

Huawei OceanStor Dorado Architecture and Key Technology



Foreword

- Huawei OceanStor Dorado is an intelligent Flash storage series with scale-up/out capability designed to support the business needs of today and tomorrow. It also adopts Huawei-developed hardware platform and the full-mesh SmartMatrix for symmetric active-active services. The systems are built on proprietary flagship hardware and FlashLink® intelligent algorithms — purpose-built for flash media. The systems also adopt end-to-end Non-Volatile Memory Express (NVMe) architecture
- This session describes the hardware and software architecture, functions and features of Huawei OceanStor Dorado .

Objectives

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

- Describe the hardware and software architecture;
- Describe the technical features of superior performance;
- Describe the technical features of high reliability;
- Describe the technical features of high efficiency;
- Describe the technical features of solid security and stability;
- List the typical scenarios

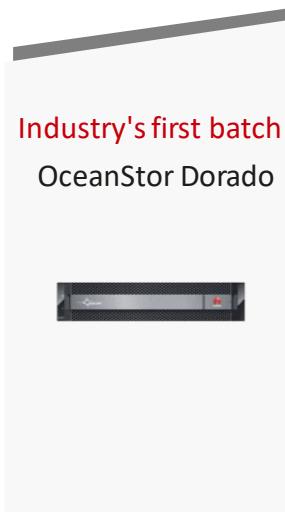
Contents

- 1. Overview**
2. Hardware Architecture
3. Software Architecture
4. Hyper Series Features
5. Smart Series Features
6. Other Key Features

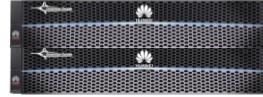
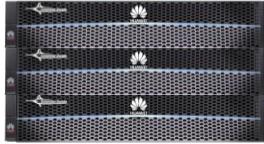
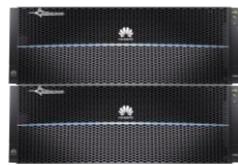
Overview and Objectives

- This section describes the overall introduction to Huawei OceanStor Dorado and the specifications of each model.
- On completion of this section, you will be able to:
 - List the full product portfolio and Highlights of OceanStor Dorado.
 - Describe the models and classifications of OceanStor Dorado;

Huawei OceanStor Dorado All-Flash Storage Sets Benchmarks with Continuous Innovation



Overview of Product Portfolio for OceanStor Dorado

Entry-Level		Mid-Range		High-End	
Dorado 3000 V6		Dorado 5000 V6		Dorado 6000 V6	
Dorado 8000 V6		Dorado 18000 V6			
	Entry-Level	Mid-Range		High-End	
Type	Dorado 3000	Dorado 5000	Dorado 6000	Dorado 8000	Dorado 18000
Height / Controllers per Engine	2 U, 2 controllers	2 U, 2 controllers	2 U, 2 controllers	4 U, 4 controllers	4 U, 4 controllers
Controller Expansion	2-4	2-16	2-16	2-32	2-32
Maximum Disks	1200	1600	2400	3200	4800-6400
Maximum Dual-Controller Effective Capacity (TiB)	500	1024/2048	2048/4096	4096/8192	4096/8192
Dual-Controller Cache	128 GB, 192 GB	256 GB, 512 GB	1024 GB	512 GB, 1024 GB, 2048 GB	512 GB, 1024 GB, 2048 GB
Front-end Ports	8/16/32 Gbit/s FC/FC-NVMe, 10GE, 25GE, 40GE, 100GE, 25/100 Gbit/s NVMe over RoCE				
Back-end Ports	SAS 3.0, 100GE				

Note: For detailed specifications, please refer to the product specification list.

OceanStor Dorado Positioning

Overall Positioning

OceanStor Dorado **all-flash unified storage** sets a new benchmark for the industry with the highest stability, best-in-class SAN and NAS performance, and intelligent, efficient management and O&M.

NAS Positioning

Focus on pure performance scenarios, such as EDA simulation, financial data exchange platforms, and carrier CDRs, with high-performance NAS

Recommended Configurations

Unified SAN and NAS storage
High-end advantages + high-density form factors

Entry-level all flash: min. 128 GB for initial config.



OceanStor Dorado 3000



OceanStor Dorado 5000



OceanStor Dorado 6000



OceanStor Dorado 8000



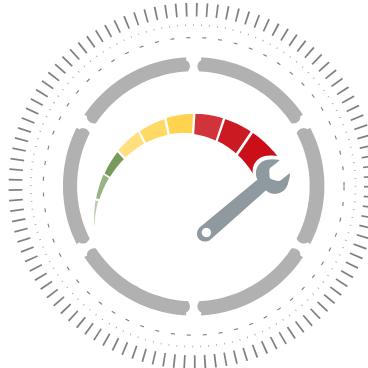
OceanStor Dorado 18000

Tolerates failure of 3 out of 4 controllers/Full interconnection

Building competitiveness based on high-end architecture
Breaking into core applications based on NVMe/intelligent tech

OceanStor Dorado All-Flash Storage Highlights

Ever Fast



Industry-leading performance and latency

21M IOPS and 0.05 ms latency
30% higher NAS performance than industry benchmark



Ever Solid



SmartMatrix fully interconnected architecture for always-on applications

Tolerates failure of 7 out of 8 controllers
Provides active-active solution for SAN and NAS



Intelligent



Intelligent full-lifecycle management

Intelligent O&M
Edge-cloud collaboration



Quiz

1. (True or False) As next generation Unified Storage, OceanStor Dorado can support SAS SSD ,NVMe SSD and HDD.
2. (Multiple-choice) Which statement is true about OceanStor Dorado?
 - A. OceanStor Dorado can deliver 21 million IOPS, 0.05 ms latency, and 30% higher NAS performance than the industry benchmark.
 - B. OceanStor Dorado can tolerate the failure of 7 out of 8 controllers and provides the industry's only active-active solution for SAN and NAS.
 - C. OceanStor Dorado supports SmartMatrix fully interconnected architecture for always-on applications.
 - D. OceanStor Dorado supports Intelligent O&M and The edge-cloud collaboration provides a smarter, simpler way to manage storage resources.

Ever Fast

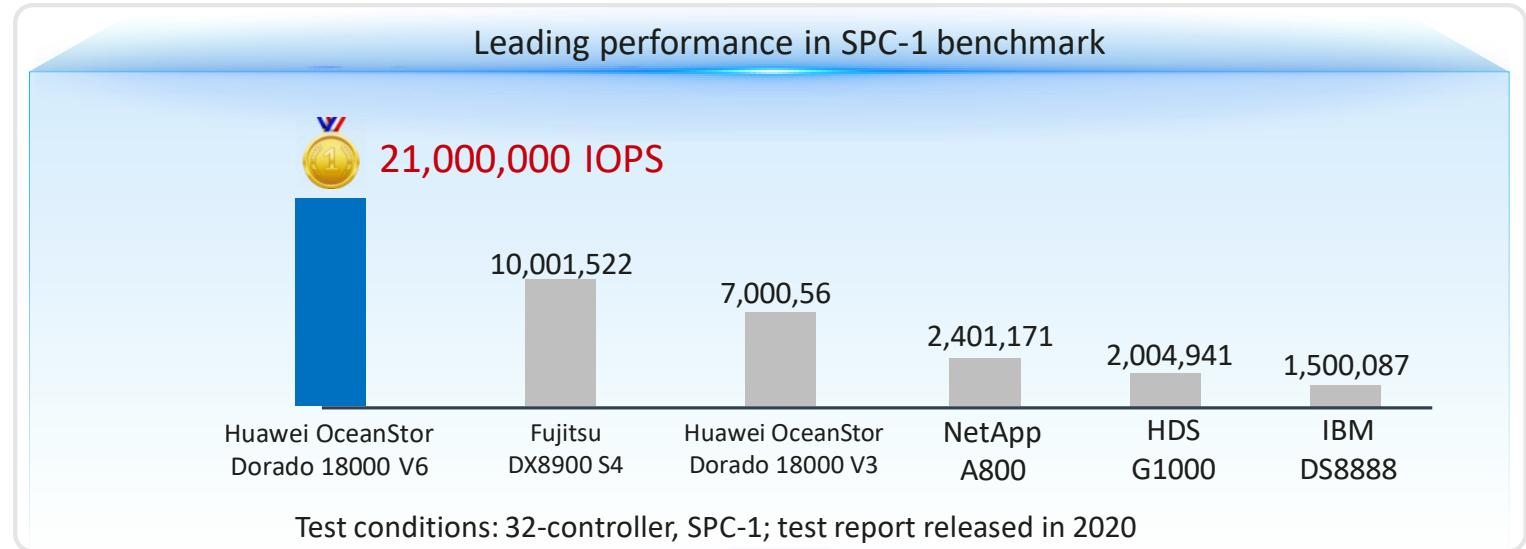


HUAWEI

OceanStor Dorado High-end: Innovative E2E Acceleration Platform Achieves 21M IOPS

2x Better

Than the
Second-Best Player



OceanStor Dorado Middle-Range: Innovative Software and Hardware Design Doubles Performance of Industry Average

5x Better User Experience

VDIs: 80% faster application response



Test conditions: dual-controller, 100 x 3.84 TB SSDs, 8 TB per LUN,
50 GB per VDI

Online transactions 5x more TPS

OceanStor Dorado 6000  57,000 TPS

Vendor E AFA  11,500 TPS

Test conditions: dual-controller, 40 x 3.84 TB SSDs, SwingBench OE2 transaction simulation system

*TPS: transactions per second

Report queries: 33% shorter batch processing



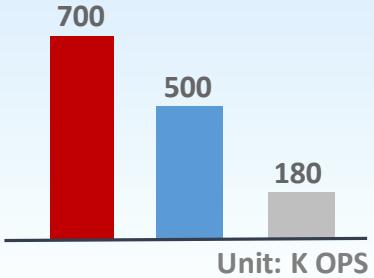
Test conditions: dual-controller, 40 x 3.84 TB SSDs, report query simulation system, 3 TB data

OceanStor Dorado: 30% Higher Performance for NAS scenario

Leading Performance in Diverse **NAS** Scenarios

- █ OceanStor Dorado 18000
- █ Vendor N A Series
- █ Vendor D F Series

40% higher in large files and small blocks



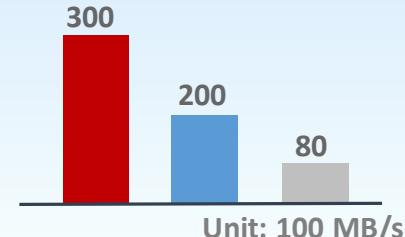
Test conditions: dual-controller, 512 x 20 GB files, 8 KB block size, 7:3 mixed read/write

30% higher in small files and small blocks



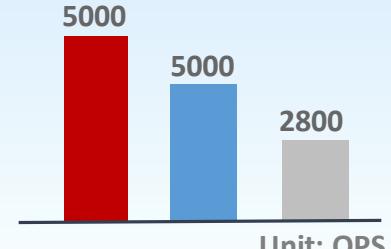
Test conditions: dual-controller, 40 million x 8 KB files, 5-level directories, 8 KB block size, 7:3 mixed read/write

50% higher in bandwidth-intensive scenarios



Test conditions: dual-controller, 512 x 20 GB files, 1 MB I/O sequential read

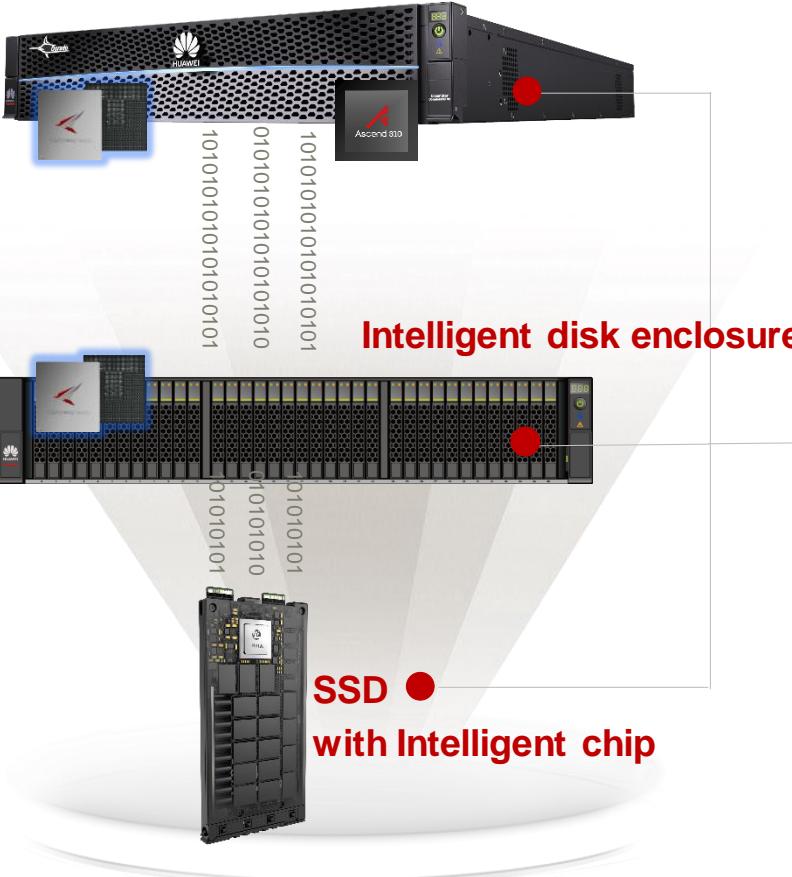
Equivalent with industry benchmark in operational I/O scenarios



Test conditions: dual-controller, and mkdir/rmdir/create/delete/readdirplus

Uncompromising Ever Fast for Innovative FlashLink® Intelligent Algorithms

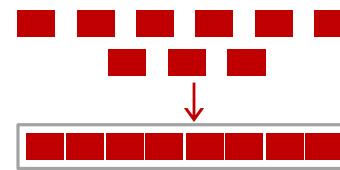
Controller enclosure with Intelligent chips



FlashLink® Intelligent Algorithms

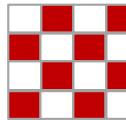
Large-block sequential writes reduce write amplification

Discrete writes of multiple small blocks

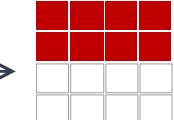


Independent metadata partitioning reduces garbage collection

Mixed

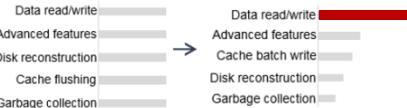


Partitioning



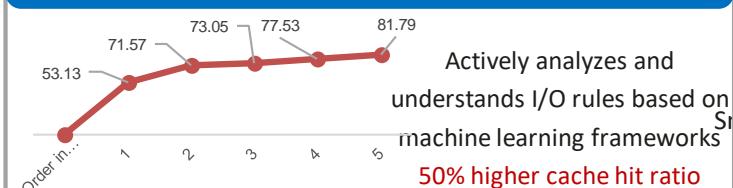
Metadata Data

Global I/O priority ensures stable and low latency

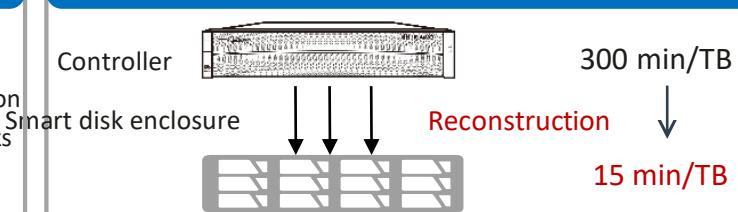


Read/write I/Os are top priority.

AI chip + Intelligent cache algorithm improve cache hit ratio

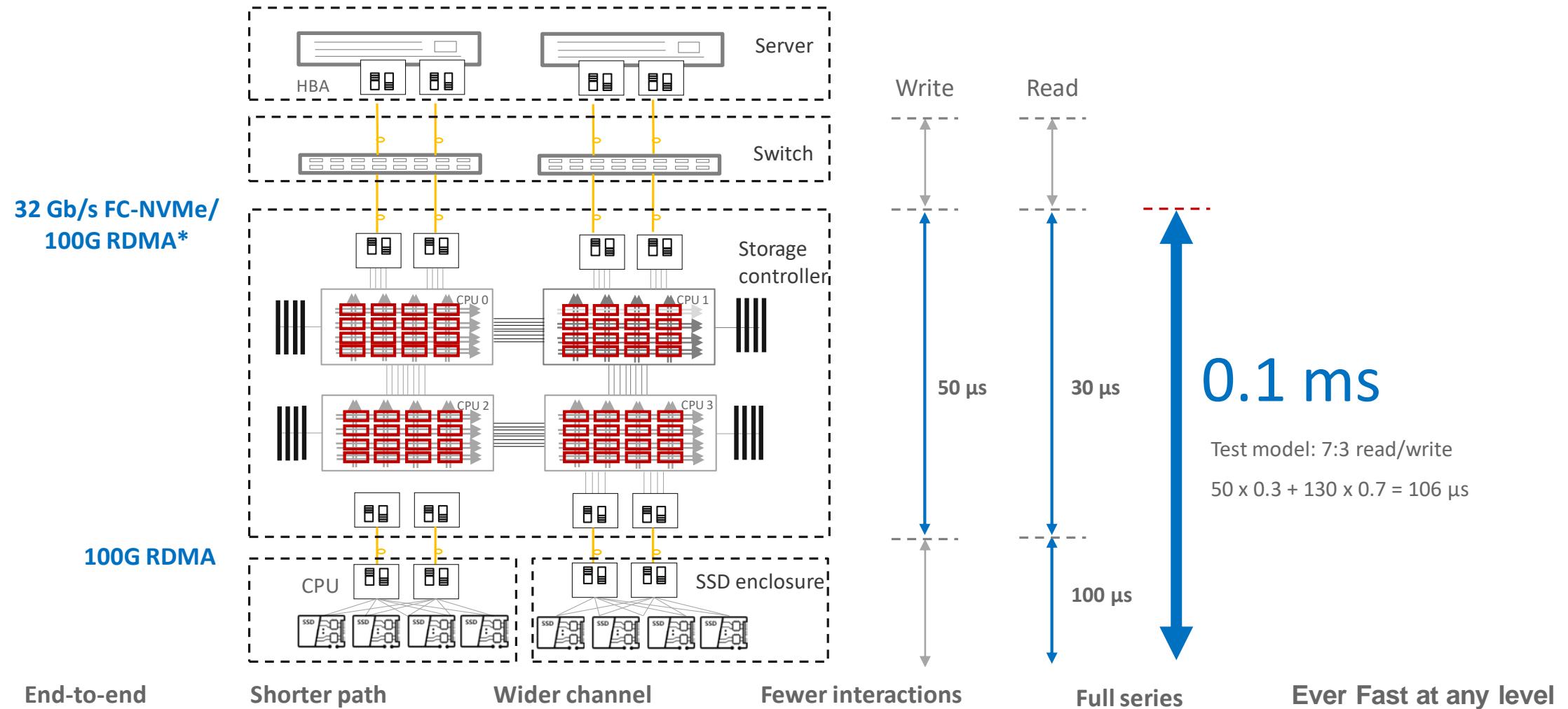


Kunpeng chip + Service splitting enable faster reconstruction

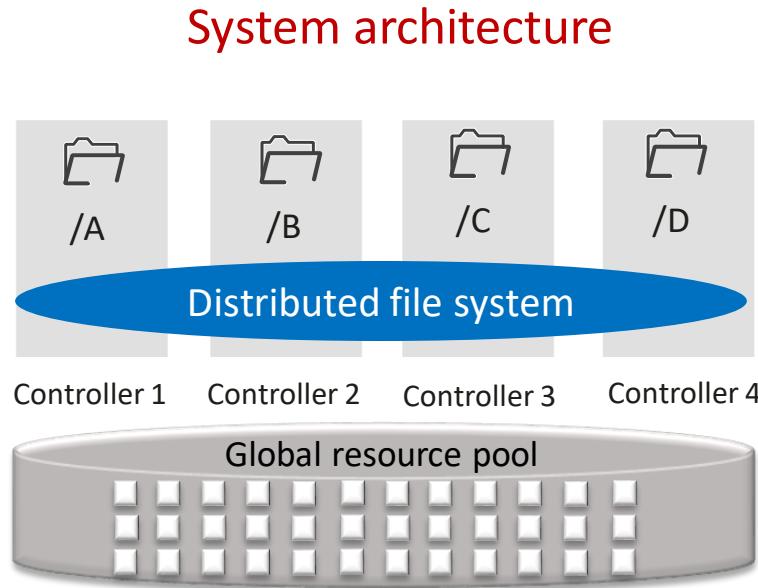


Flash-oriented design & disk and controller collaboration ensure stable application performance on the live network.

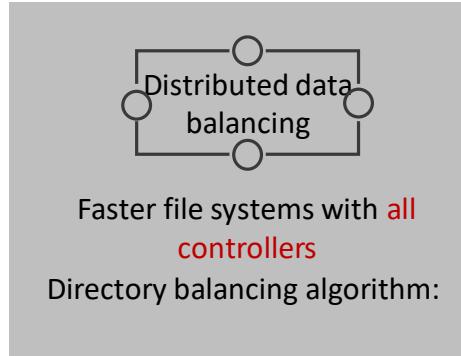
Ever Fast at any level for Pioneering Full-Series with E2E NVMe Architecture



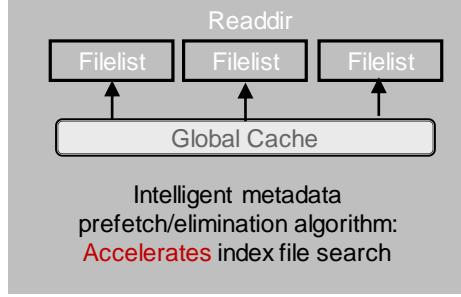
Globally Shared Distributed File System Enables Fast NAS Performance



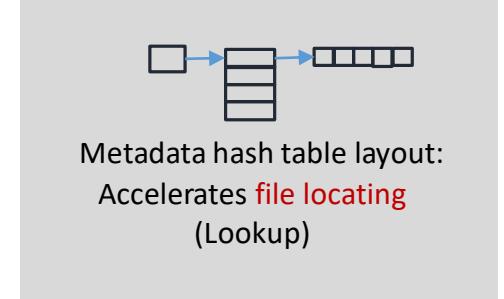
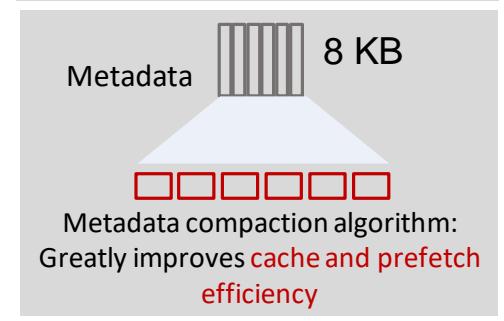
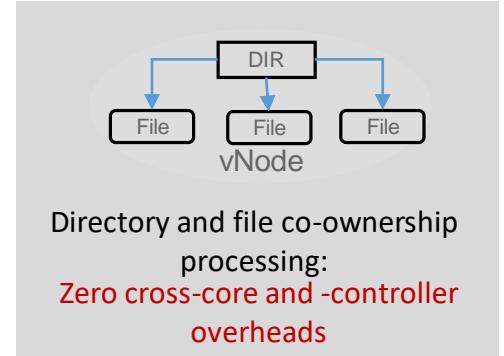
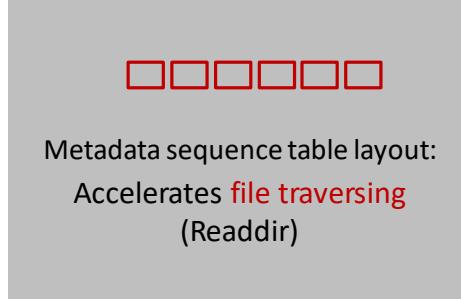
Intelligent balancing



Intelligent cache



Intelligent layout



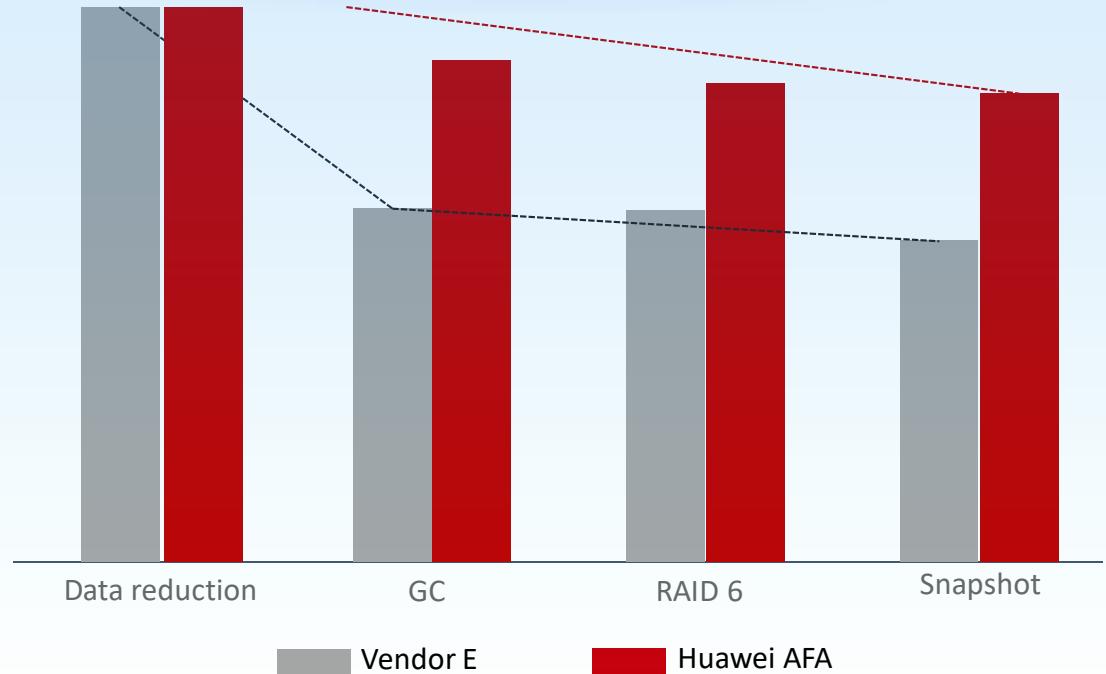
Benefits from All-Flash Native Architecture: <10% Performance impact

