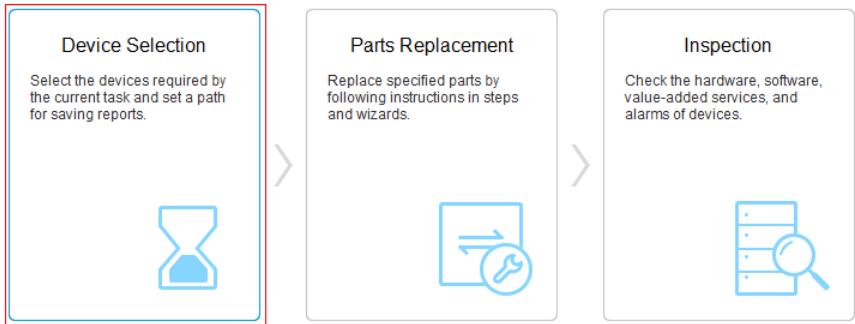
- When handling a disk module, hold only its edge to prevent damage.
- Remove and insert a disk module with even force. Excessive force may damage the appearance of the disk module or cause faults.
- To avoid damaging disk modules, wait at least 1 minute between removal and insertion.
- To prevent data loss, replace only a disk module of which the alarm/location indicator is steady yellow.
- Complete the replacement within five minutes after removing a disk module. Otherwise, the system heat dissipation is compromised.
- Use SmartKit to replace a risky disk (not faulty).
- Ensure that the new disk is inserted into the same slot as the replaced disk. Otherwise, the system may work abnormally.

Disk Replacement Using SmartKit

- Start SmartKit and select the parts replacement tool.



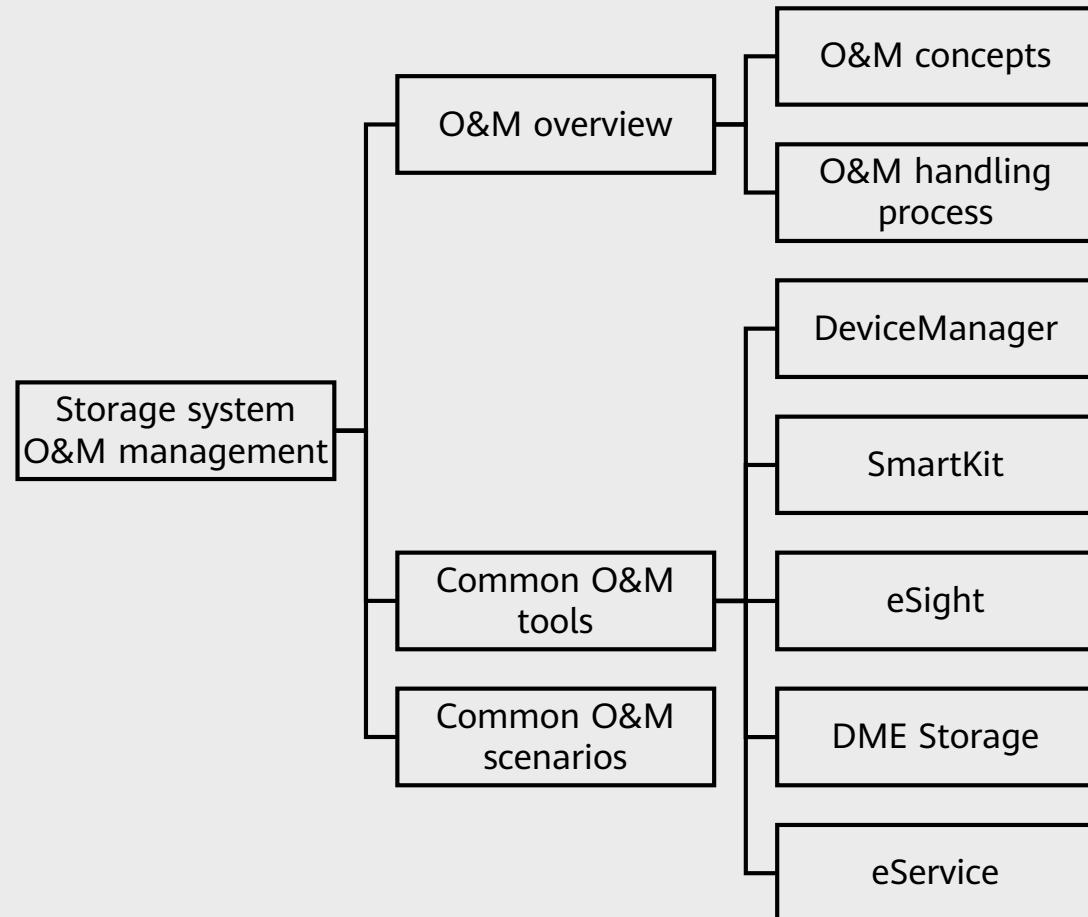
Disk Replacement Wizard



Quiz

1. (Choosing multiple options) Which of the following are common management software? ()
 - A. DeviceManager
 - B. eSight
 - C. SmartKit
 - D. eService
2. (True or false) SmartKit integrates various tools required for deploying, maintaining, and upgrading IT devices, helping product users, and service and maintenance engineers perform precise operations on these devices, simplifying operations and improving work efficiency. ()

Summary



Recommendations

- Huawei official websites:
 - Enterprise business: <http://enterprise.huawei.com/en/>
 - Technical support: <https://support.huawei.com/enterprise/en/index.html>
 - Online learning: <https://www.huawei.com/en/learning>
- Popular tools
 - HedEx Lite
 - Network documentation tool center
 - Information query assistant

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2022 Huawei Technologies Co., Ltd.
All Rights Reserved.

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.