Uncompromising Performance

<10% Performance impact
with rich features
in Both SAN and NAS Scenarios

Performance with value-added features enabled

Vendor's AFA: 20% to 30% decrease; Huawei AFA: < 10% decrease

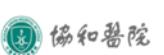


Test devices: dual-controller, 1 TB cache, NVMe backend, 25 x 3.84 TB SSDs

Ever Solid

12,000+ DCs

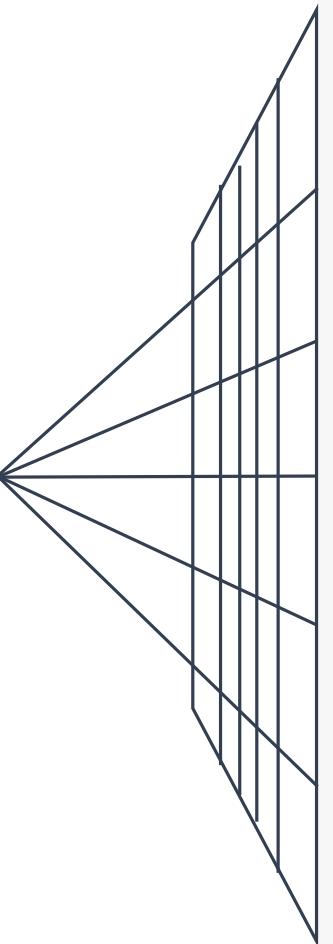
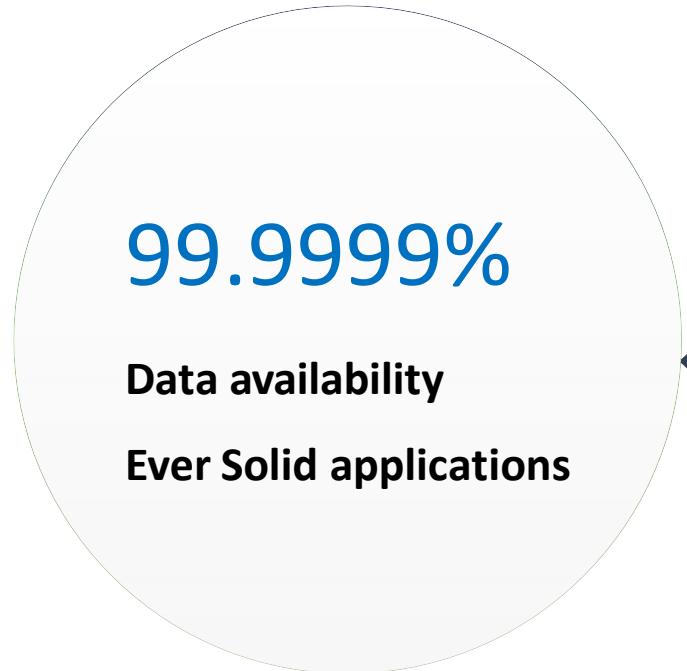
13-year stable operation on the live network



Helsana



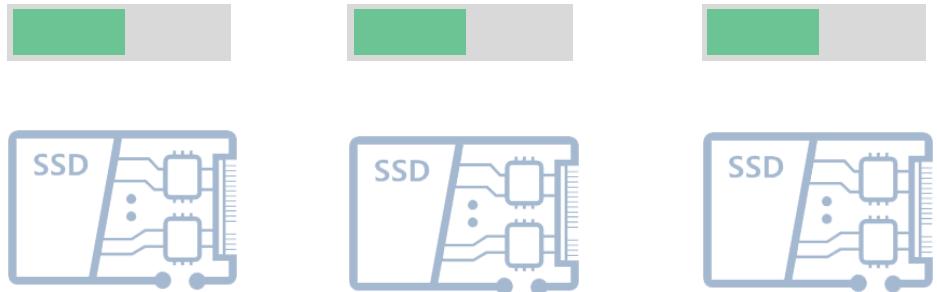
Ever Solid Applications with Five Reliability Layers



-  **Ever Solid Cloud backup**
 - **Gateway-free** cloud backup
 - 30x higher backup frequency and 20x faster backup speeds
-  **Ever Solid Solution**
 - Gateway-free active-active solution (1 ms latency)
 - FlashEver **without data migration**
-  **Ever Solid System**
 - **Comprehensive enterprise-class features**
 - RAID-TP tolerates simultaneous failure of 3 disks
 - **Only 15-minute** reconstruction time of 1 TB of data
-  **Ever Solid Architecture**
 - **SmartMatrix** fully-interconnected architecture tolerates failure of 7 out of 8 controllers
 - E2E A-A design for ever solid applications
-  **Ever Solid SSD**
 - Global wear leveling
 - **Huawei-patented global anti-wear leveling**

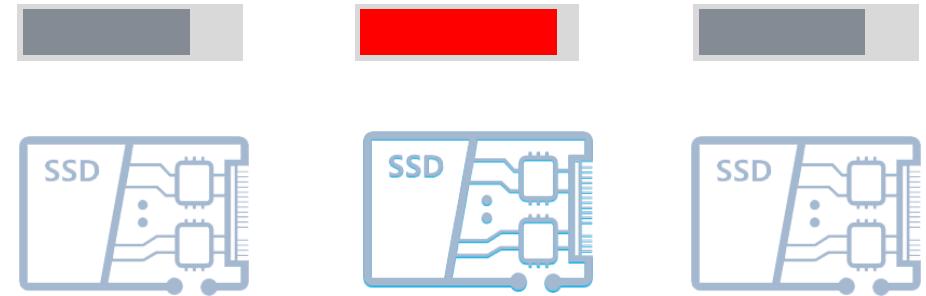
Ever Solid Disk: Global Wear and Anti-Wear Leveling

Early SSD life: global wear leveling



RAID 2.0+ improves SSD reliability by evenly distributing data to SSDs with fingerprints for **wear leveling**.

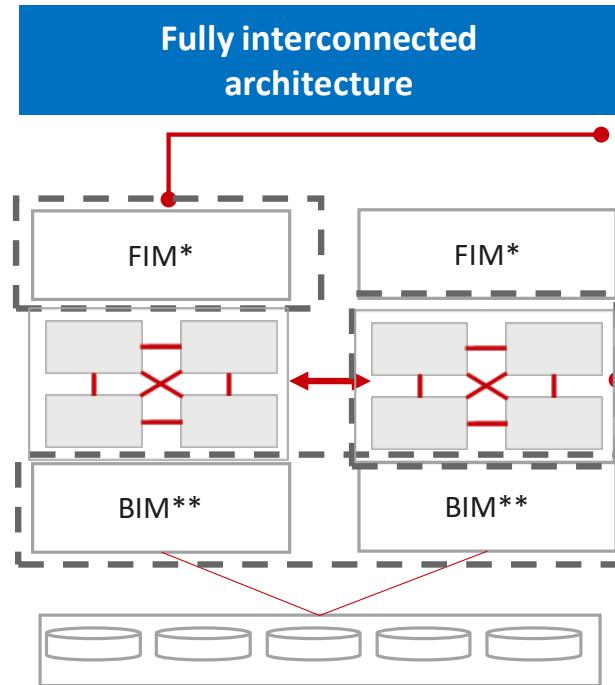
Late SSD life: Huawei-patented global anti-wear leveling



The workload of one SSD increases to **prevent service downtime from simultaneous failures of multiple SSDs**.

Extend SSD service life and improve reliability

Ever Solid Architecture: SmartMatrix Sets A New Benchmark of Reliable



FIM sharing

- A FIM connects to 4 controllers through PCIe ports to access all the controllers in active-active mode using multi-channel technology.

Fully interconnected controllers

- The controllers in an enclosure are fully interconnected through a passive backplane.
- Cross-enclosure expansion: 100 Gbit/s RDMA shared interface modules connect to 8, 12, or 16 controllers.

2 controller enclosures connected to 1 smart SSD enclosure

- A BIM is installed in a controller enclosure. All controllers can simultaneously access an SSD enclosure connected to the BIM.
- A smart SSD enclosure has 2 groups of uplink ports that connect to 2 controller enclosures. The SSD enclosure connects to 8 controllers.

*Front-end interconnect I/O module (FIM)

**Back-end interconnect I/O module (BIM)

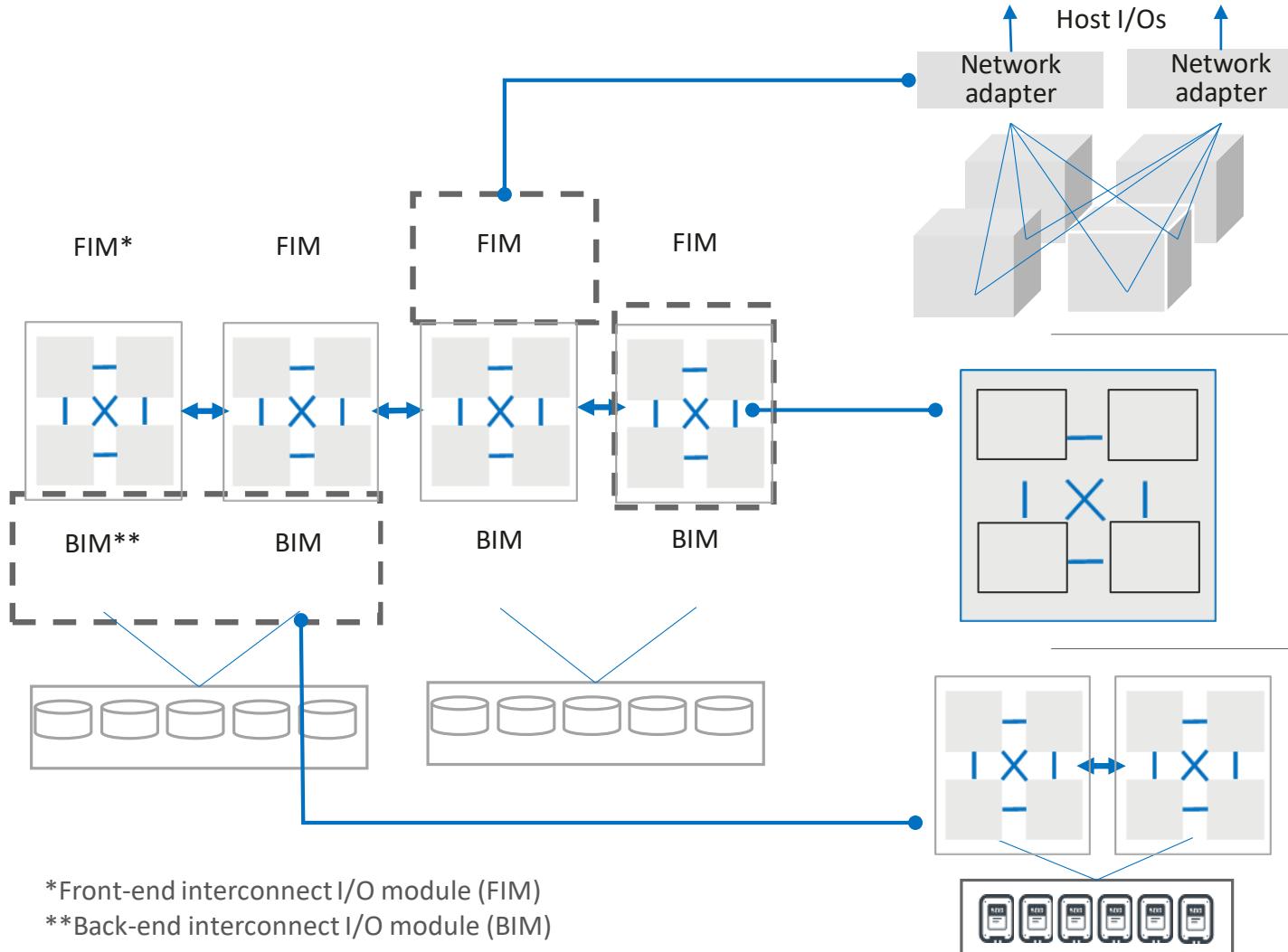
Ever Solid Architecture

Tolerates failure of up to **7 controllers**

Tolerates failure of **1 controller enclosure**

	OceanStor Dorado V6	Vendor E	Vendor H
Tolerates failure of 1 controller	✓	✓	✓
Tolerates failure of 2 controllers	✓	✗	✓
Tolerates failure of 7 controllers	✓	✗	✗

Architecture Reliability: E2E Full Mesh for Service Continuity



One FIM shared by 4 controllers

- A FIM connects to 4 controllers through PCIe ports to access all the controllers in active-active mode using multi-channel technology.

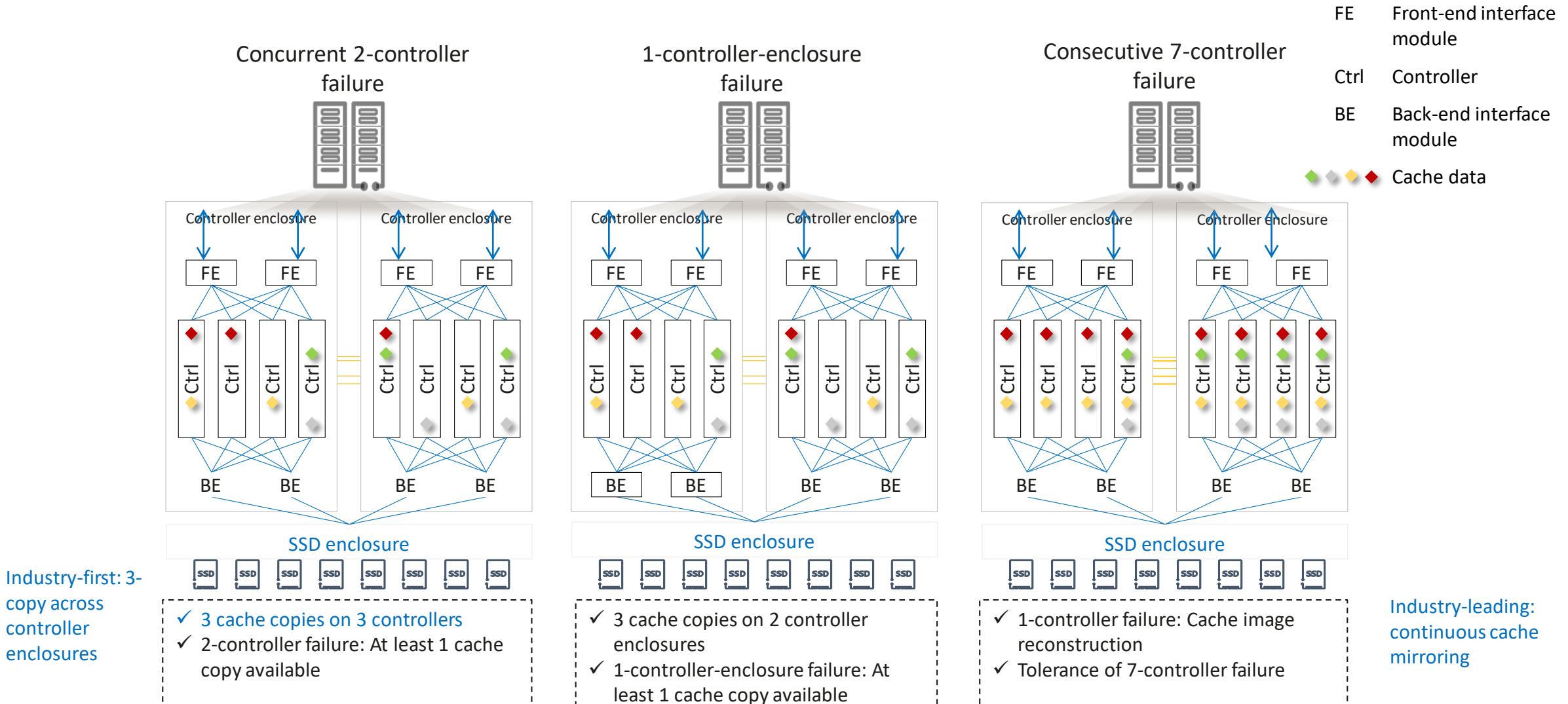
Fully interconnected controllers

- The controllers in an enclosure are fully interconnected through a passive backplane.
- Cross-enclosure expansion: 100 Gbit/s RDMA shared interface modules connect to 8, 12, or 16 controllers.

2 controller enclosures connected to 1 smart SSD enclosure

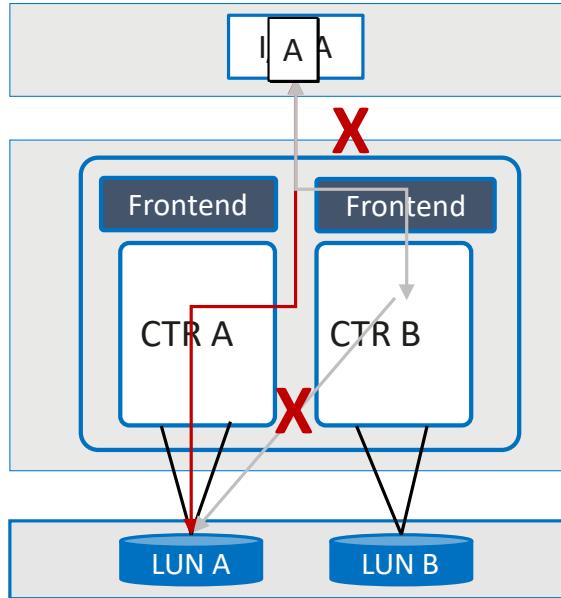
- A BIM is installed in a controller enclosure. All controllers can simultaneously access an SSD enclosure connected to the BIM.
- A smart SSD enclosure has 2 groups of uplink ports that connect to 2 controller enclosures. The SSD enclosure connects to 8 controllers.

Architecture Reliability: Tolerance for Failures of Up to 7 Controllers

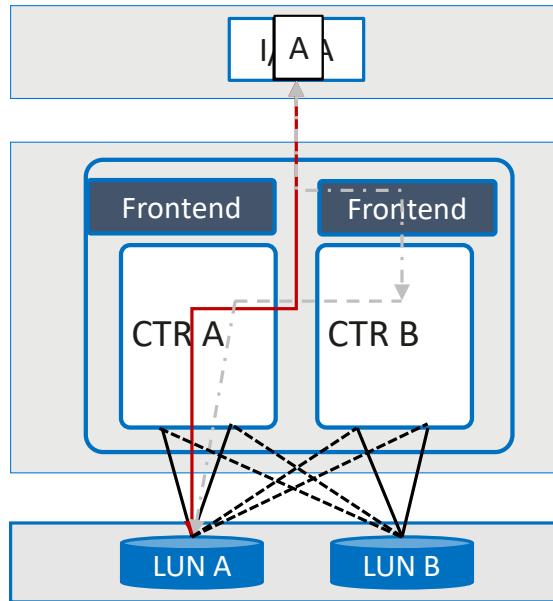


System Architecture Comparison of Mid-Range Storage Controllers

A-P

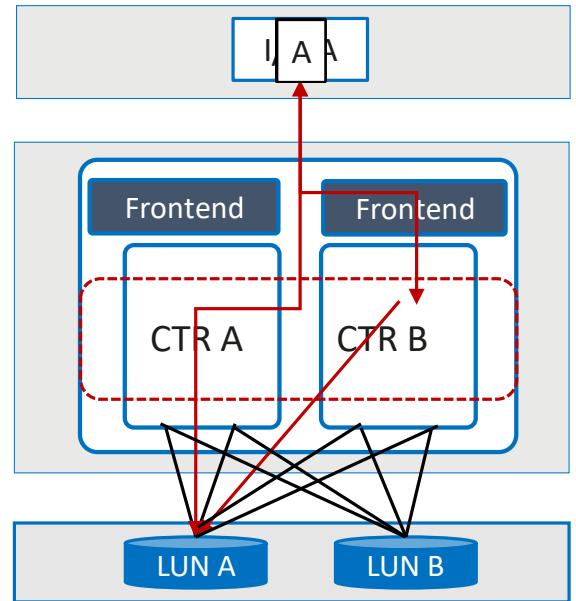


ALUA*



- LUNs have no owning controllers.
- Only the primary controller processes I/Os.
- Single LUN has a processing performance bottleneck.
- Failover takes dozens of seconds, affecting services.

A-A

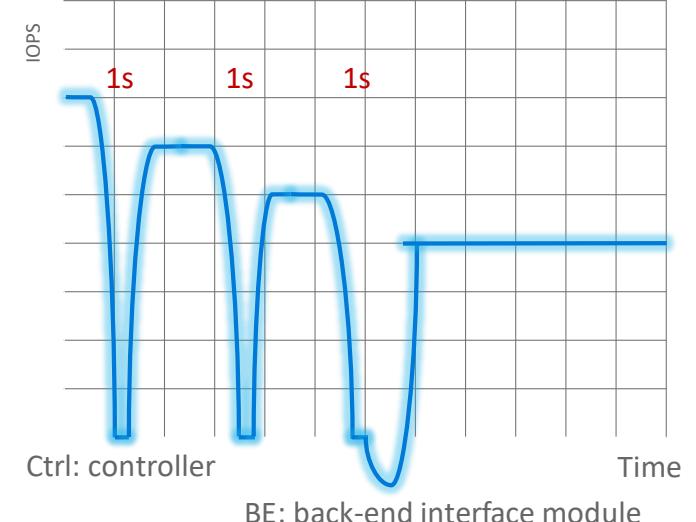
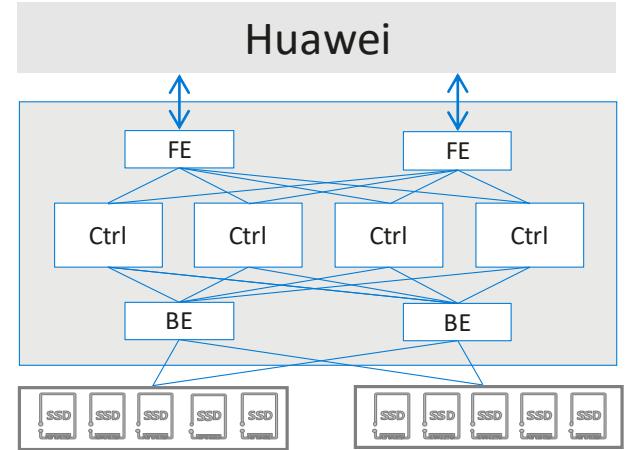
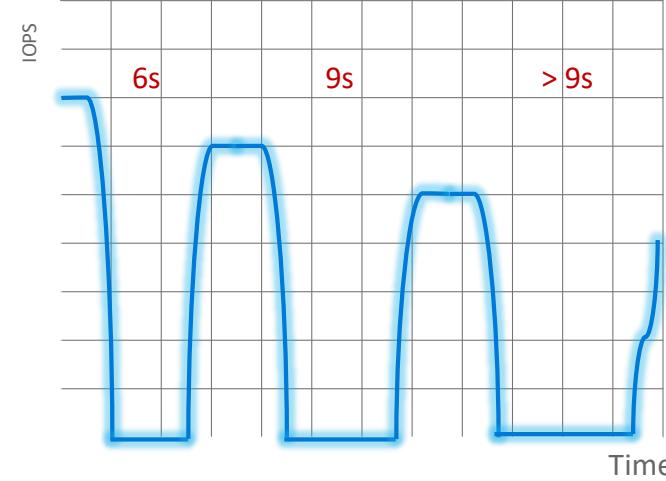
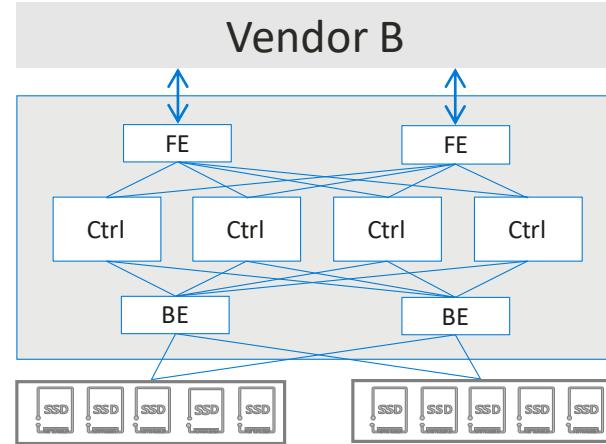
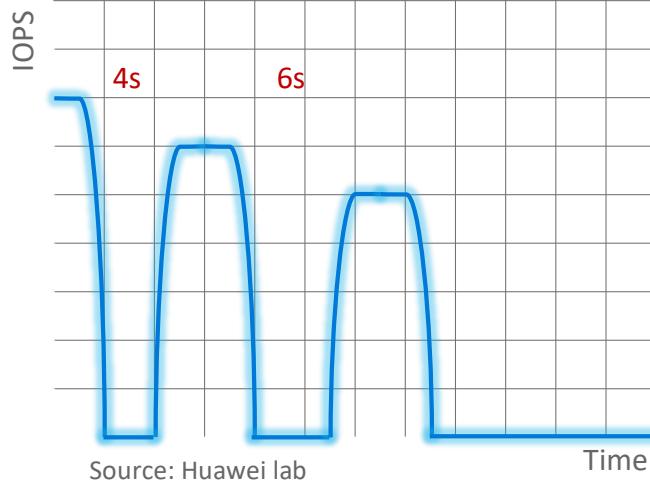
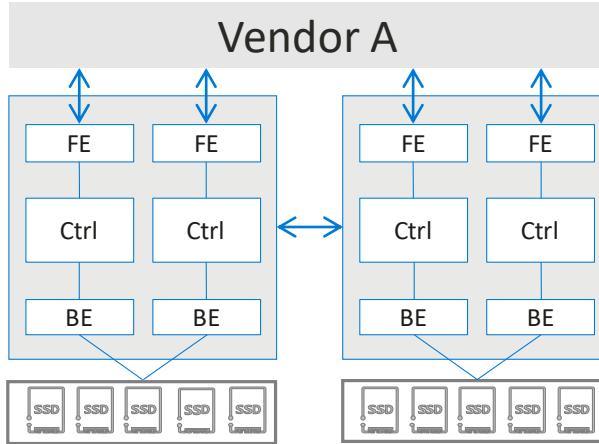


- LUNs have no ownership.
- All front-end controllers receive I/Os and transfer them to the owning controller for processing.
- Single LUN has a processing performance bottleneck.
- Failover takes dozens of seconds, affecting services.

- LUNs have no ownership.
- All controllers participate in I/O processing.
- Single LUN has no performance bottleneck.
- **Failover in seconds** upon a controller failure, with no impact on services.

System Architecture Comparison of High-end Storage Controllers

– Seamless Service Switchover in Seconds



Remarks: It is not supported to tolerate three controllers fail at the same time.

FE: front-end interface module

BE: back-end interface module

Ever Solid System Reliability: RAID-TP Provides Protection for SSDs

Large-capacity SSDs lead to double the capacity (up to 32 TB) and 5x to 10x the failure rate

Simultaneous 3-disk failure without service interruption



SSD failure toleration

Traditional RAID: up to 2 SSDs

Huawei RAID-TP: simultaneous **3-SSD** failure

15 minutes

RAID-TP

5 hours

Traditional RAID

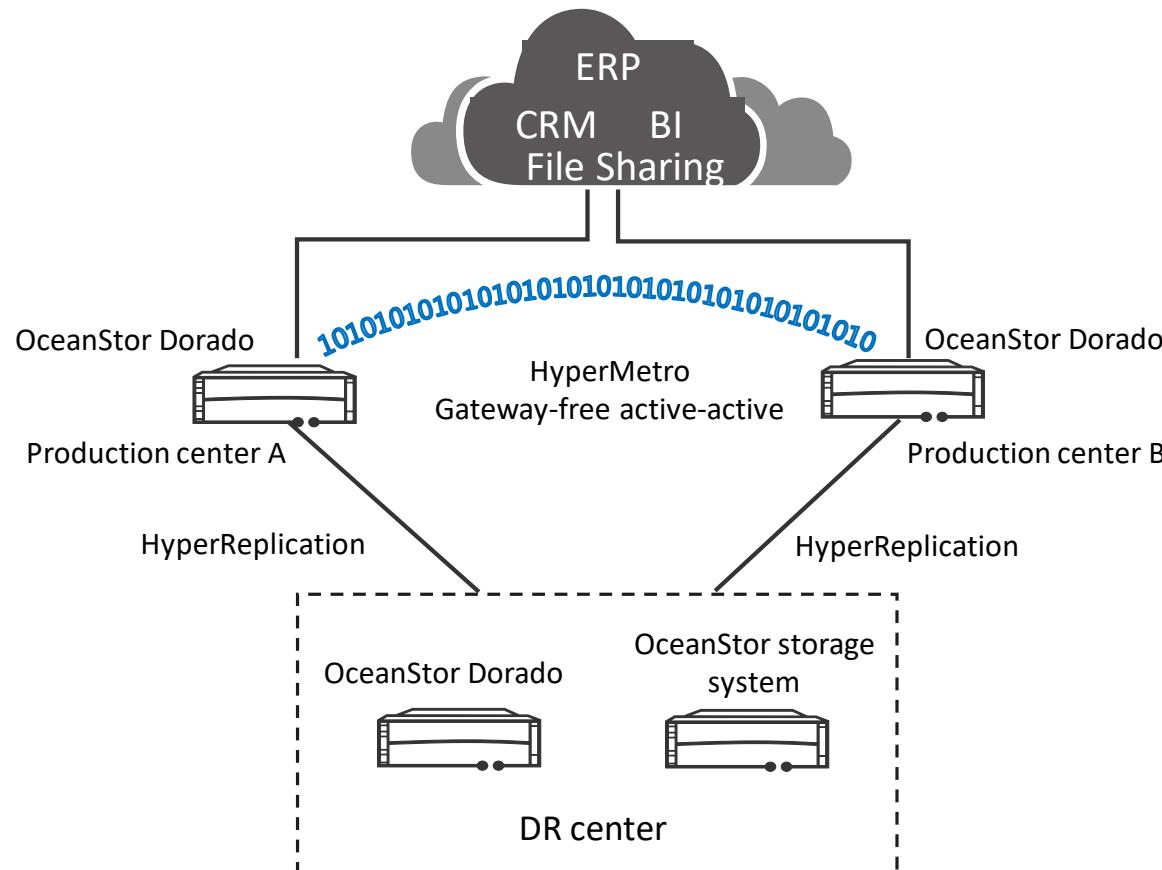
Reconstruction of 1 TB data

Data reconstruction

Traditional RAID: 5 hours

RAID-TP: 1 TB of data within **15 minutes**

Ever Solid Solution Reliability: Active-Active Solution for SAN and NAS

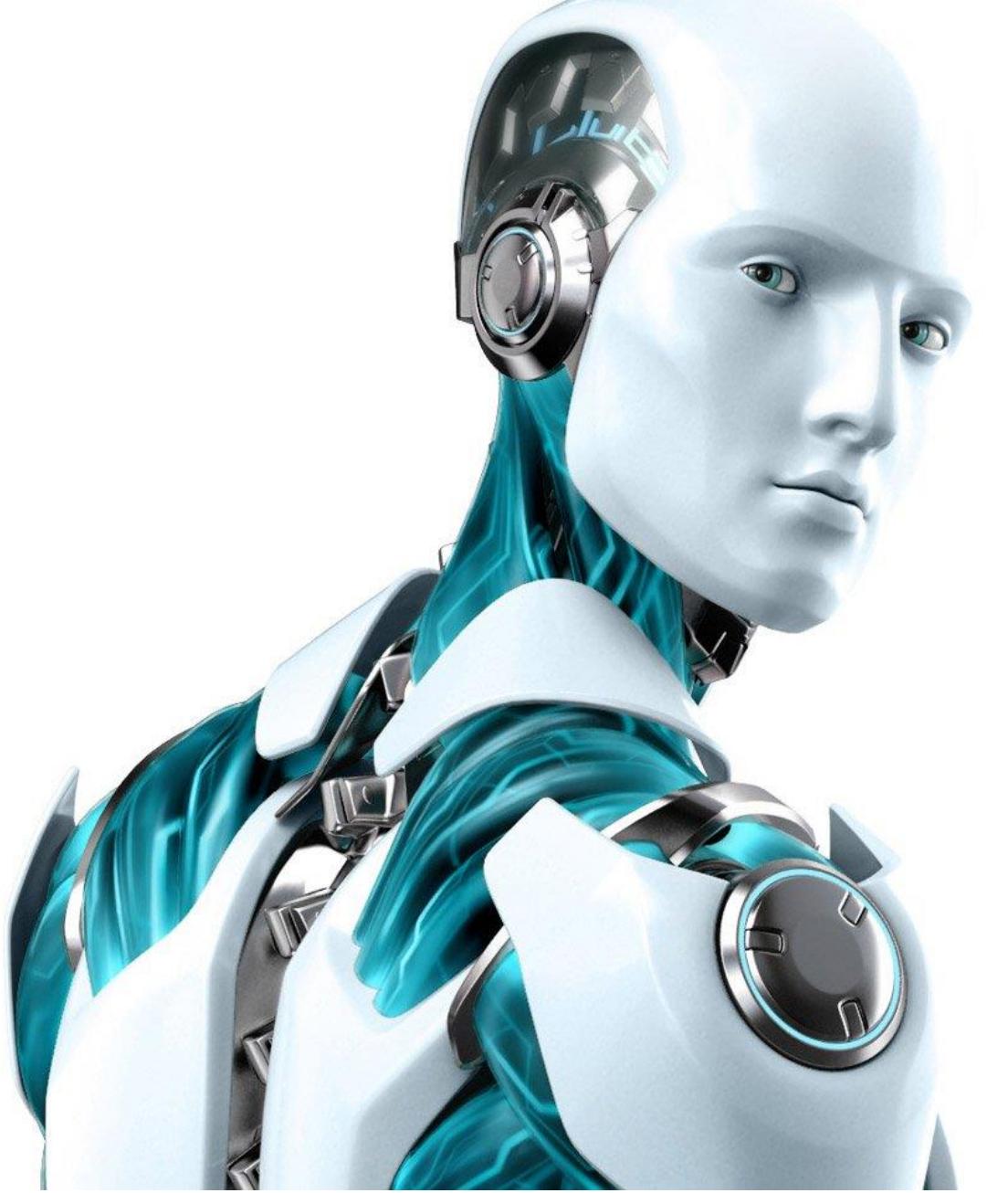


Lightning-fast and rock-solid

- **Gateway-free:** SAN and NAS convergence, fewer nodes, and simplified management
 - **SAN active-active:** Load balancing between sites, RPO = 0, RTO ≈ 0, 50% higher all-IP A-A performance compared to traditional IP solution
 - **Industry's only NAS A-A solution**

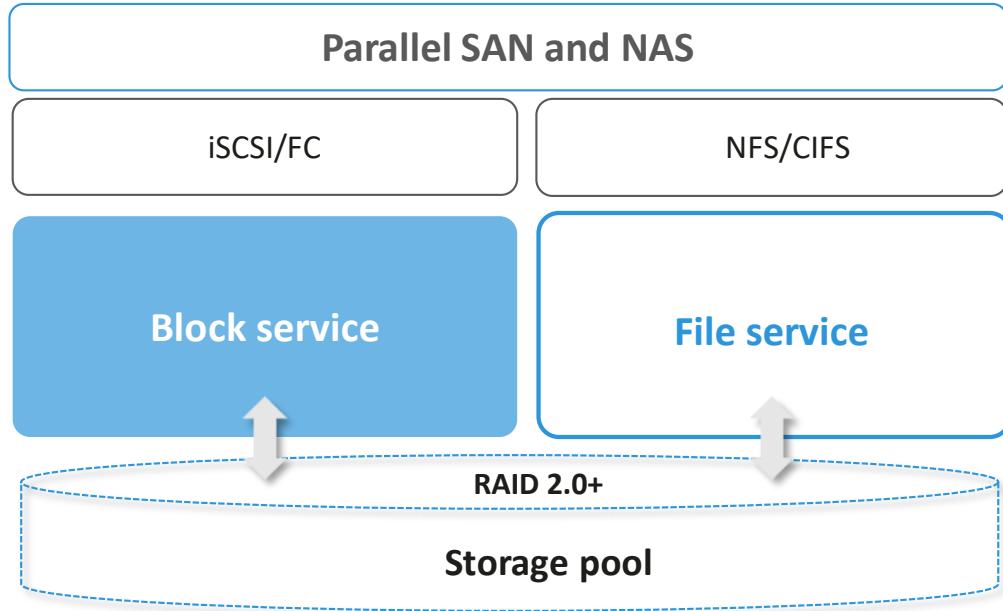
Easy-to-scale

- Scalability to 3DC (SAN) improves reliability.
 - Serial, parallel, and star networking (SAN) meets the most demanding requirements for enterprise reliability.



Intelligent

Converged SAN and NAS: One System for Multiple Workloads



Gateway-free

All-in-one block and file storage and NAS gateway-free design for **20%** lower procurement costs.

Parallel architecture

Parallel SAN and NAS provide different workloads with optimal access paths for the highest access performance.

Healthcare
HIS/PACS



Finance
Cheque imaging

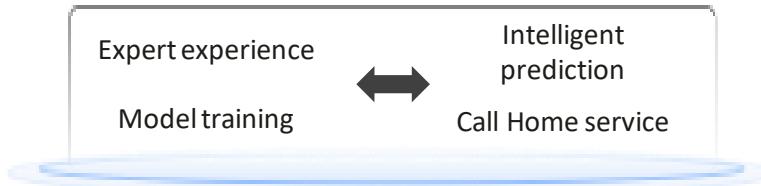


Government
E-Gov./OA systems

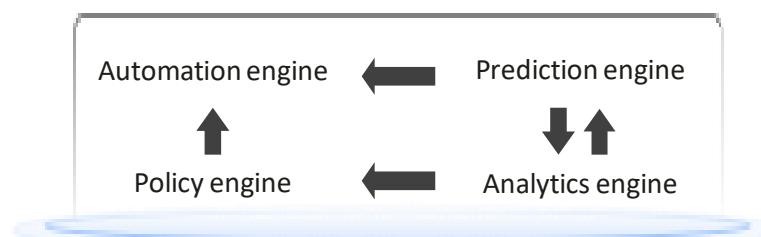
Three Intelligent Layers: Full-Lifecycle Data Management for the Intelligent Era



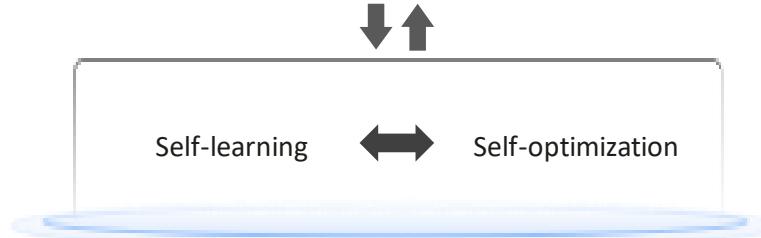
Cloud Intelligent
Intelligent model training



Center Intelligent
Intelligent model application



Device Intelligent
Intelligent hardware and algorithms



DME IQ

Cloud Brain for intelligent O&M



DME Storage

Execution engine and automatic management

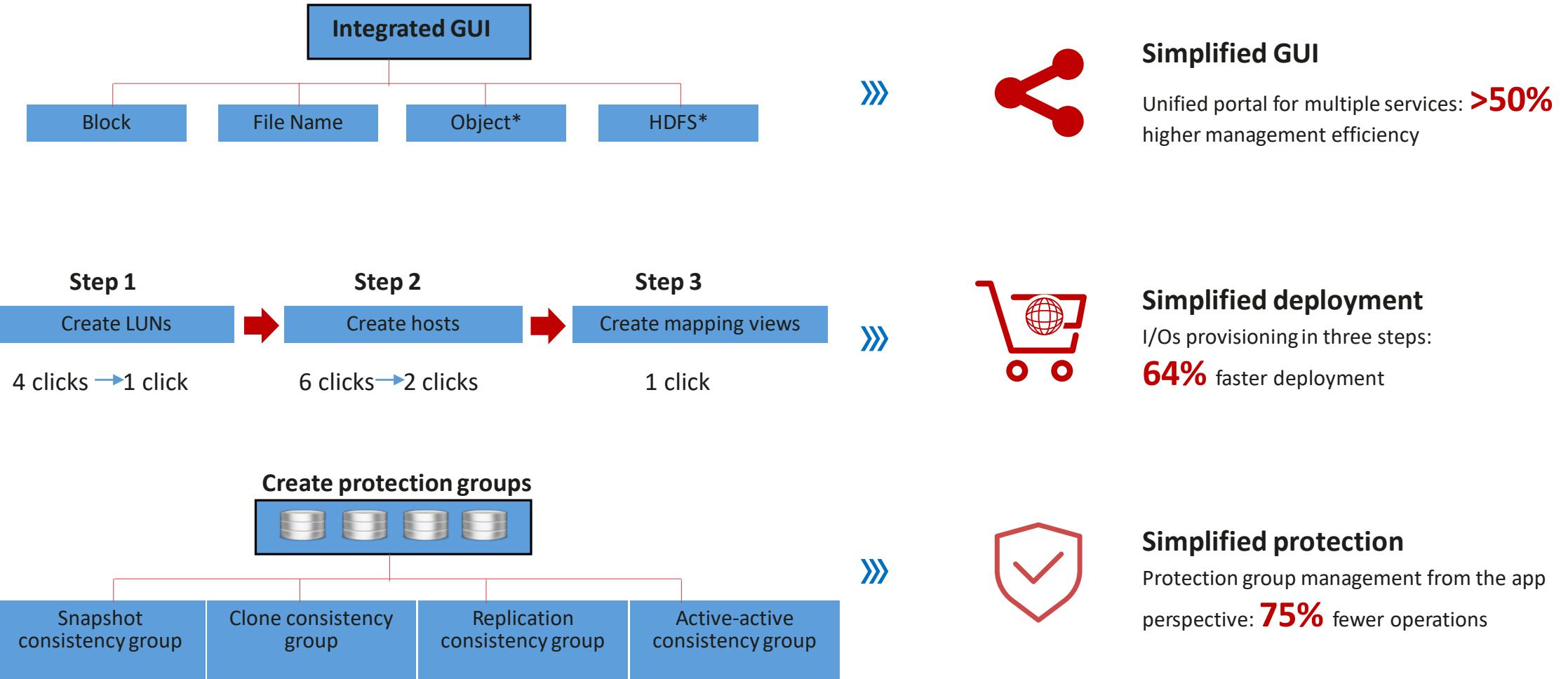


DeviceManager

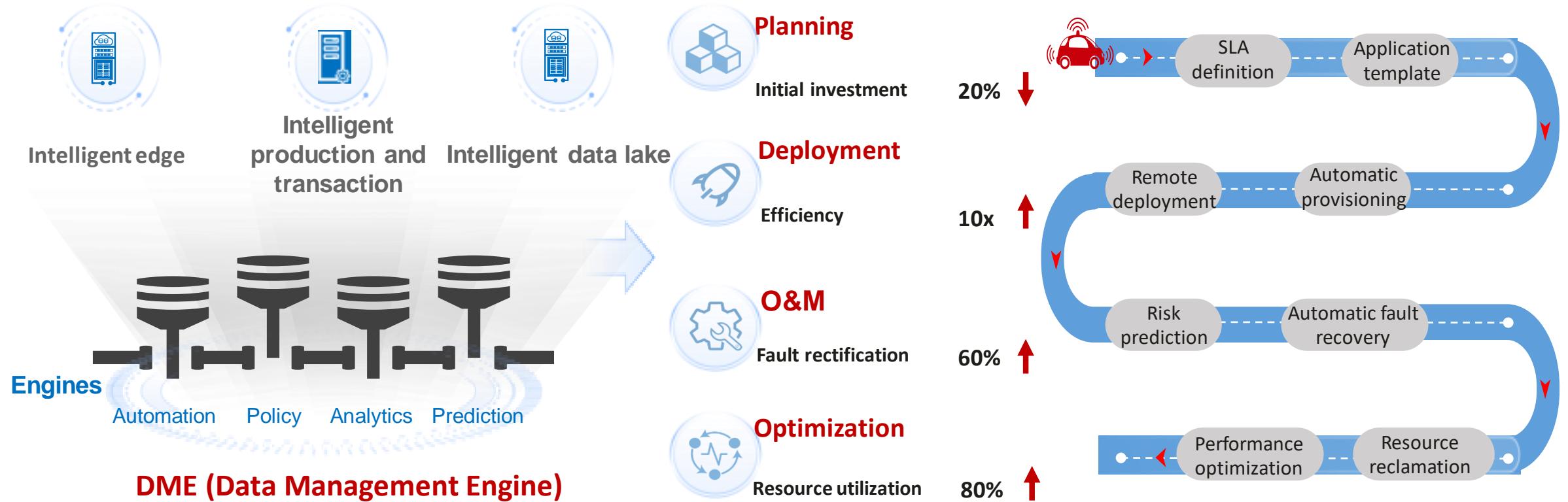
Intelligent devices and simplified configuration

* DME: Data Management Engine

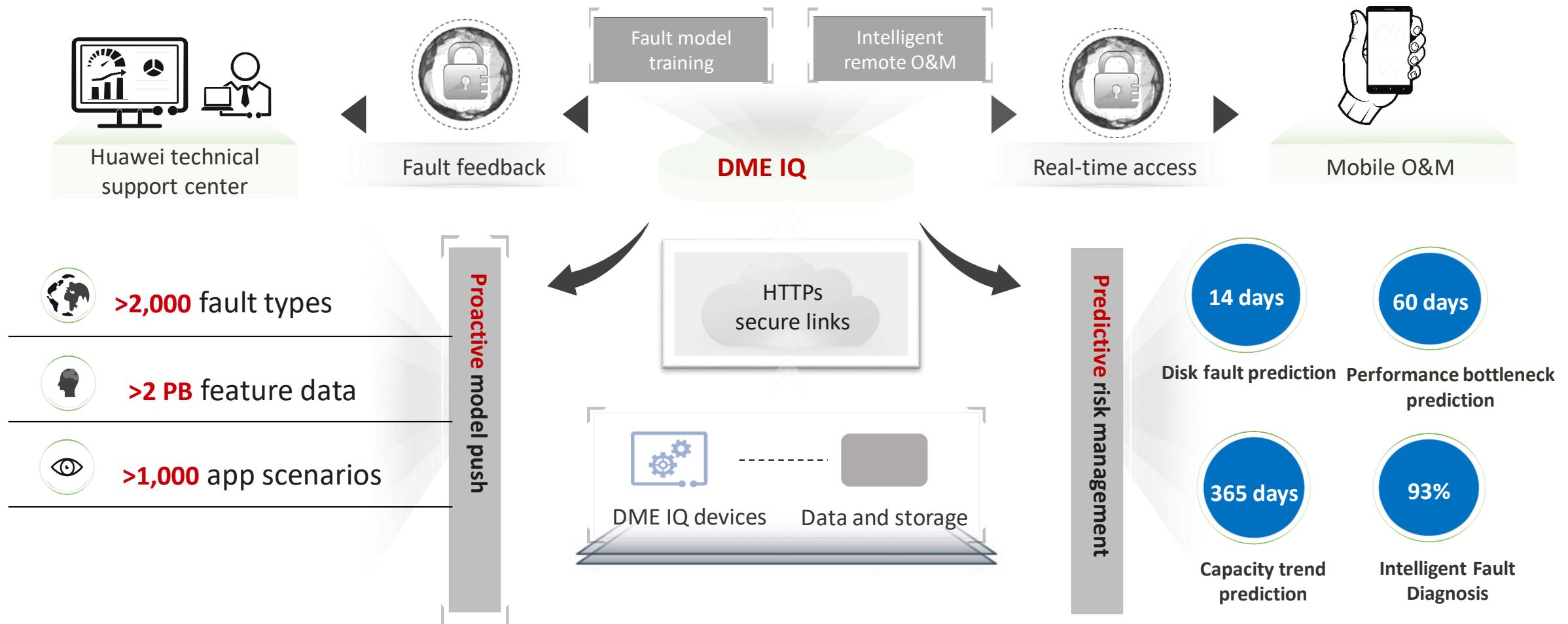
DeviceManager: Easy-to-Use, Intelligent Device Management



DME Storage: Intelligent Management Engine for Full-Lifecycle Automation



DME IQ: Cloud Brain for Proactive, Predictive, and Intelligent O&M



Quiz

1. (Multiple-choice) Which of the following key hardware designs are used for Industry-leading performance in OceanStor Dorado ?
 - A. FlashLink® Intelligent Algorithms
 - B. E2E NVMe Architecture
 - C. Globally Shared Distributed File System
 - D. All-Flash Native Architecture

2. (Multiple-choice) Which of the following key designs are used for Solid Stability in OceanStor Dorado ?
 - A. Global Wear and Anti-Wear Leveling to enable Ever Solid Disk
 - B. SmartMatrix design to enable Ever Solid Architecture
 - C. RAID-TP feature to support Ever Solid System with fast data reconstruction
 - D. HyperMetro Active-Active Solution for SAN and NAS to enable Ever Solid Solution

Contents

1. Overview
- 2. Hardware Architecture**
3. Software Architecture
4. Smart Series Features
5. Hyper Series Features
6. Other Key Features

Overview and Objectives

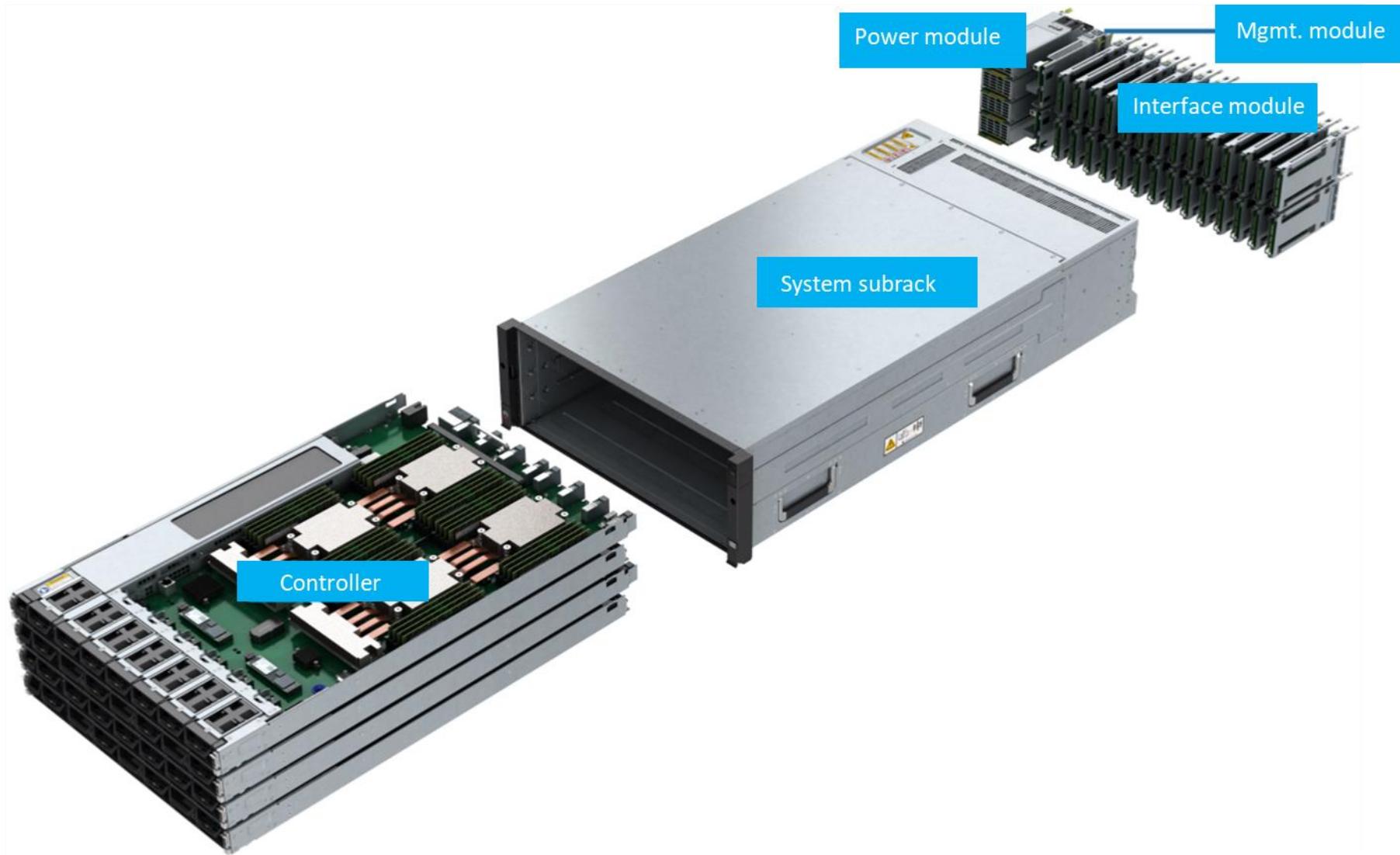
- This section describes the hardware architecture of Huawei OceanStor Dorado series.
- On completion of this section, you will be able to:
 - Describe the hardware structure, components, and key hardware design of the OceanStor Dorado High-end;
 - Describe the hardware structure, components, and key hardware design of the OceanStor Dorado Mid-range and Entry-level.

New Generation Innovative Hardware Platform

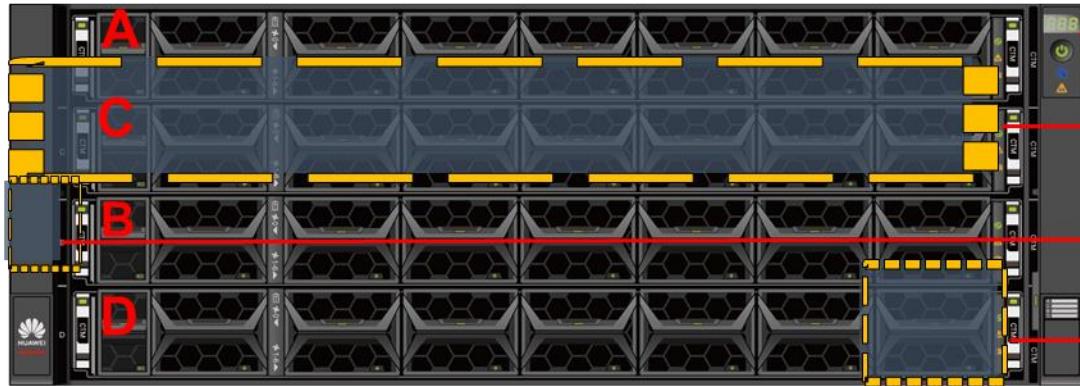
	Rear Panel	Front Panel
High-end controller enclosure		
Mid-range controller enclosure		
Entry-level controller enclosure		
Smart disk enclosure		

Device 3D display: https://support-it.huawei.com/3d-center/#/home?series=dorado_v6

OceanStor Dorado 8000/18000 V6 Controller Architecture



OceanStor Dorado 8000/18000 V6 Form Factor



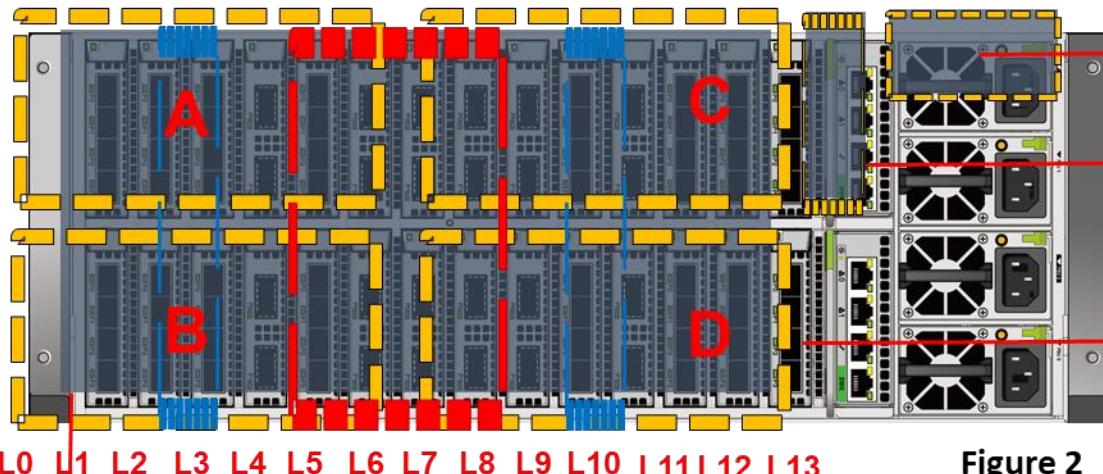
H0 H1 H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 H12 H13 Figure 1

4U independent controller enclosure (four controllers)

Controller modules: A-C-B-D (top-to-bottom), A&B and C&D are mirrored and configured in pairs.

BBU: One BBU per controller, which powers the corresponding controller independently.

Fan module: 6+1 redundancy for each controller



L0 L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12 L13 Figure 2

- Power supply: 2+2 redundancy, supporting 100–240V AC and 240V HVDC

Management board: 1+1 redundancy, hot swappable. Each module provides 1 serial port + 1 management port + 2 maintenance ports.

Interface module

- Each controller enclosure supports up to 28 interface modules which are shared by all controllers. The upper and lower slots are marked as Hs and Ls, respectively.
- Port type: 8/16/32 Gb FC/FC-NVMe, 10/25/40/100 Gb ETH, 12 Gb SAS
- Four scale-out interface modules**, it can be installed from H3,&L3 and H10&L10.
- H5 to H8 and L5 to L8 are back-end interface slots and support up to 800 disks per enclosure.
- Supports SAS 3.0 back-end interface modules, and 100G RDMA back-end interface modules using PCIe.

- If only controllers A and B are configured, install interface modules in the upper and lower symmetrically in areas A and B. The modules are shared only by A and B.

OceanStor Dorado 5000/6000 V6 Controller Architecture

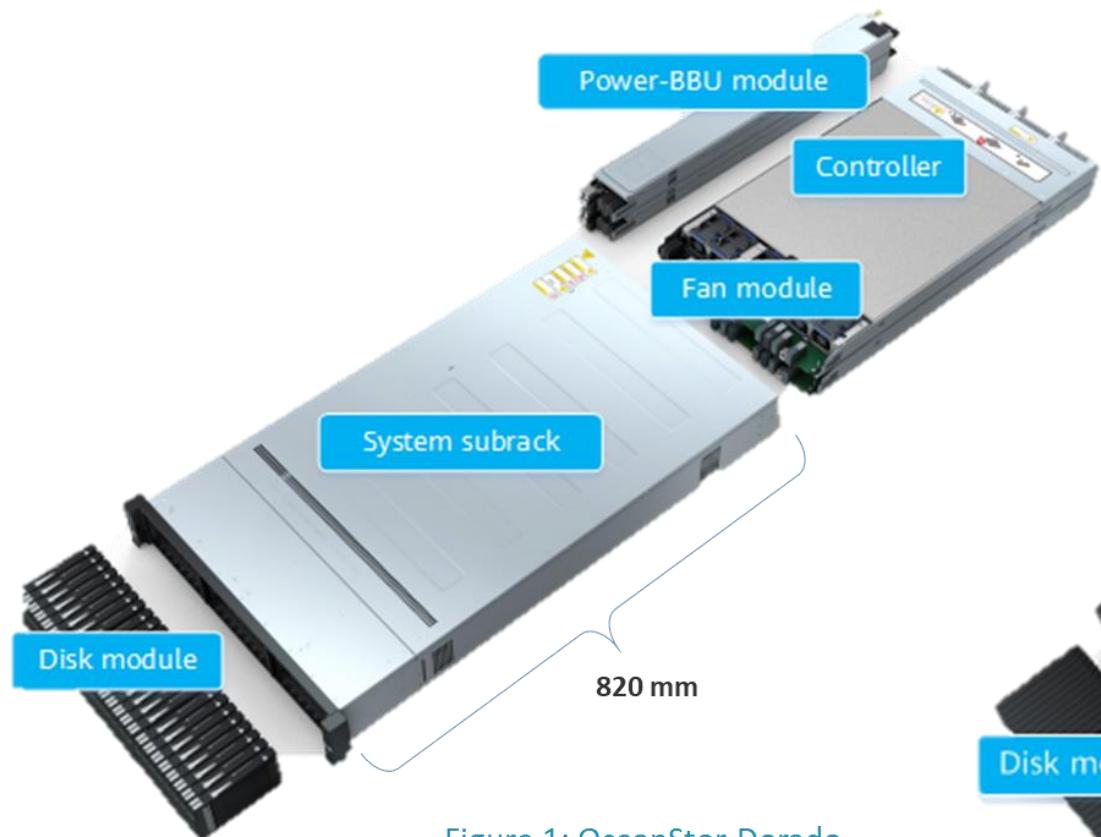


Figure 1: OceanStor Dorado
5000/6000 (SAS)

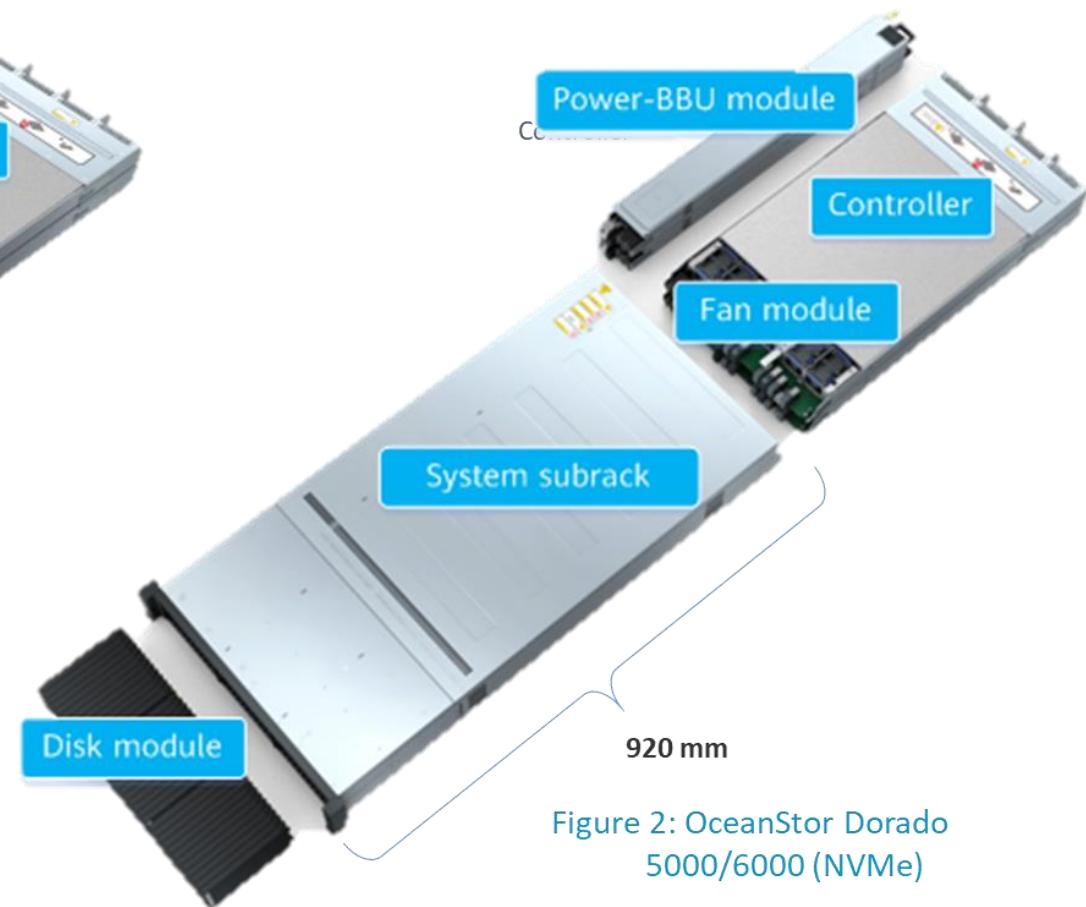
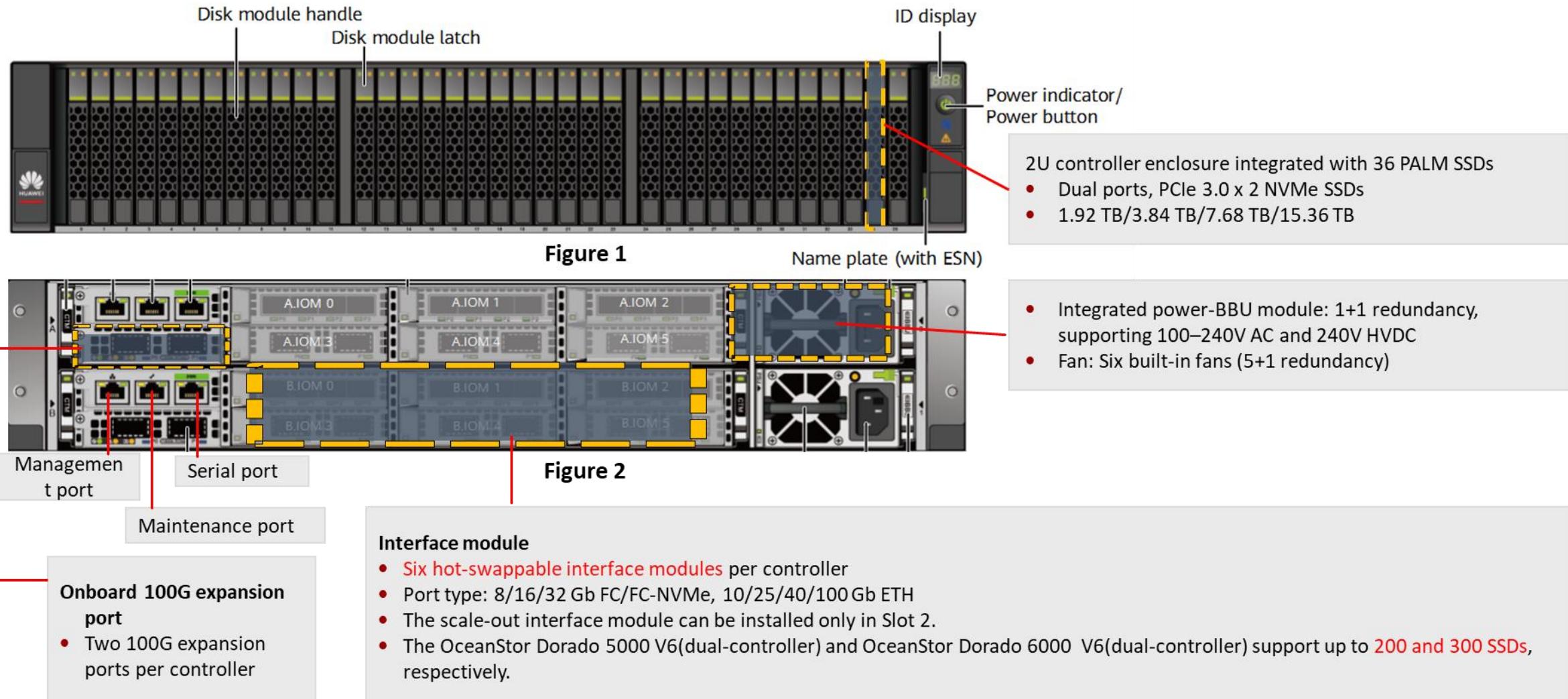
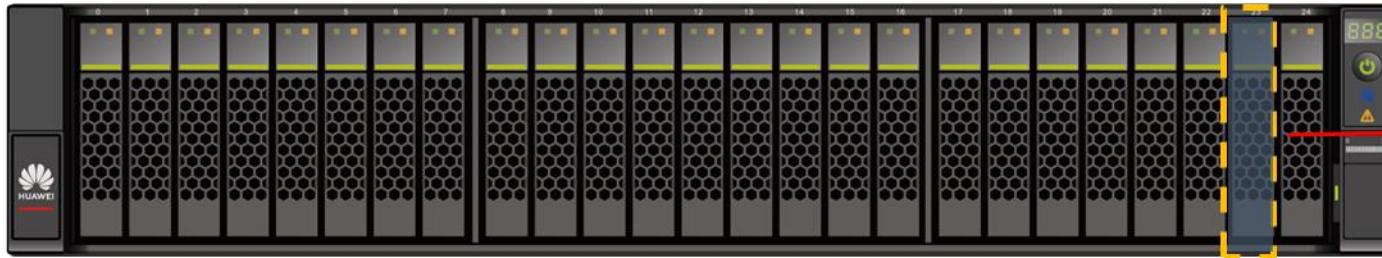


Figure 2: OceanStor Dorado
5000/6000 (NVMe)

OceanStor Dorado 5000/6000 V6 Form Factor (NVMe)



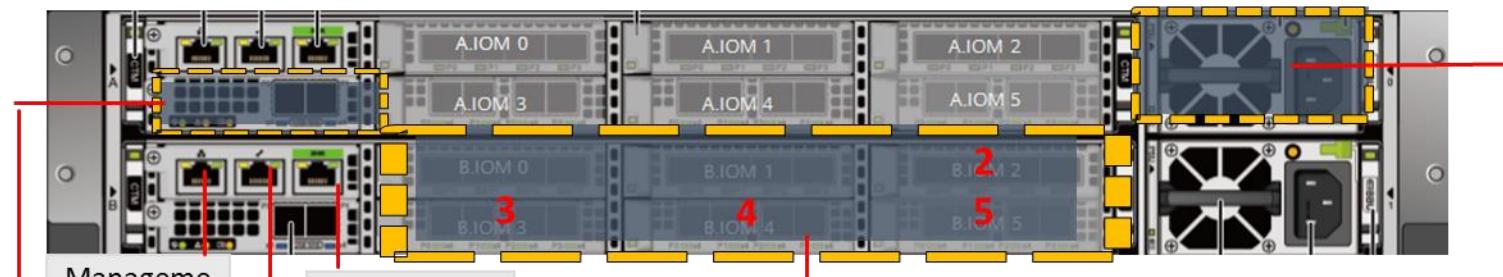
OceanStor Dorado 5000 V6/6000 V6 Form Factor (SAS)



2U controller enclosure with 25 x 2.5-inch integrated SSDs

- 12 Gb SAS SSDs
- 960 GB/1.92 TB/3.84 TB/7.68 TB/15.36 TB/30.72 TB

Figure 1



- Integrated power-BBU module: 1+1 redundancy, supporting 100–240V AC and 240V HVDC
- Fan: Six built-in fans (5+1 redundancy)

Figure 2

Interface module

- Six hot-swappable interface modules per controller
- Port type: 8/16/32 Gb FC/FC-NVMe, 10/25/40/100 Gb ETH, 12 Gb SAS
- The scale-out interface module can be installed only in Slot 2. The scale-up interface modules (at most 3) can be installed only in Slot 3, 4, and 5.
- The OceanStor Dorado 5000 V6(dual-controller) and OceanStor Dorado 6000 V6(dual-controller) support up to 200 and 300 SSDs, respectively.

Onboard SAS expansion port

- Two SAS expansion ports per controller
- *(or 2 100 Gbt/s RDMA ports)

OceanStor Dorado 3000 V6 Controller Architecture

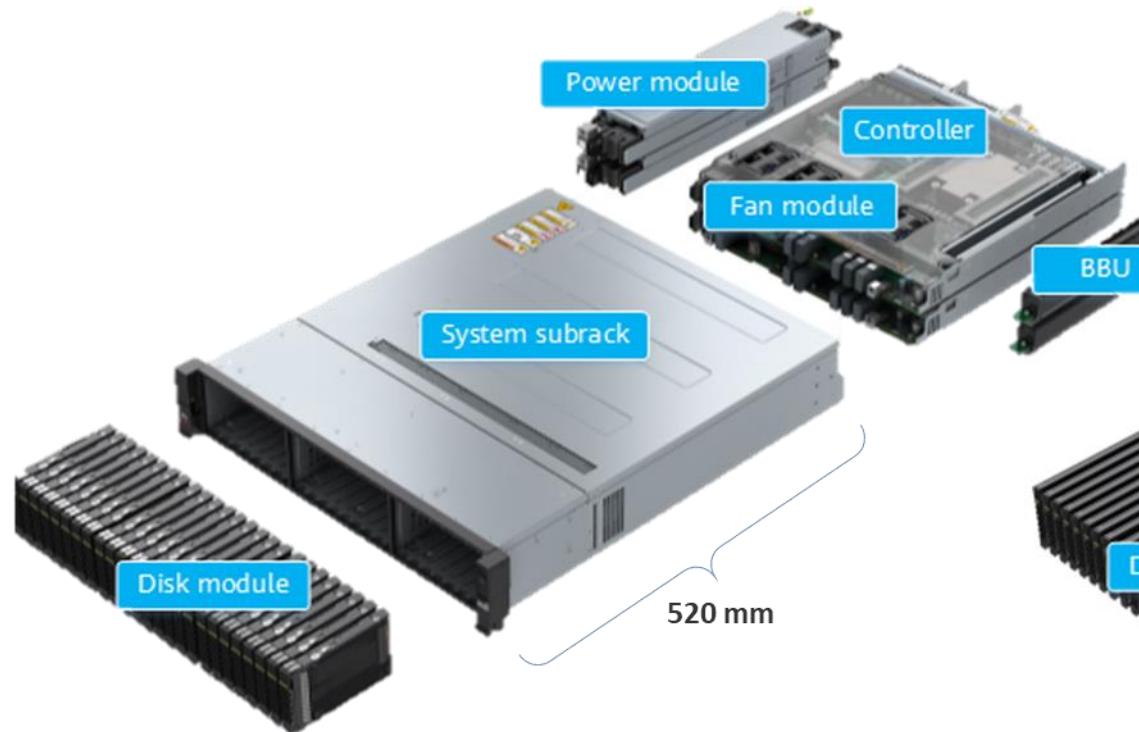


Figure 1: OceanStor Dorado 3000 (SAS)

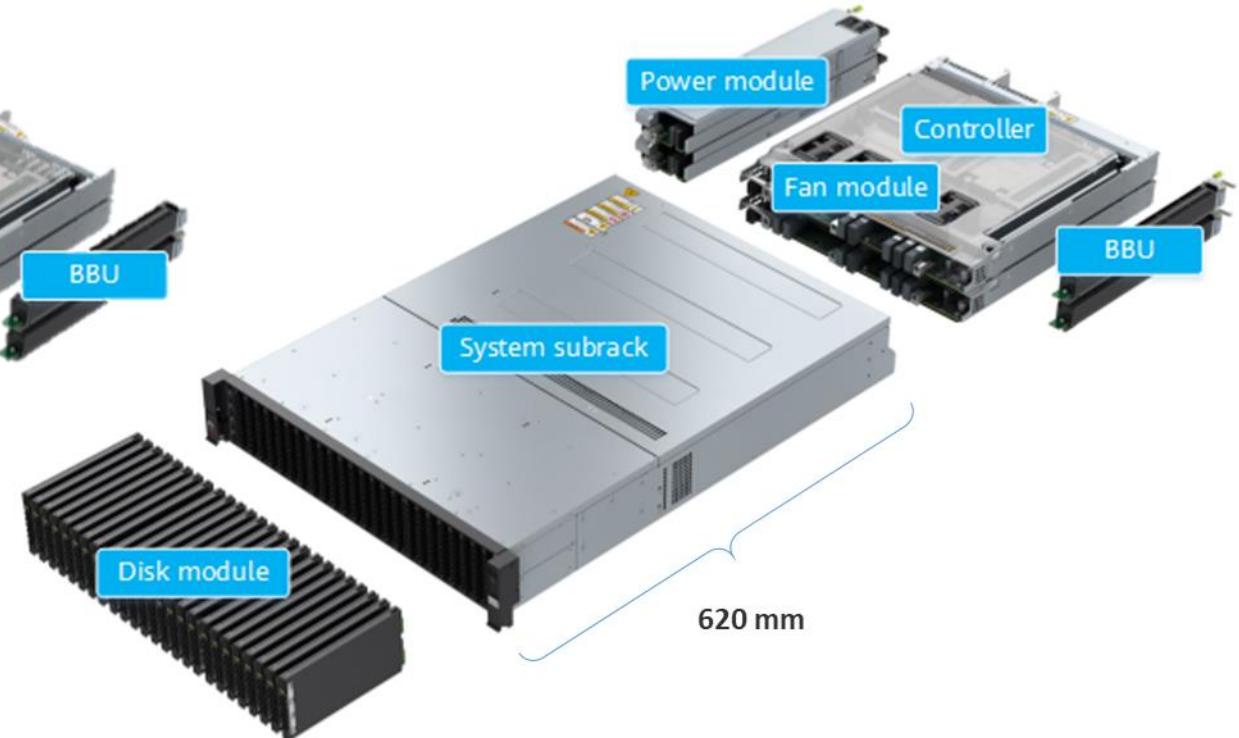


Figure 2: OceanStor Dorado 3000 (NVMe)

OceanStor Dorado 3000 V6 (NVMe) Form Factor

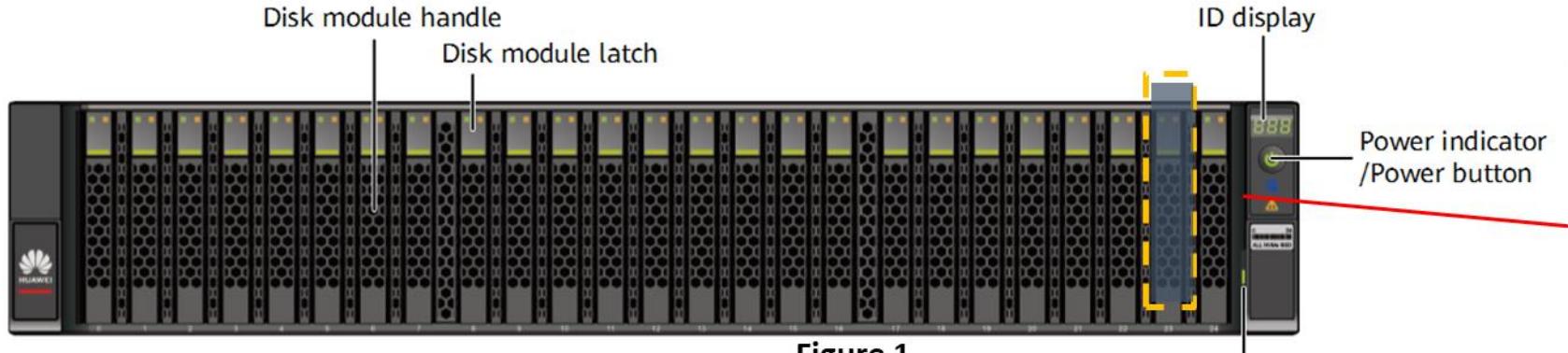


Figure 1

2 U controller enclosure with 25 x palm-sized SSDs

- Dual-port PCIe 3.0 x 2 NVMe SSDs
- 1.92 TB/3.84 TB /7.68 TB/15.36 TB
- SCM drives are supported (conditional sales).
- 800 GB/1600 GB

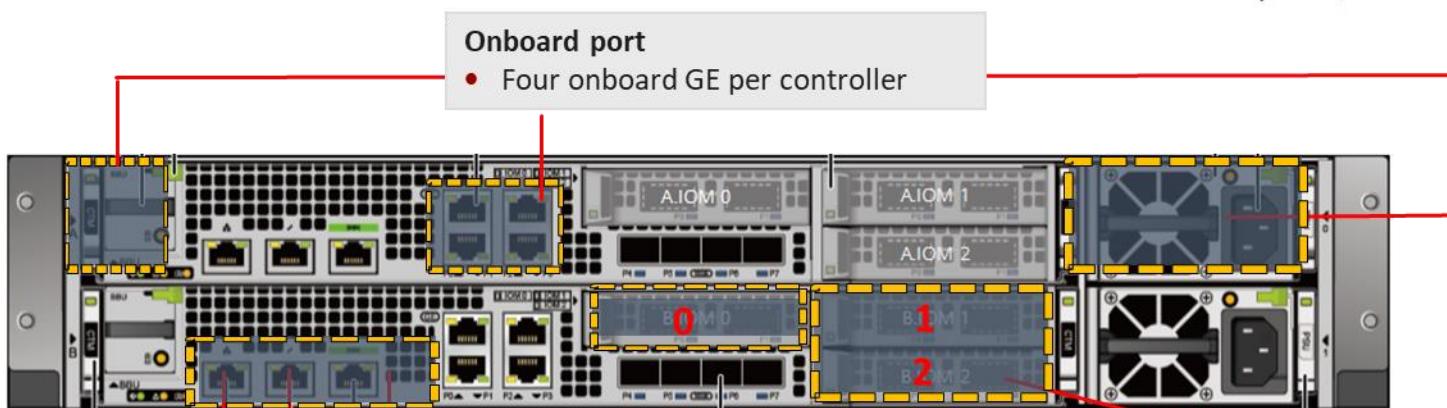


Figure 2

Onboard port

- Four onboard GE per controller

- One BBU per controller, supporting hot swap

Integrated power-fan module

- Power supply: 1+1 redundancy, supporting 100–240 V AC and 240 V HVDC
- Fan: Four built-in fans (3+1 redundancy)

Interface module

- Three hot-swappable interface modules per controller
- Port type: 8/16/32 Gb FC/FC-NVMe, 10/25/40/100GE, **10GE electrical ports, 25 Gb RoCE**, and 100 Gb RDMA
- The scale-out and scale-up interface modules can be installed only in Slots 1 and 2 respectively.
- Dual controllers support up to **150 NVMe SSDs**.

*No backend LOM interfaces are provided. Back-end interface modules must be configured for expansion.

OceanStor Dorado 3000 V6 Product Form Factor

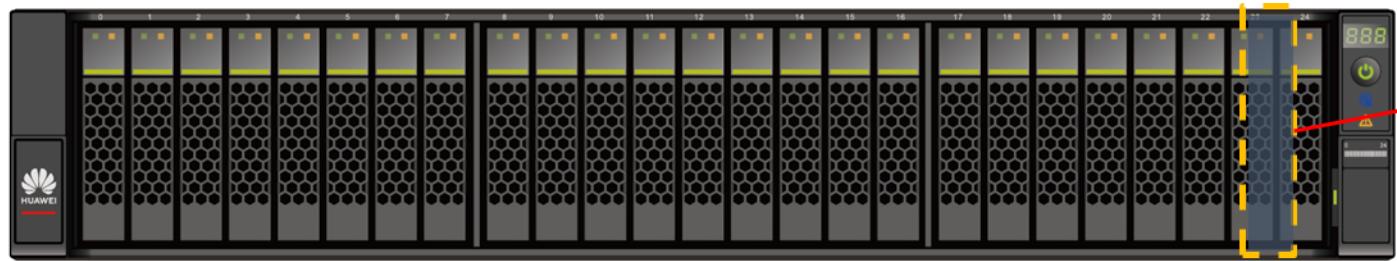


Figure 1

- 2U controller enclosure with 25 x 2.5-inch integrated SSDs
 - 12 Gb SAS SSDs
 - 960 GB/1.92 TB/3.84 TB/7.68 TB/15.36 TB/30.72 TB

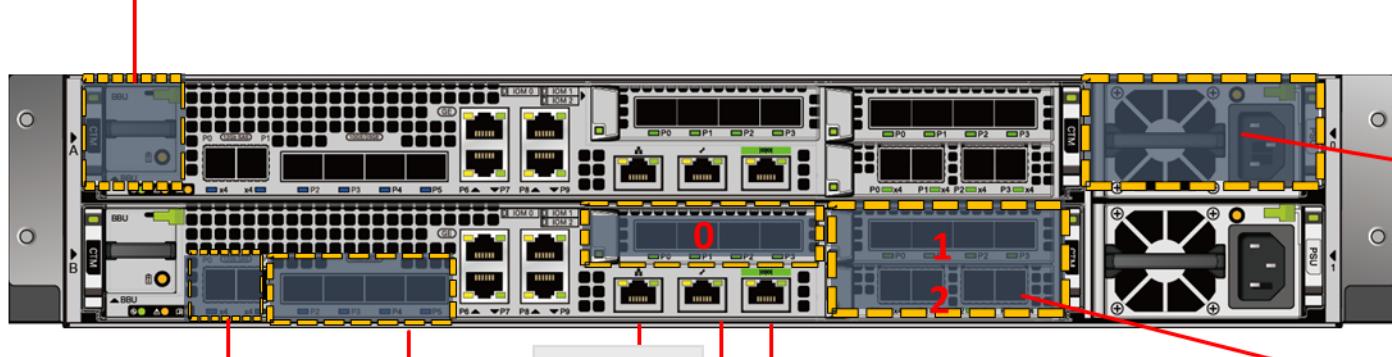


Figure 2

- One BBU per controller, supporting hot swap

Integrated power-fan module

- Power supply: 1+1 redundancy, supporting 100–240V AC and 240V HVDC
- Fan: Four built-in fans (3+1 redundancy)

- Onboard SAS expansion port
 - Two SAS expansion ports per controller

- Onboard port
 - Four onboard 10Gb ETH per controller

- Management port
- Serial port
- Maintenance port

Interface module

- Three hot-swappable interface modules per controller
- Port type: 8/16/32 Gb FC/FC-NVMe, 10/25/40/100 Gb ETH, and 12 Gb SAS
- The scale-out and scale-up interface modules can be installed only in Slot 1 and Slot 2 respectively.
- Dual controllers support up to 150 SAS SSDs.

OceanStor Dorado Disk Enclosure Form Factor

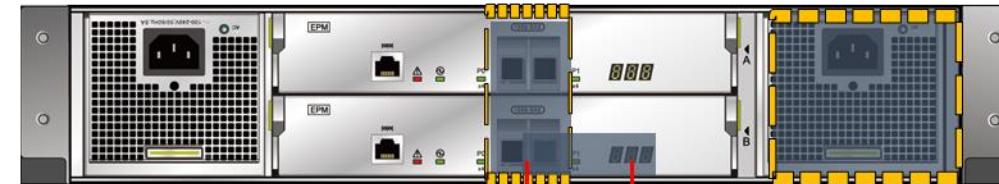
SAS enclosure front panel (common or intelligent)



- 25 2.5-inch SAS SSDs
- 12 Gb SAS SSDs
 - 960 GB/1.92 TB/3.84 TB/7.68 TB/15.36 TB/30.72 TB

Figure 1

SAS enclosure rear panel (common)



- Two onboard SAS 3.0 expansion ports per controller

Management port

- Power supply: 1+1 redundancy, supporting 100–240V AC and 240V HVDC

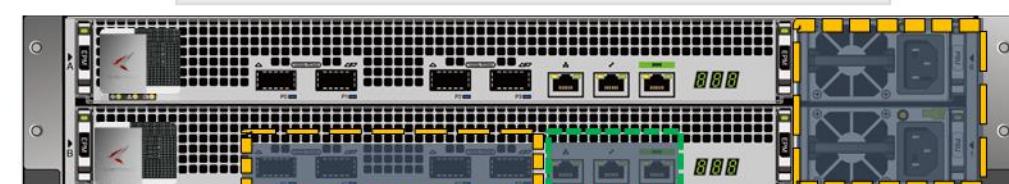
NVMe enclosure front panel



- 36 palm-sized NVMe SSDs
- Supports dual-port PCIe 3.0 x 2 NVMe SSDs
 - 1.92 TB/3.84 TB/7.68 TB/15.36 TB

Figure 2

Rear panel of an NVMe or intelligent SAS enclosure



- Four onboard 100G RDMA expansion ports per controller

Management port

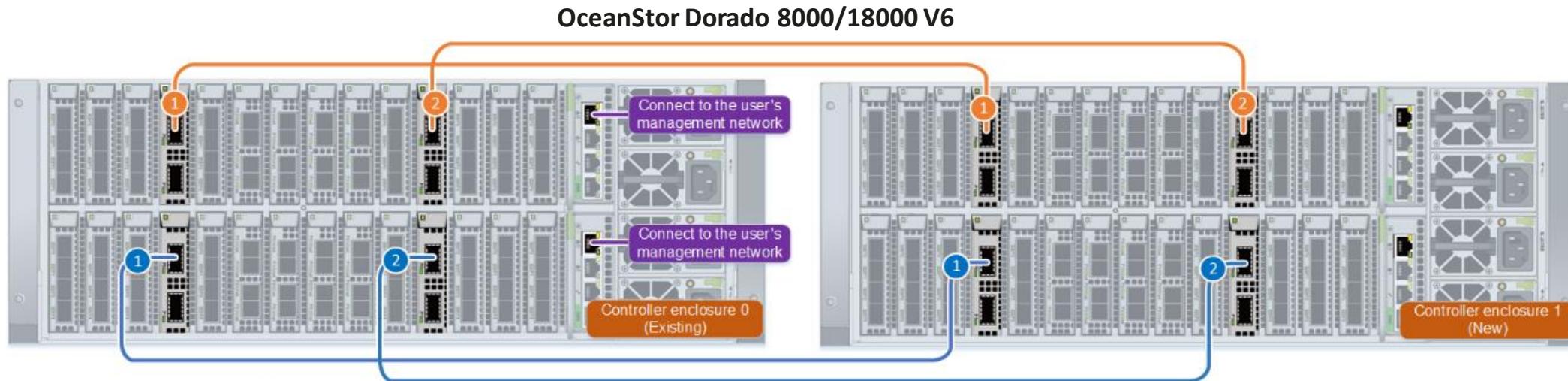
- Power supply: 1+1 redundancy, supporting 100–240V AC and 240V HVDC

OceanStor Dorado SSD Form Factor



	2.5-inch SSD	Palm-sized SSD
Dimensions (mm)	100.6 x 70 x 14.8	160 x 79.8 x 9.5
Thickness with handle (mm)	25	18.5
Weight/disk	0.25 kg	0.2 kg
Number of disks (2U disk enclosure)	25	36
Type	SAS SSD	NVMe SSD
Capacity	960 GB/1.92 TB/3.84 TB/7.68 TB/15.36 TB/30.72 TB	1.92 TB/3.84 TB/7.68 TB/15.36 TB

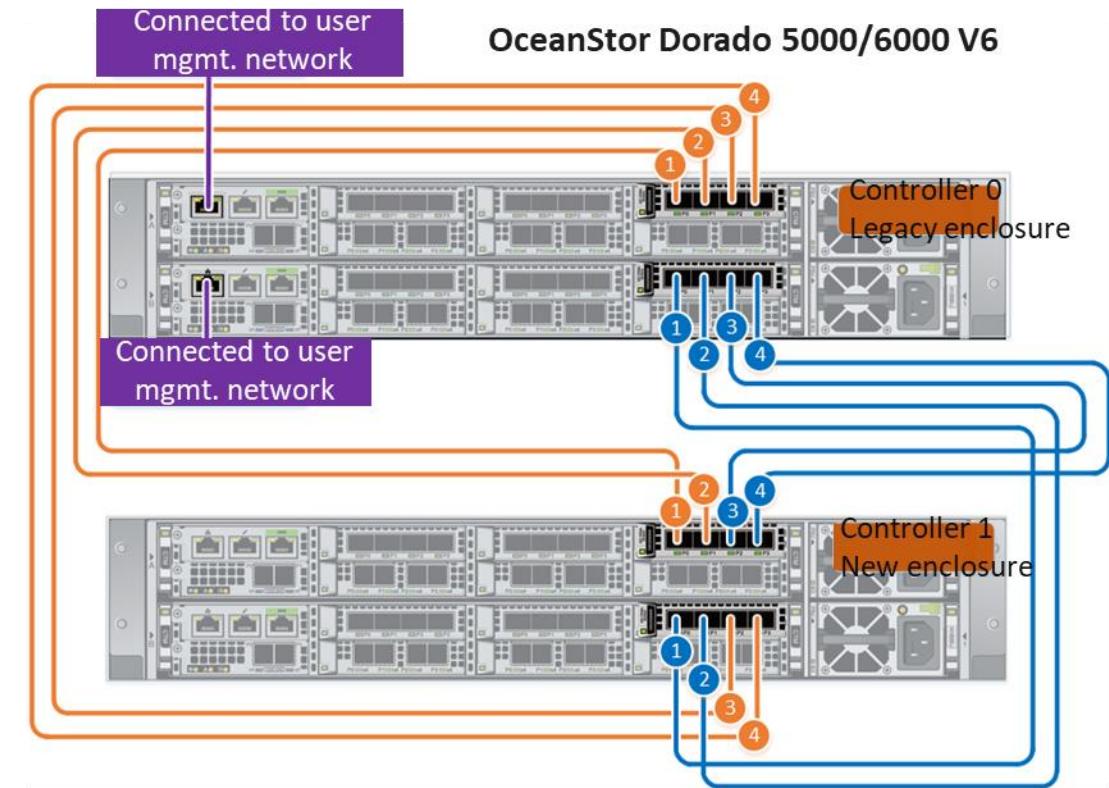
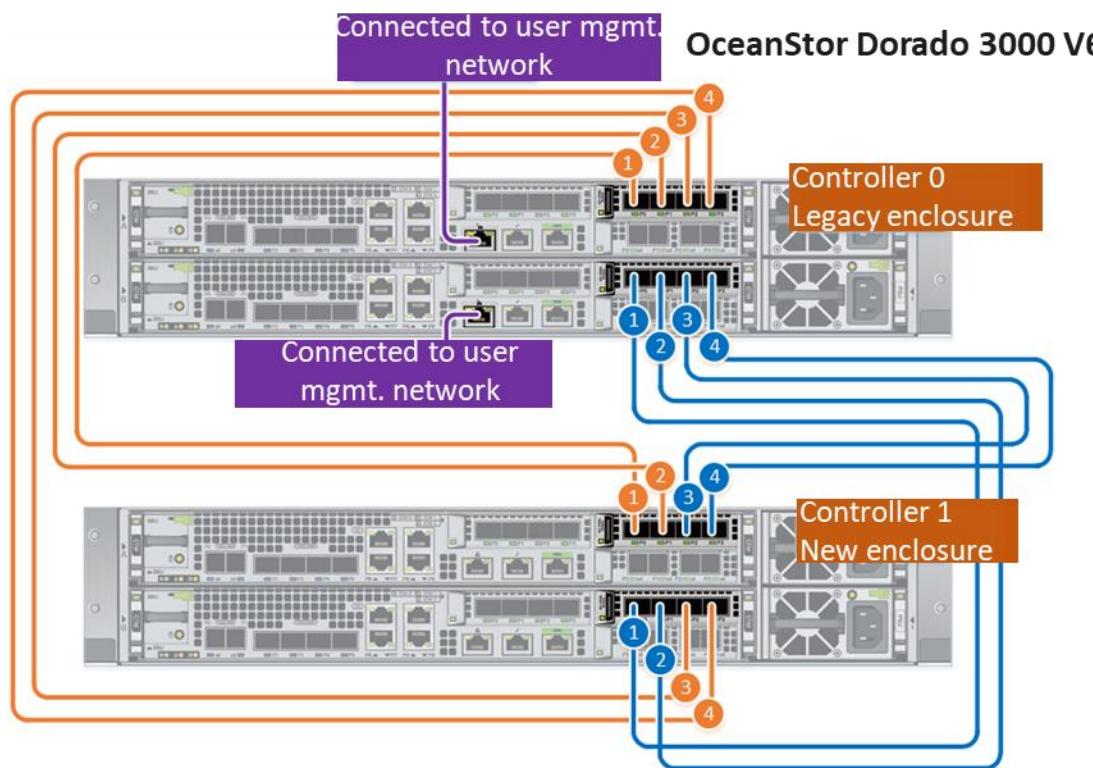
OceanStor Dorado High-End Storage Scale-Out Networking



- The current version supports the direct connection of 8 controllers using 4 x 100G RDMA cables. 8 controllers comprise a cluster system with active-active load balancing.
- The current version supports switch-based interconnection of 16 controllers (32 controllers need to be evaluated separately). Each enclosure connects to the switches through 8 100G RDMA cables.
- Switch model: CE8850-SAN (without optical modules)



OceanStor Dorado Mid-Range & Entry Level Storage Scale-out Networking (Direct Connection)



- ❑ The current version supports the direct connection of 4 controllers using 8 25G RDMA cables. 4 controllers comprise a cluster system with active-active load balancing. **The direct connection model does not support online controller expansion via switches.**
- ❑ OceanStor Dorado 5000/6000 6.1.2 versions support 2/4/6/8/10/12/14/16 controllers through switches. By default, 100 Gb switches are used for new deployment projects. 25 Gb switches delivered earlier than version 6.1.2 can be upgraded to version 6.1.2 (no need to replace and use 100 Gb switches).

Quiz

1. (Multiple-choice) Which of the following Slots on OceanStor Dorado 8000 can be used Scale out?
 - A. H3&L3
 - B. H3&L10
 - C. H10&L3
 - D. H10&L10

2. (Single-choice) How many SSDs can be housed in OceanStor Dorado 3000 NVMe Enclosure?
 - A. 24
 - B. 25
 - C. 30
 - D. 36

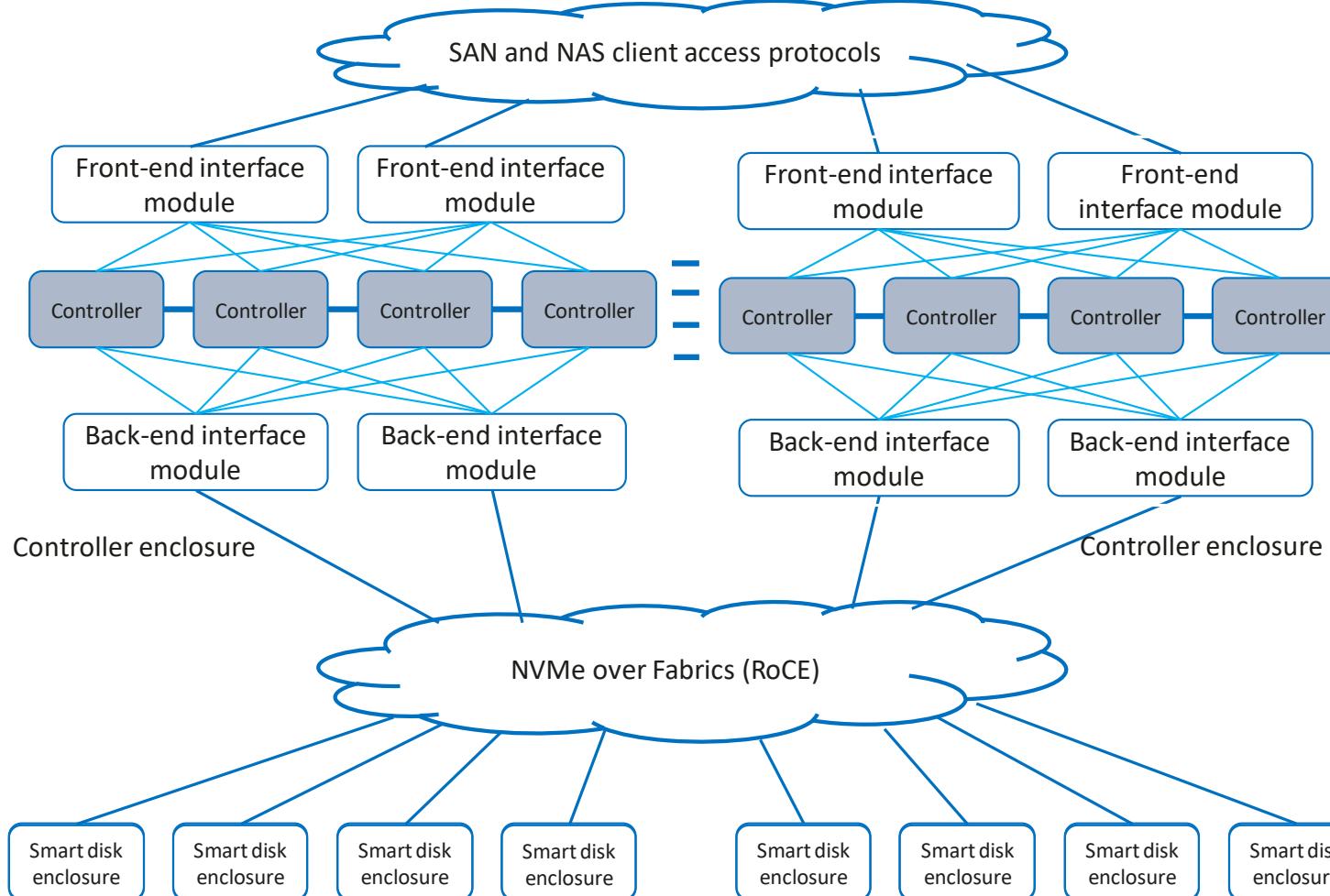
Contents

1. Overview
2. Hardware Architecture
- 3. Software Architecture**
4. Smart Series Features
5. Hyper Series Features
6. Other Key Features

Overview and Objectives

- This section describes the software architecture of Huawei OceanStor Dorado.
- On completion of this section, you will be able to:
 - Describe the overall software architecture of unified Storage;
 - Describe the design of key technologies for convergence and software features matrix.

Design Principle: Distributed, End-to-End NVMe, and Global Shared Resource (High-End)



Distributed Active-Active Architecture

- Distributed active-active architecture, symmetric host/client access
- Load balancing among all controllers and auto-rebalancing upon scale-out, failover, and fallback. When a host reads and writes one LUN on the OceanStor Dorado, the loads are balanced among all storage controllers.

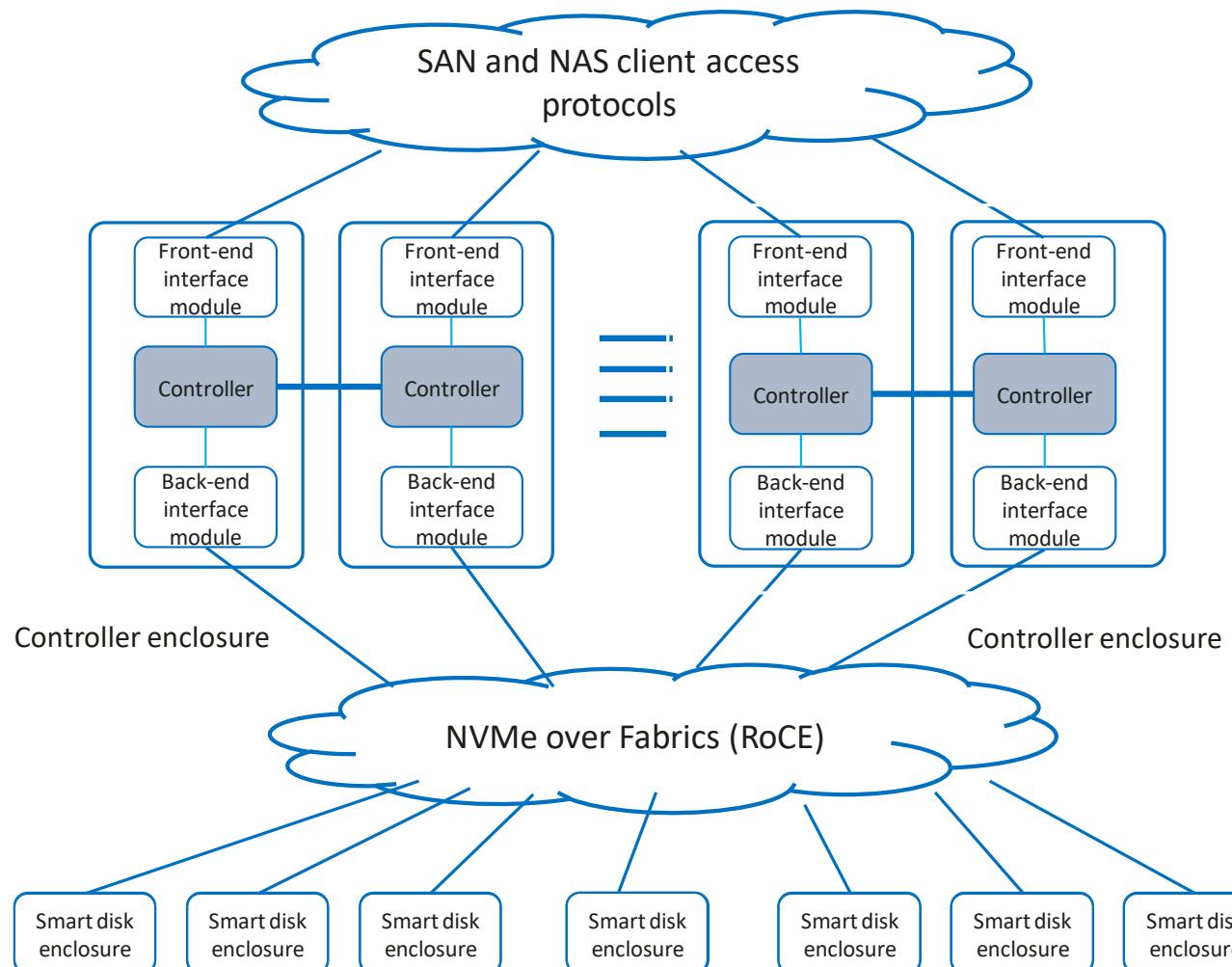
E2E Full-Stack Offload

- Front end: DTOE NICs offload the TCP/IP protocol stack.
- Back end: Smart disk enclosures and NVMe SSDs, high-end storage supports at most 8 controllers concurrently access one SSD

Global Resource Sharing

- Global sharing of cache and pool resources
- High-end devices support back-end interconnect I/O modules (BIMs, 100 Gbit/s RDMA).

Design Principle: Symmetric, End-to-End NVMe, No.1 NVMe scalability (Mid-Range & Entry-Level)



Symmetric A-A architecture

- Global cache and service in one engine processing involving all controllers. Workloads on a single LUN are balanced among multiple controllers.
- Distributed file system: A single file system can be accessed from multiple controllers.

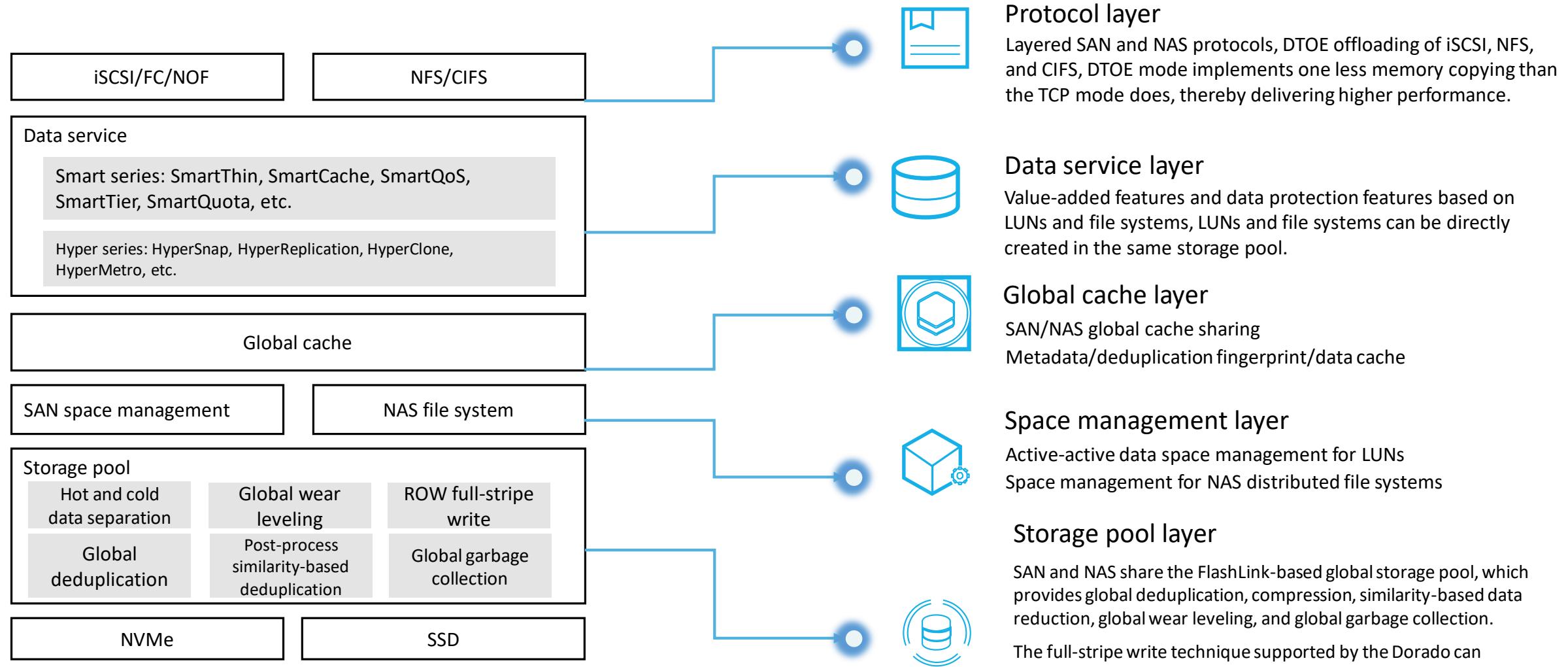
Scale-out/scale-up expansion

- Scale-out performance: RDMA high-speed interconnection between 16 controllers
- Scale-up capacity: A single engine supports a maximum of eight NVMe disk enclosures.

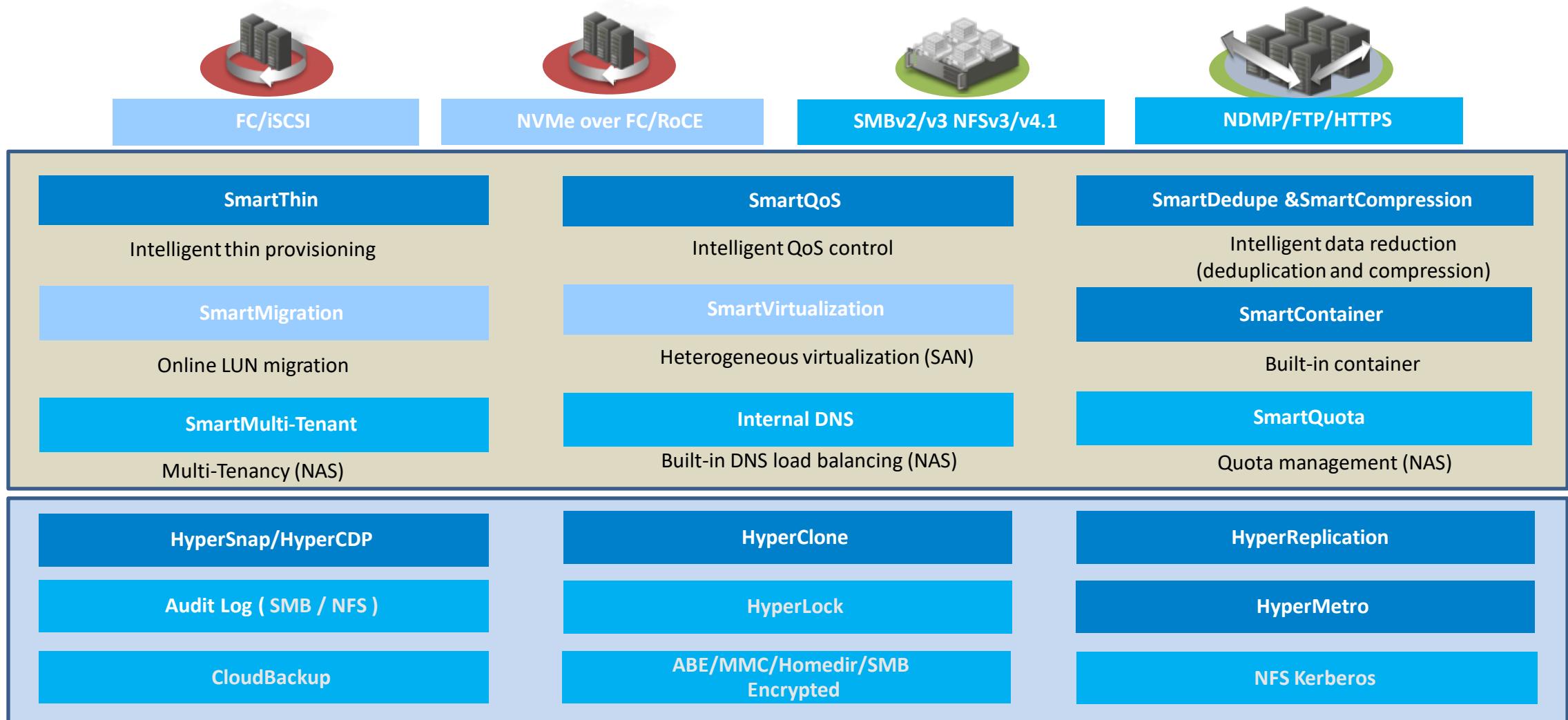
Superior performance

- FlashLink is used for intelligent disk-controller cooperative algorithm, doubling computing power.
- Optimized performance, 2x higher than the industry average

OceanStor Dorado Converged SAN and NAS Architecture



OceanStor Dorado Key Features



Note: For details about the protocol versions and delivery of minor functions, see the product specifications list.

Quiz

1. (Multiple-choice) Which of the Hyper features on OceanStor Dorado 8000 can be used for both SAN and NAS?
 - A. HyperSnap
 - B. HyperLock
 - C. HyperReplication
 - D. HyperMetro

2. (Multiple-choice) Which layer is included in OceanStor Dorado Converged SAN and NAS Architecture?
 - A. Application Layer
 - B. Protocol layer
 - C. Data service layer
 - D. Global cache layer
 - E. Space management layer
 - F. Storage pool layer

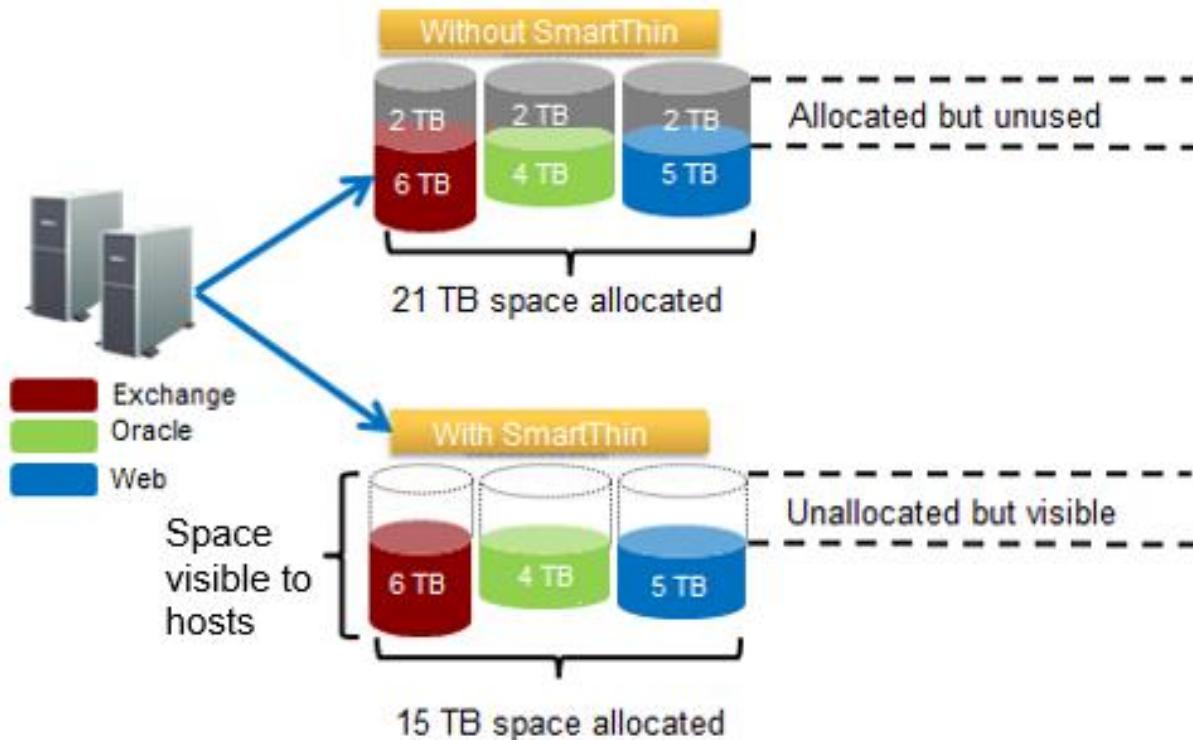
Contents

1. Overview
2. Hardware Architecture
3. Software Architecture
- 4. Smart Series Features**
5. Hyper Series Features
6. Other Key Features

Overview and Objectives

- This section describes the software features of Smart series of Huawei OceanStor Dorado.
- On completion of this section, you will be able to:
 - Describe the design of key technologies for Smart features.

SmartThin: The Thin Provisioning Function



SmartThin functions:

- Thin LUN capacity virtualization:** SmartThin allows the capacity detected by a host to be larger than the actual capacity of a thin LUN.
- Capacity-on-write:** SmartThin allocates space to a thin LUN when a host writes data to the thin LUN. The space allocated equals to the space required.
- Online thin LUN capacity expansion:** SmartThin supports storage pool expansion and thin LUN expansion.
- Thin LUN space reclamation:** SmartThin provides two space reclamation methods: standard Small Computer System Interface (SCSI) command reclamation and all-zero data space reclamation. Space reclaimed becomes the unused capacity of storage pools.

SmartThin improvements:

- Real space is allocated on demand when LUNs are being used.
- LUN space can be expanded dynamically.

SmartVirtualization: Manage Resources of Heterogeneous Storage

Relationship Between
an eDevLUN and an External LUN

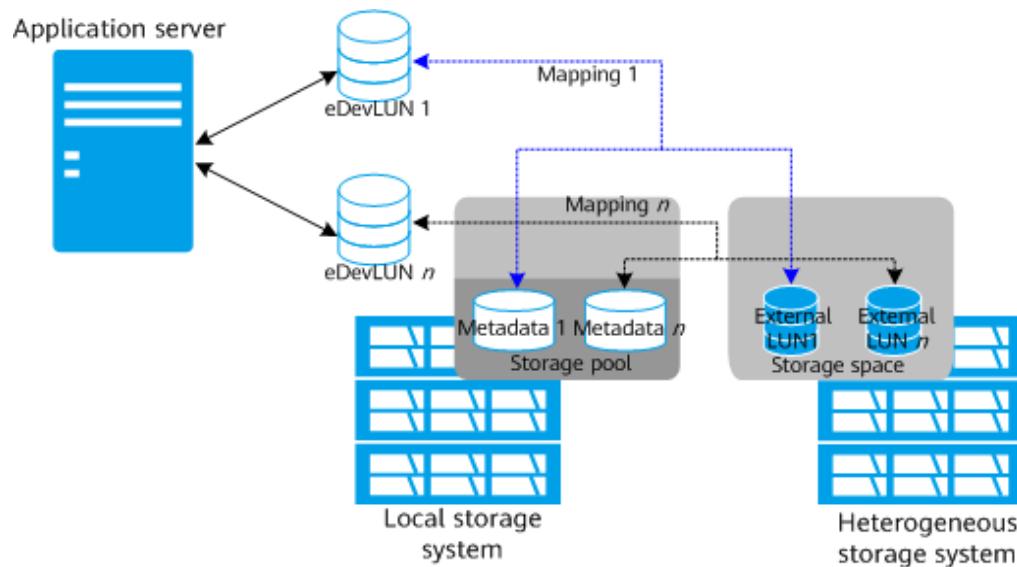


Figure1

Legend:
— Metadata space
— Data space
↔ Mapping
↔ Read and write I/O

Storage resource management
across storage systems

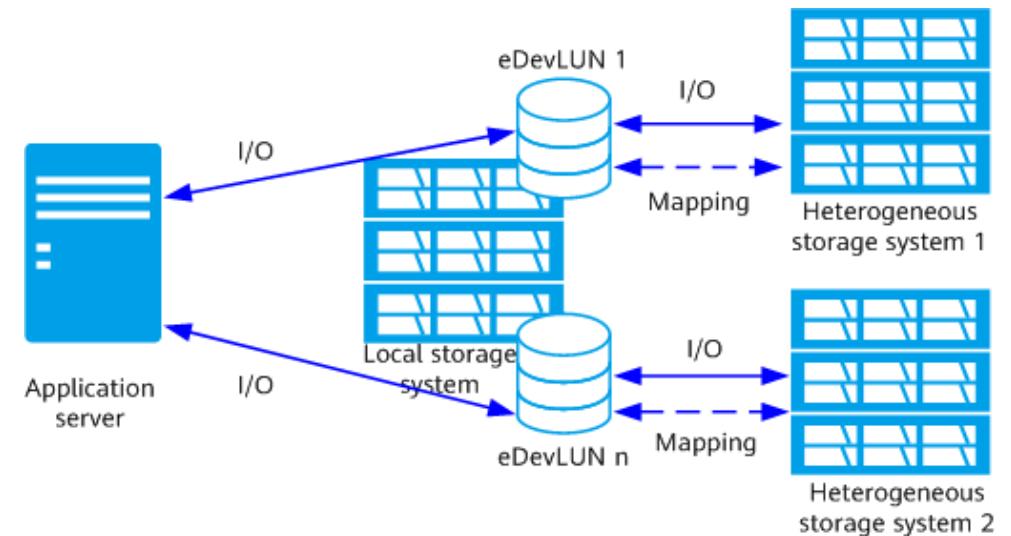


Figure 2:

SmartVirtualization: enables the local storage system to use and manage storage resources of the peer storage system as local storage resources despite of the different software and hardware architectures

SmartMigration: Non-Disruptive Migration



Figure1:Data Migration

Source LUN:

LUN from which service data is migrated.

Target LUN:

LUN to which service data is migrated.

LM module:

Manages SmartMigration in the storage system.

Pair:

In SmartMigration, a pair indicates the data migration relationship between the source LUN and target LUN. A pair can have only one source LUN and one target LUN.

Dual-write:

The process of writing data to the source and target LUNs at the same time during service data migration.

SmartMigration: It is a key technology for service migration. It can non-disruptively migrate service data within a storage system and between storage systems.

SmartMigration: Non-Disruptive Migration

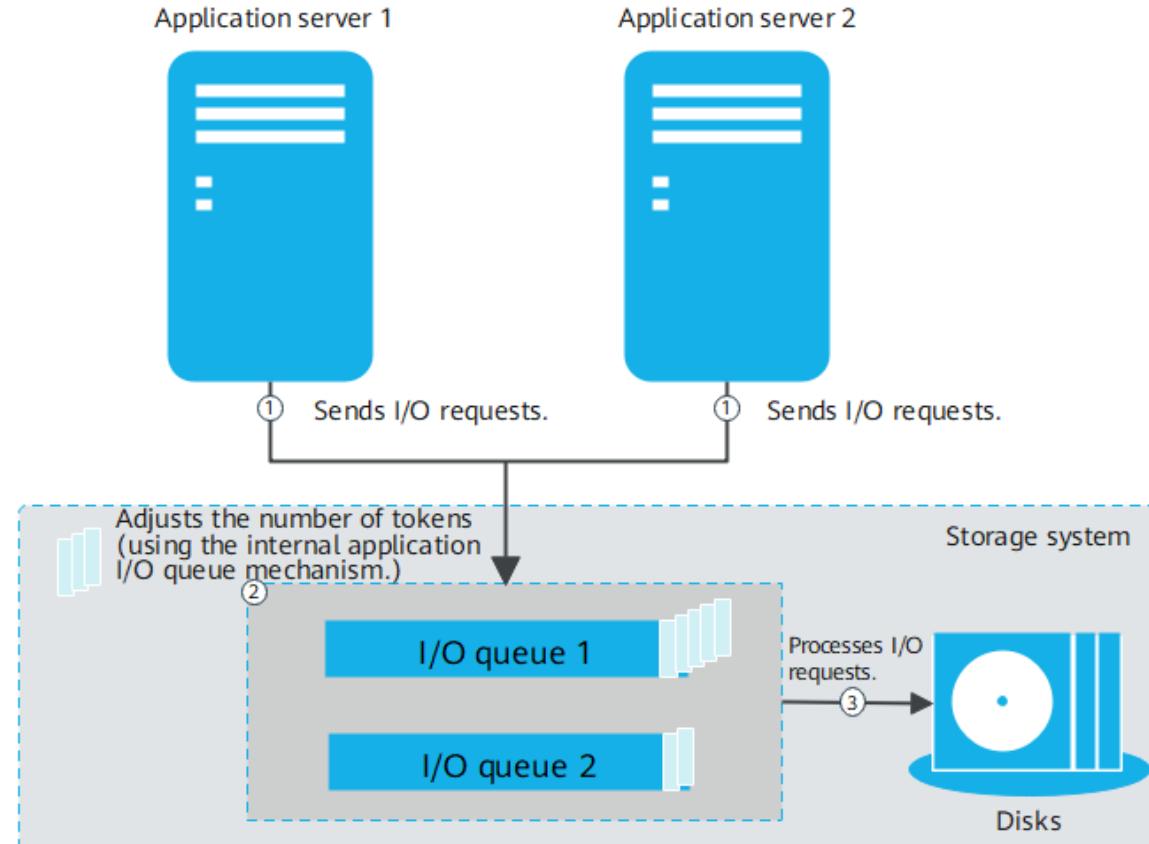


Figure1: I/O Traffic Control

SmartQoS (intelligent service quality control)

sets different performance indicators for applications to ensure the performance of mission-critical applications

1. **Priority policy:** The system allocates internal I/O queues into different classes (high, medium, and low), and
2. **Upper limits set by the flow control mechanism:** Users can limit the performance of non-critical applications by setting upper limits for their **IOPS** and **bandwidth** and **latency**, preventing these applications from occupying too many system resources.

SmartDedupe & SmartCompression: Intelligent Data Reduction

SmartDedupe

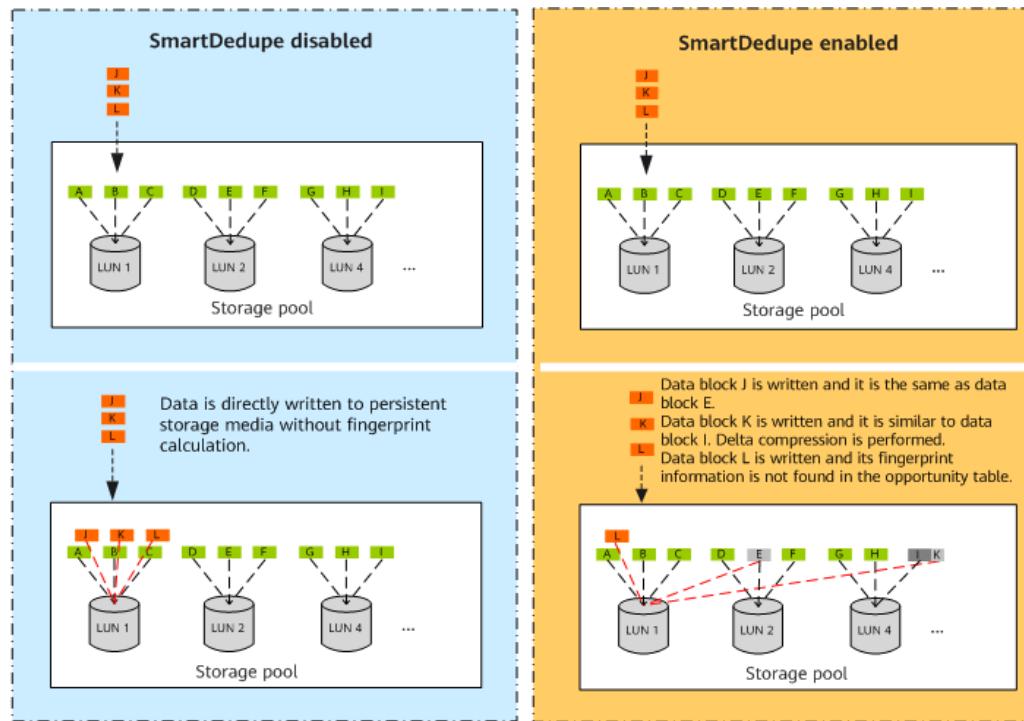


Figure1

SmartDedupe deletes duplicate data blocks from a storage system, reducing storage space consumption. Storage systems support Inline deduplication, similarity-based deduplication and Post-process deduplication.

SmartCompression

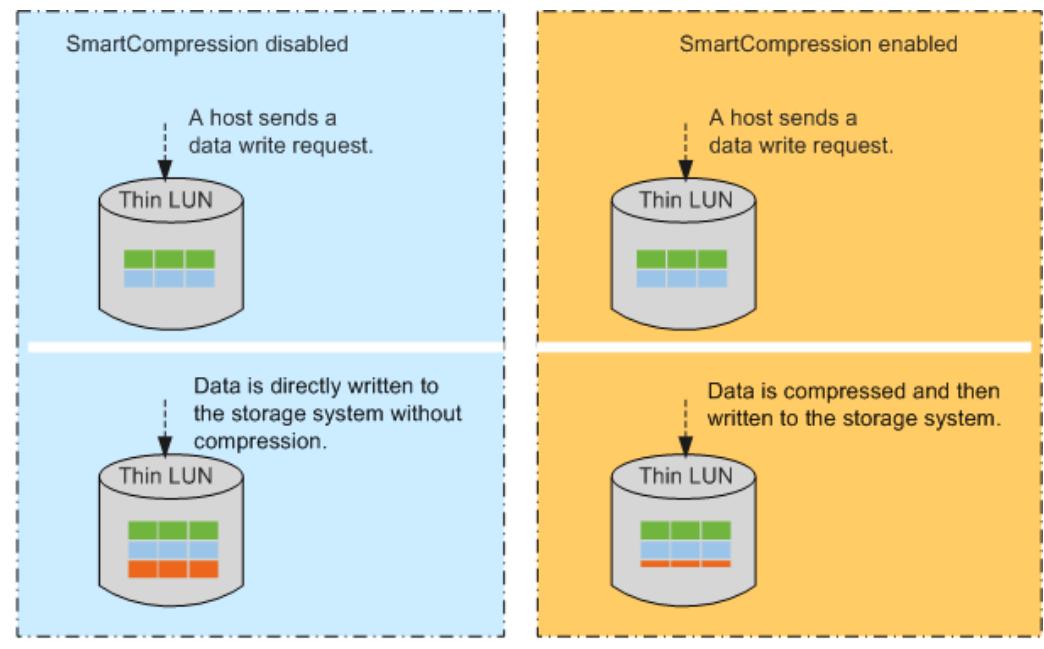


Figure2

SmartCompression reorganizes data to save storage space and improve the data transfer, processing, and storage efficiency without any data loss. Storage systems support Data compaction and inline compression.

SmartMulti-Tenant for SAN: Logical Service & Network Isolation

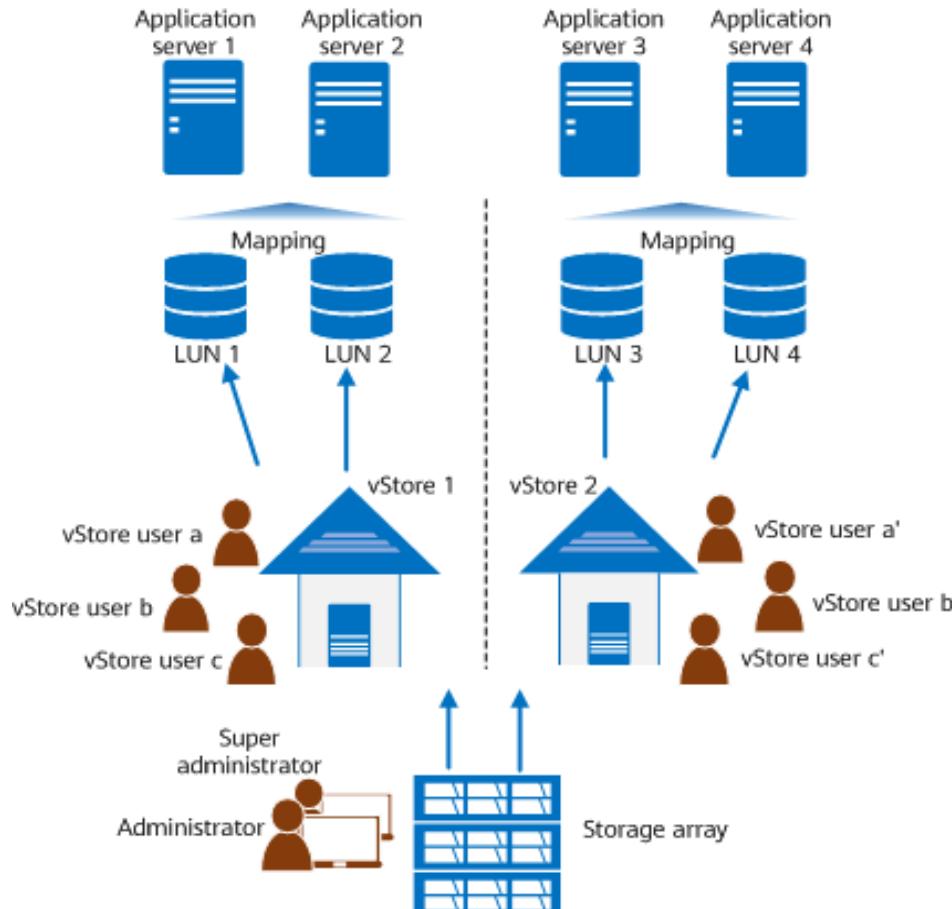


Figure1

SmartMulti-Tenant allows vStores to create multiple virtual storage systems in one physical storage system. With SmartMulti-Tenant, vStores can share hardware resources and safeguard data security and confidentiality in a multi-protocol unified storage architecture.

Advantages of SmartMulti-Tenant In real practices, SmartMulti-Tenant aims to improve resource utilization efficiency and reduce the per-unit cost by fully consolidating resources, it implements logical isolation to ensure service and network security of vStores. Resources of one vStore are invisible to other vStores.

System role: a system default or user-defined role that can create vStores and allocate storage resources on a storage system.

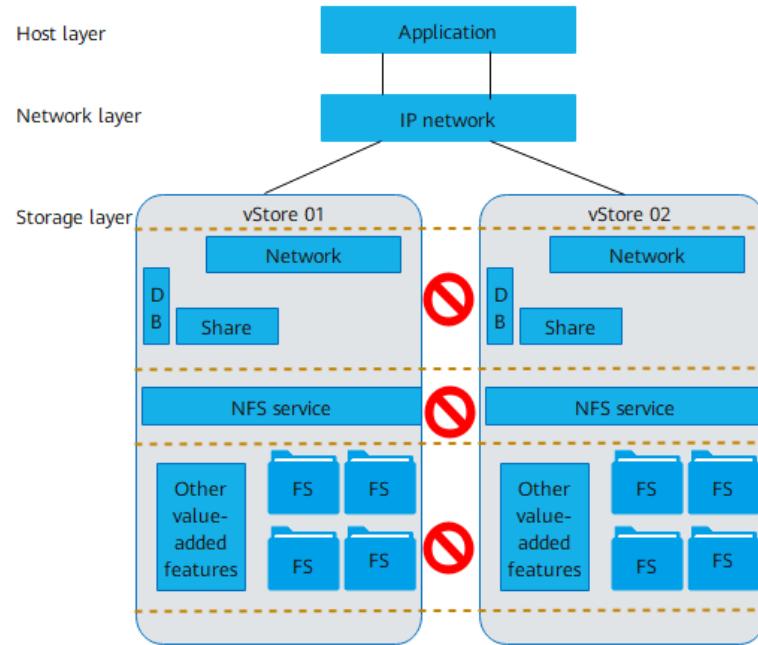
vStore role: a default or user-defined role that can complete vStore settings in the vStore view.

System view: A super administrator allocates and manages global resources of a storage system through this view.

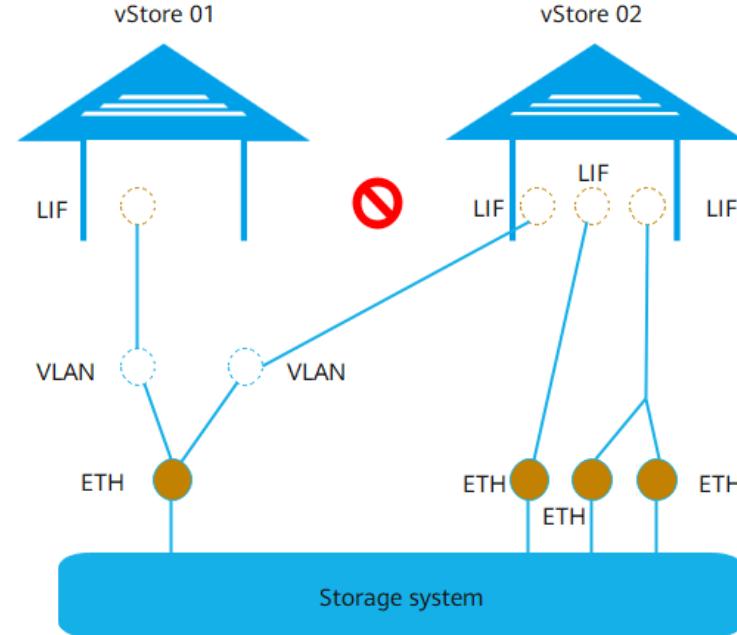
vStore view: A super administrator and a vStore user allocate and manage storage resources of vStores through this view.

SmartMulti-Tenant for NAS: Management/Network/Service Isolation

Service isolation: Different AD, LDAP, and NIS domain services can be configured for each vStore. Independent local users and user mappings to provide NFS and SMB services can be also configured separately.



Network isolation: Each vStore provides service access only through its own LIFs. A client that connects to a specified LIF can only access the file systems and snapshots of the vStore that owns the LIF.

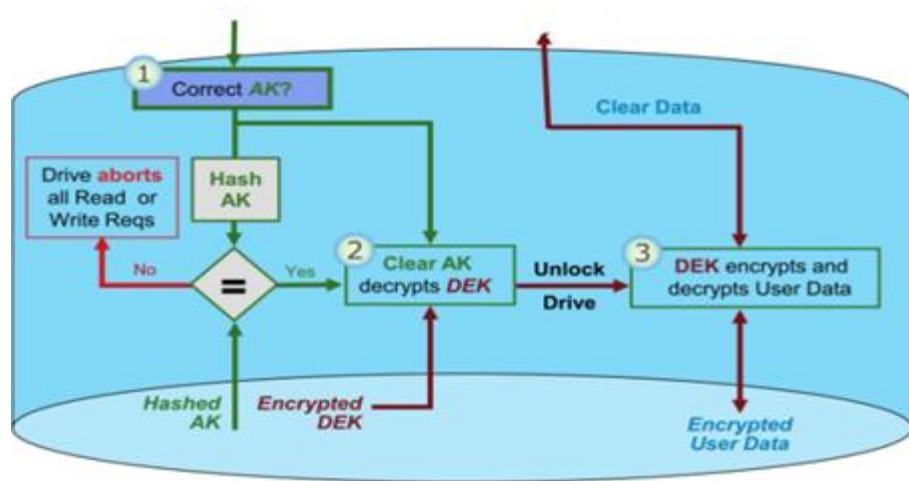


Management isolation: The vStore administrators can configure and manage their own storage resources only through the GUI or RESTful APIs.

SmartEncrypt and SmartErase: Encryption and Destruction

Data encryption:

Self-encrypting drives (SEDs) + key management server (internal or external) implement static encryption of persistent data. Data encryption complies with information security standards and protocols in computer-related fields defined by the Trusted Computing Group (TCG).

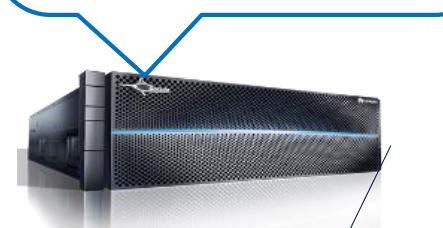


SED encryption principle:

- **Authentication key (AK)**: Each disk has an AK, which is stored on both the disk and key management server. It is used to authenticate the system for accessing disks.
- **Data encryption key (DEK)**: It is used to encrypt the data stored on disks. The data does not provide disk access interfaces.

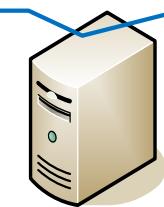
Internal key management service

NIST SP 800-57 provides layered protection and key generation, update, backup, recovery, and destruction functions. It is easy to use, configure, and manage.

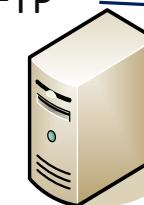


External key management service

It is mainly used in centralized key management or scenarios that require FIPS 140-2 certification. It provides layered protection and key generation, update, backup, recovery, and destruction functions.



SFTP



KMIP+TLS



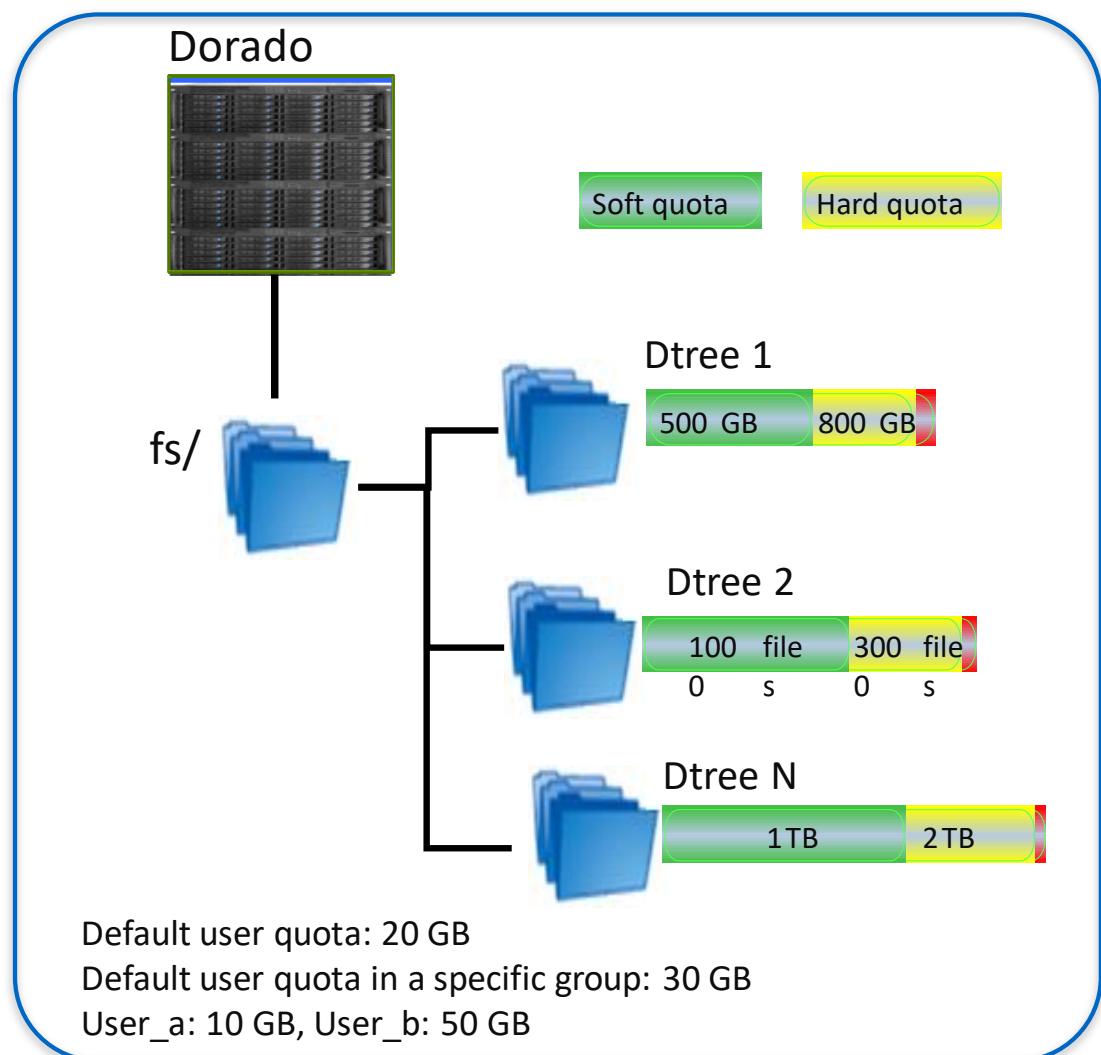
External key management service

Data destruction: Effectively erases disk data to prevent data leakage.

- Non-SED and SED: Block-level erasure or overwrite
- SED: DEK destruction

Note: The overwrite standards are **DoD 5220.22-M (E)**, **DoD 5220.22-M (ECE)**, **VSITR**, and **Custom**.

SmartQuota



		Directory Quota	User Quota	User Group Quota
Quota type	Space	Yes	Yes	Yes
	File quantity	Yes	Yes	Yes
Threshold	Space hard quota	Yes	Yes	Yes
	File quantity hard quota	Yes	Yes	Yes
Threshold	Space soft quota	Yes	Yes	Yes
	File quantity soft quota	Yes	Yes	Yes

- Dtrees can be created for multiple directories.
- File system quotas and dtree quotas can both take effect.
- Supports the default user quota, default group quota, and default user quota in a specific group.

Quiz

1. (Multiple-choice) Which Smart series features are supported by OceanStor Dorado?
 - A. SmartThin
 - B. SmartQos
 - C. SmartPartition
 - D. SmartVirtualization

2. (Multiple-choice) Which technologies does OceanStor Dorado support to increase data reduction ratio?
 - A. SmartDedupe
 - B. SmartMigration
 - C. SmartMulti-Tenant
 - D. SmartCompression

Contents

1. Overview
2. Hardware Architecture
3. Software Architecture
4. Smart Series Features
- 5. Hyper Series Features**
6. Other Key Features

Overview and Objectives

- This section describes the software features of Hyper series of Huawei OceanStor Dorado.
- On completion of this section, you will be able to:
 - Describe the design of key technologies for Hyper features.

HyperSnap (SAN&NAS): Lossless High-Density Snapshot based on ROW

What is ROW ?

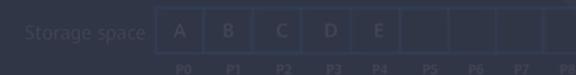
Working Principles



A snapshot and its source data share the storage space. You do not need to plan independent storage space for snapshots.

How does it achieve Performance Lossless?

Working Principles — Zero Performance Loss



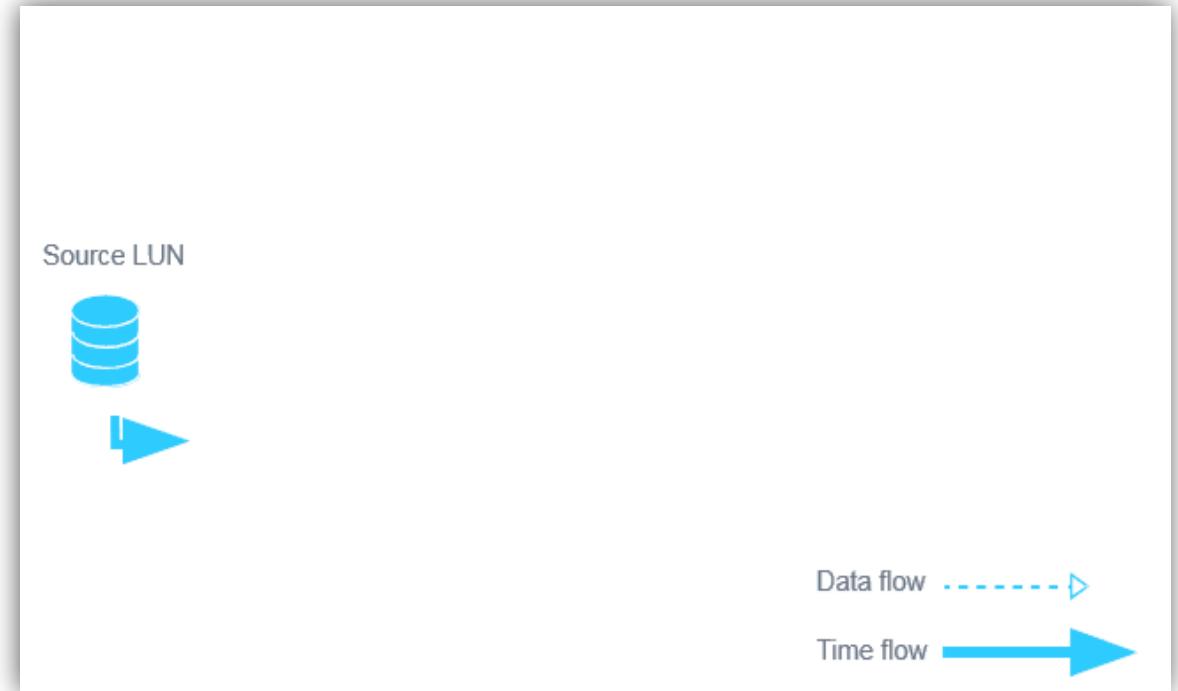
Data snapshots are independent of each other. Users can access snapshots created at different points in time to complete different application purposes.

HyperCDP(SAN&NAS): Creates High-Density Snapshots to Provide Continuous Data Protection.

Rapid Data Backup and Restoration



Re-purposing of Backup Data

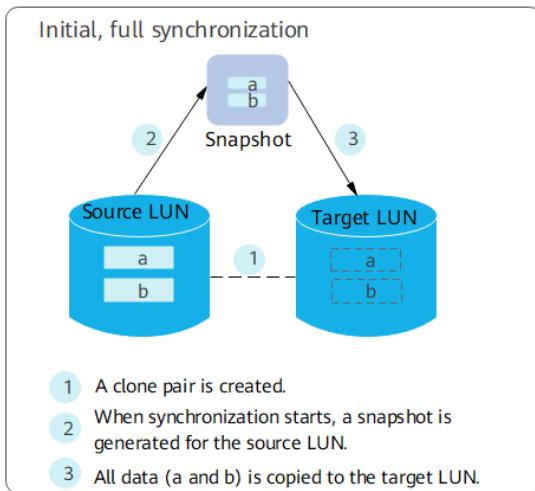


* When configuring the **HyperCDP Schedule**, activate **Secure Snapshot** feature directly to automatically create snapshots.

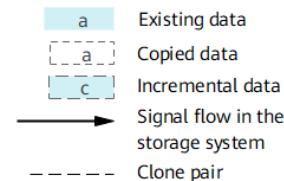
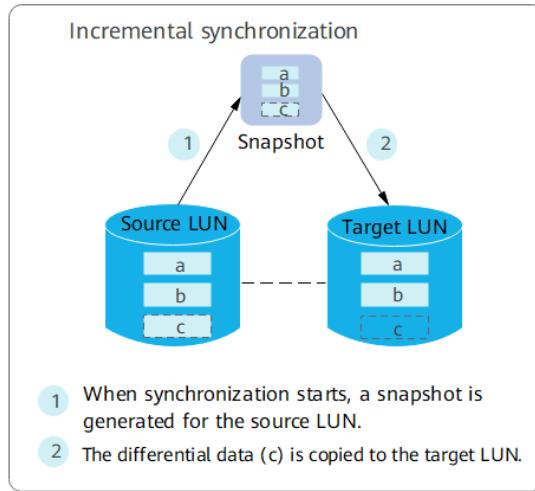
** Secure Snapshot Protection period setting (1 day to 20 years).

HyperClone(SAN&NAS): Physical Isolation of Data

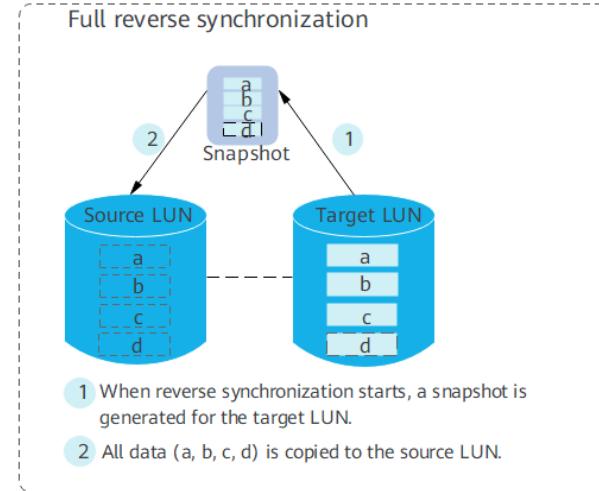
HyperClone Synchronization



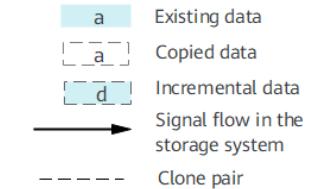
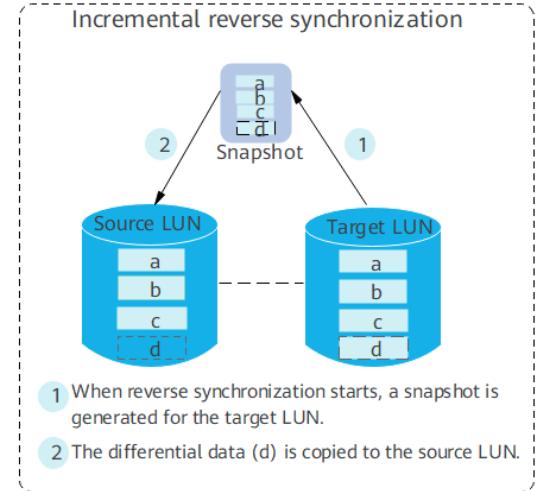
HyperClone creates a full copy of the source LUN's data on a target LUN at a specified point in time (synchronization start time). The target LUN can be read and written immediately, and you do not need to wait for the copy process to complete.



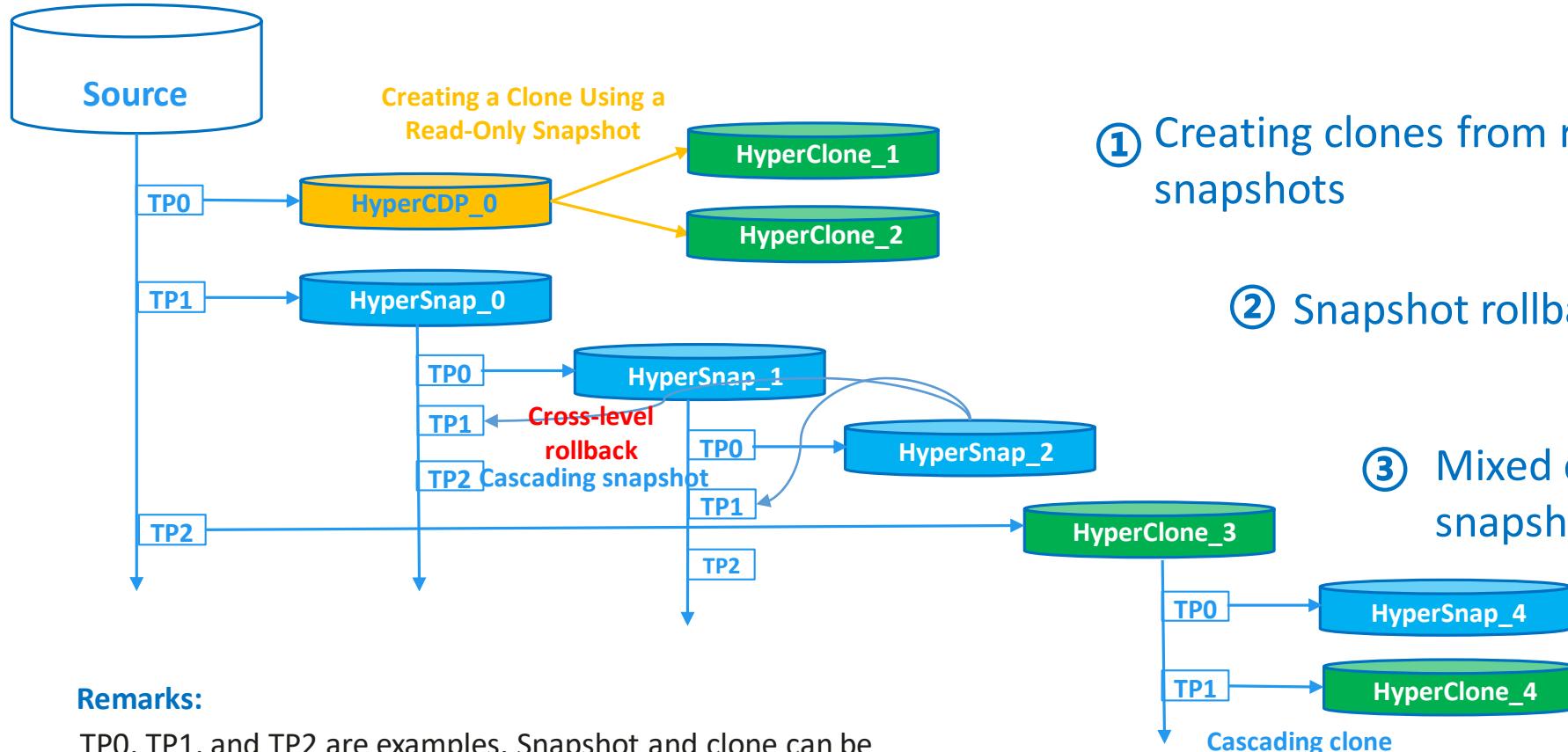
HyperClone Reverse Synchronization



HyperClone not only synchronizes data from the source LUN to the target LUN, but can also reversely synchronize data from the target LUN to the source LUN. If the source LUN suffers data corruption, you can restore the data from the target LUN.

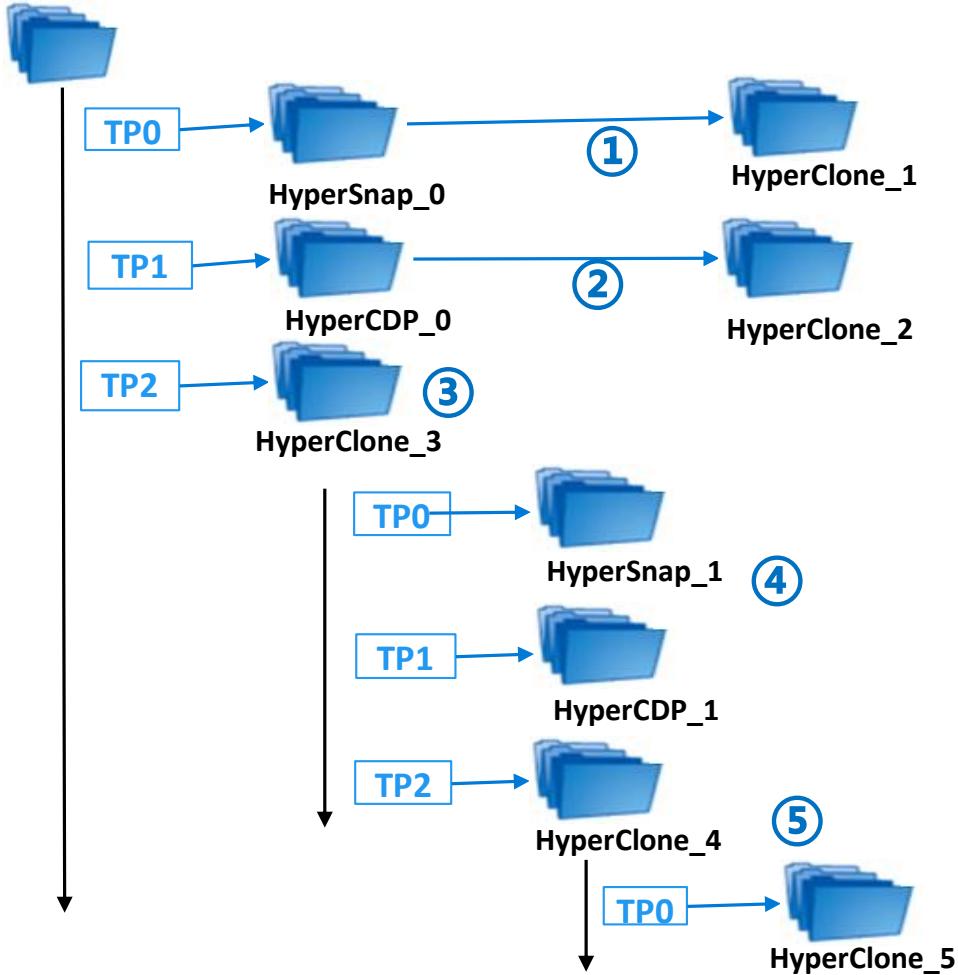


HyperSnap/HyperCDP/HyperClone (SAN)



- ① Creating clones from read-only snapshots
- ② Snapshot rollback as required
- ③ Mixed cascading of snapshots and clones

HyperSnap/HyperCDP/HyperClone (NAS)

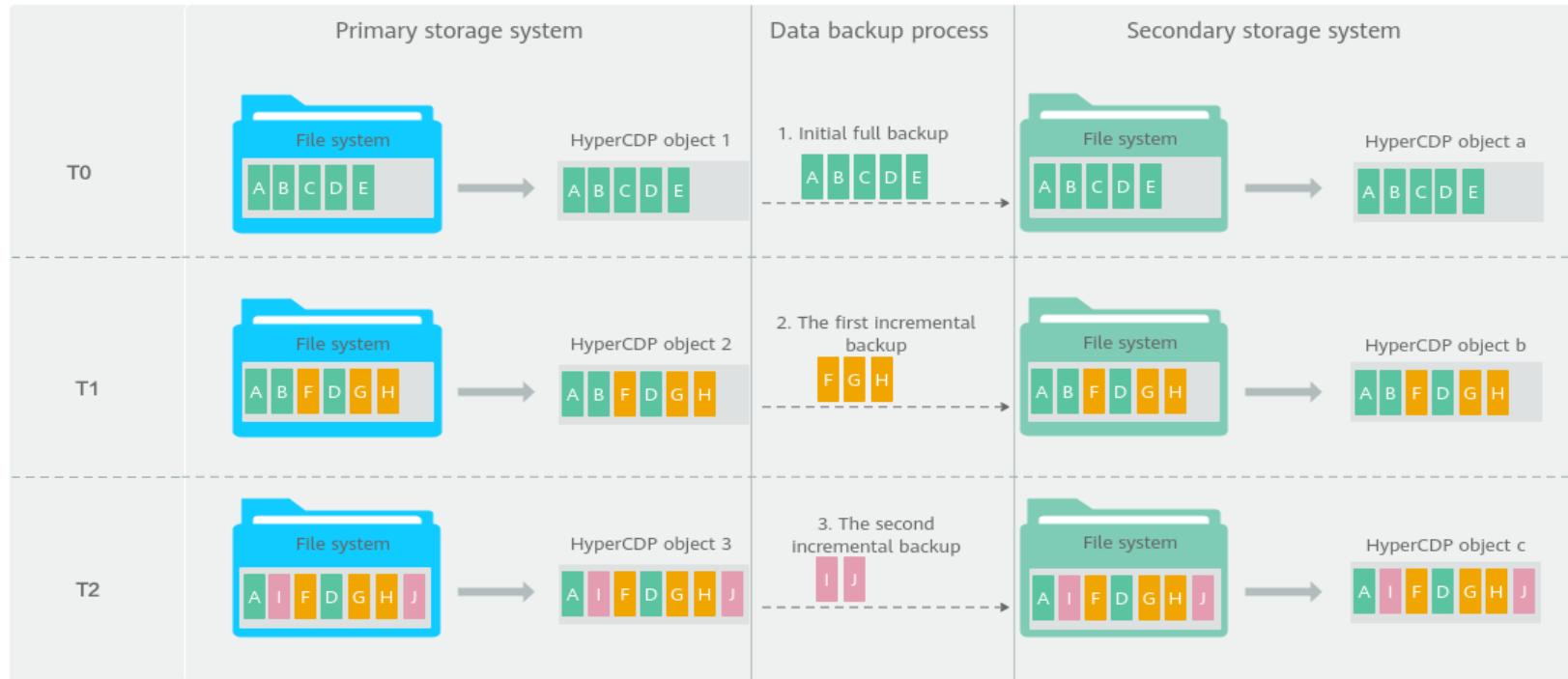


- ① Creating a clone file system based on an existing snapshot
- ② Creating a clone file system based on a HyperCDP object
- ③ Creating a clone file system based on a source file system
- ④ Creating a snapshot or HyperCDP object for a clone file system
- ⑤ Cascading clones (up to 8 levels)

HyperVault (HyperCDP + HyperReplication)

Working Principles

- Local backup uses the local HyperCDP policy.
- Remote backup synchronizes snapshots with tags asynchronously and backs them up remotely.



- HyperVault implements data backup and recovery within a storage system and between storage systems based on **HyperCDP** and the snapshot synchronization function of **HyperReplication**.
- Backup and recovery within a storage system is **called local backup and recovery**, and backup and recovery between storage systems is called **remote backup and recovery**.

HyperReplication for SAN: For Remote DR and Backup of Data

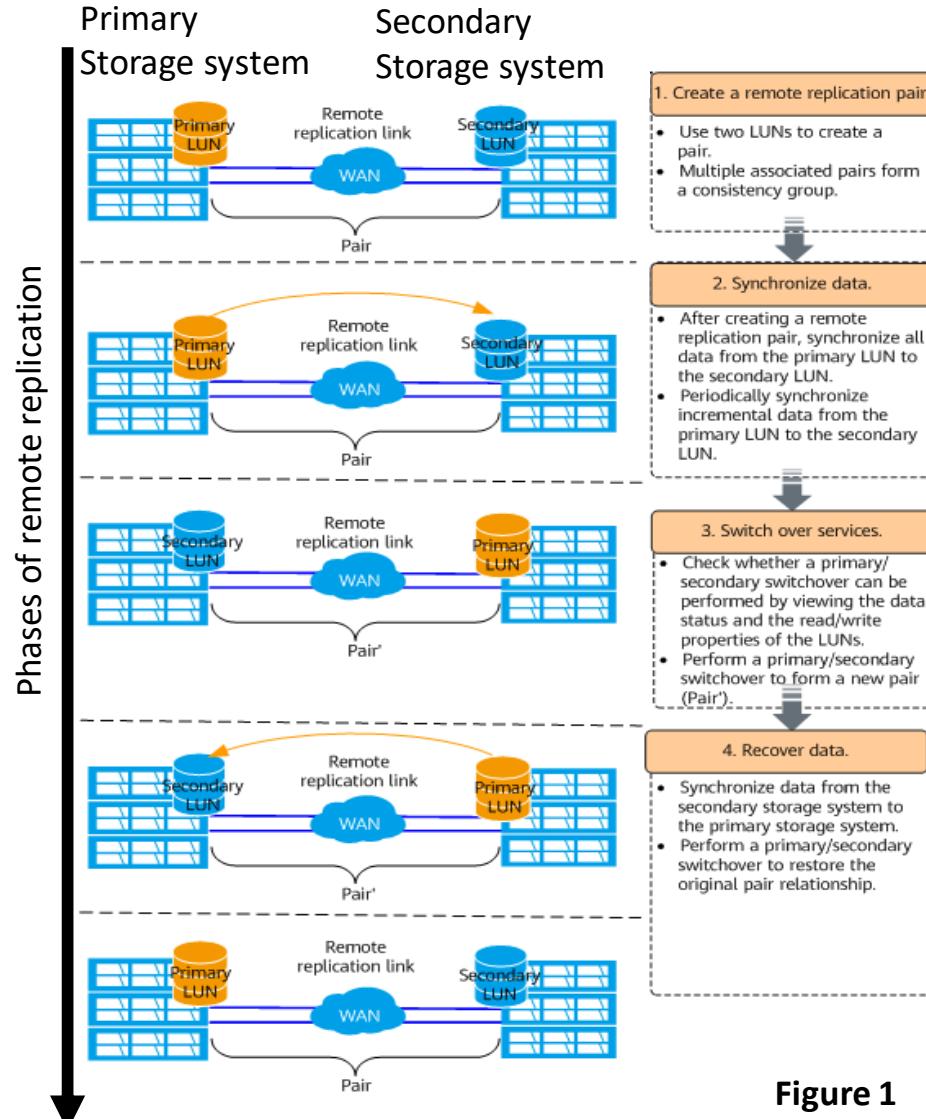


Figure 1

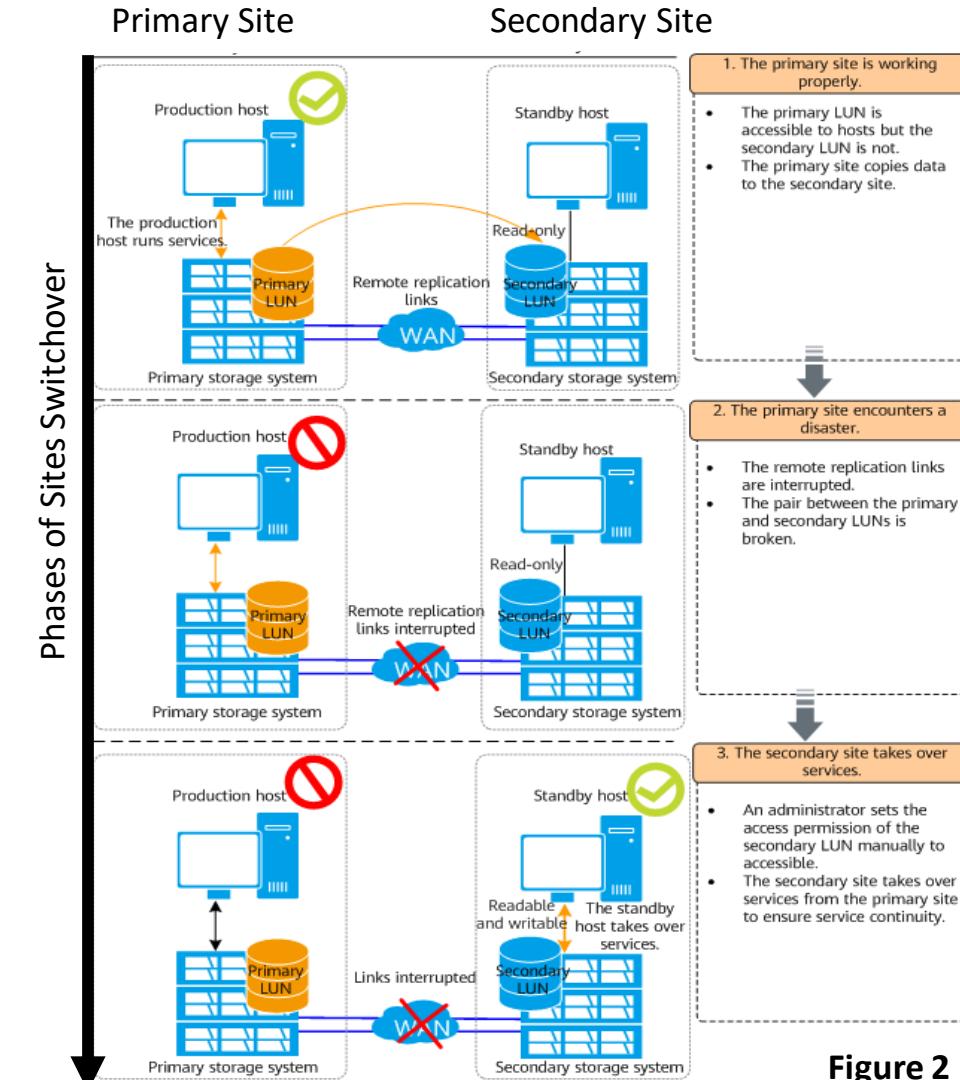
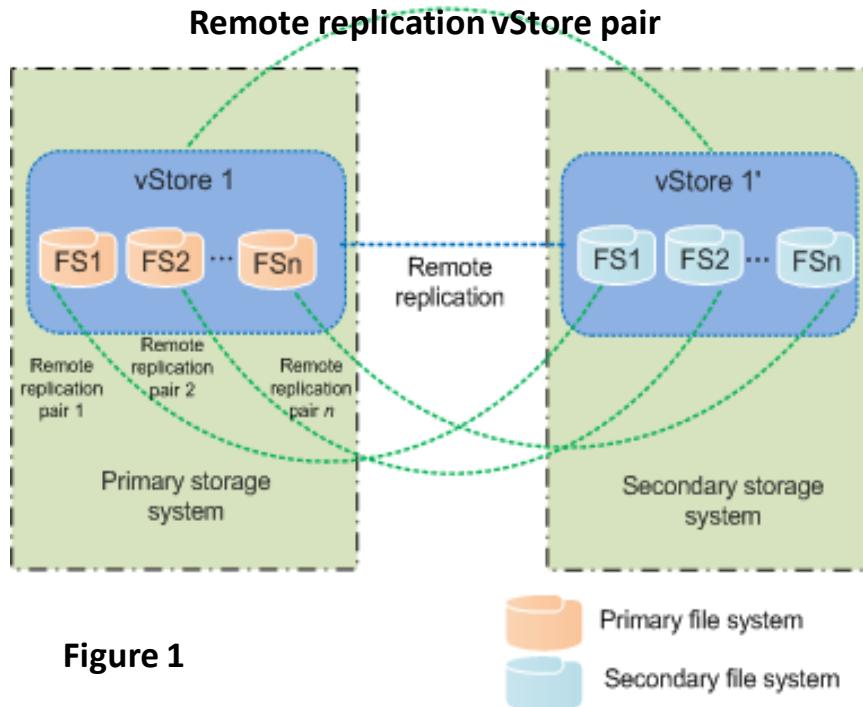


Figure 2

HyperReplication for NAS: For Remote DR and Backup of Data



- A **remote replication vStore pair** indicates the relationship between vStores on the primary and secondary storage systems.
- A remote replication vStore pair synchronizes vStore configuration **information** and file system **data**.
- A remote replication pair only synchronizes **data** based on the file systems.

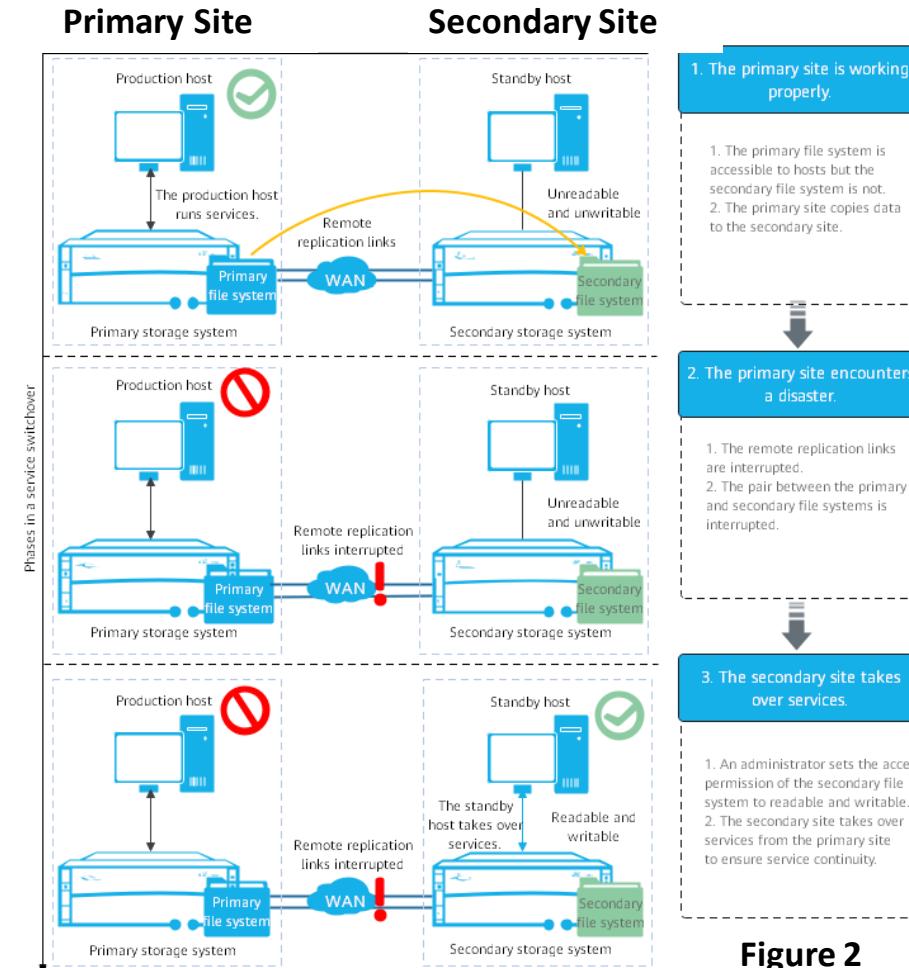
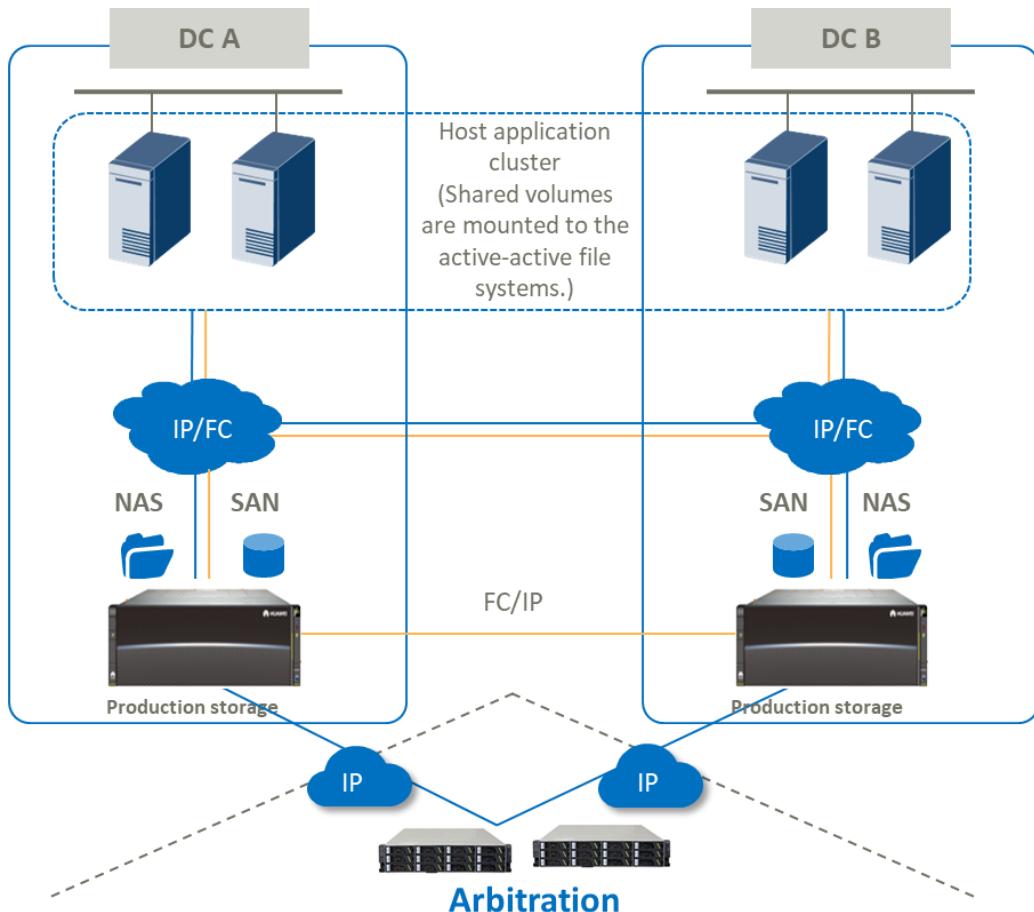


Figure 2

HyperMetro (SAN & NAS): Active-Active Storage Solution



***NVMe over Fabrics (NoF)** maps NVMe to various types of physical networks to extend NVMe implementation on the PCIe bus and achieve shared access of high-performance storage devices over fabrics.

****Remote Direct Memory Access over Converged Ethernet (RoCE)**

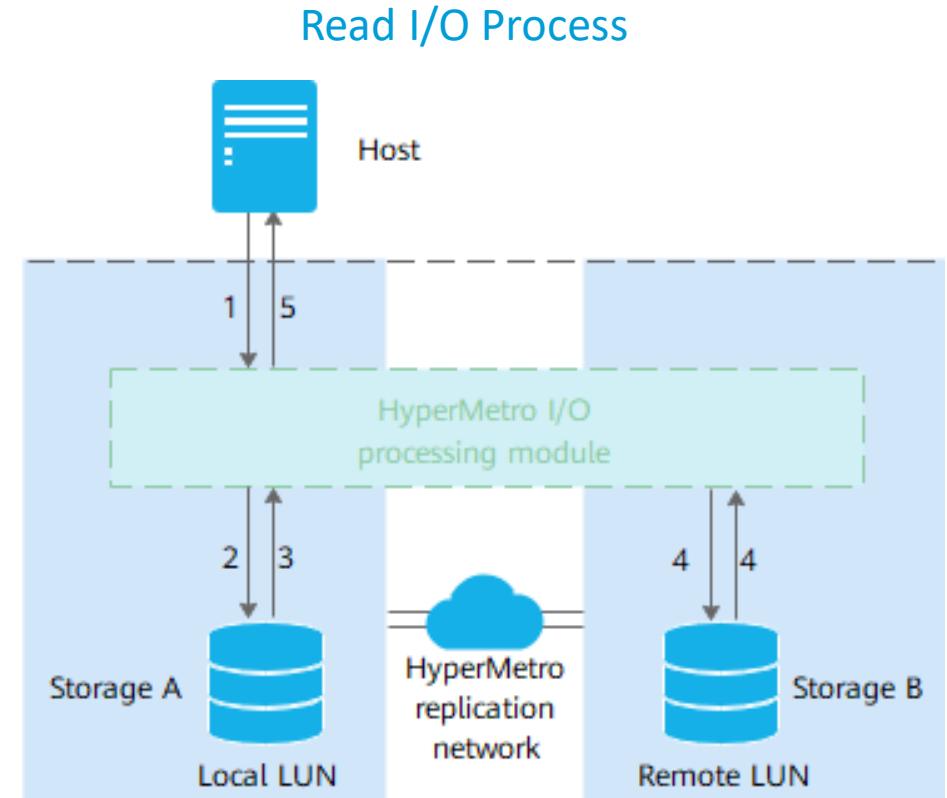
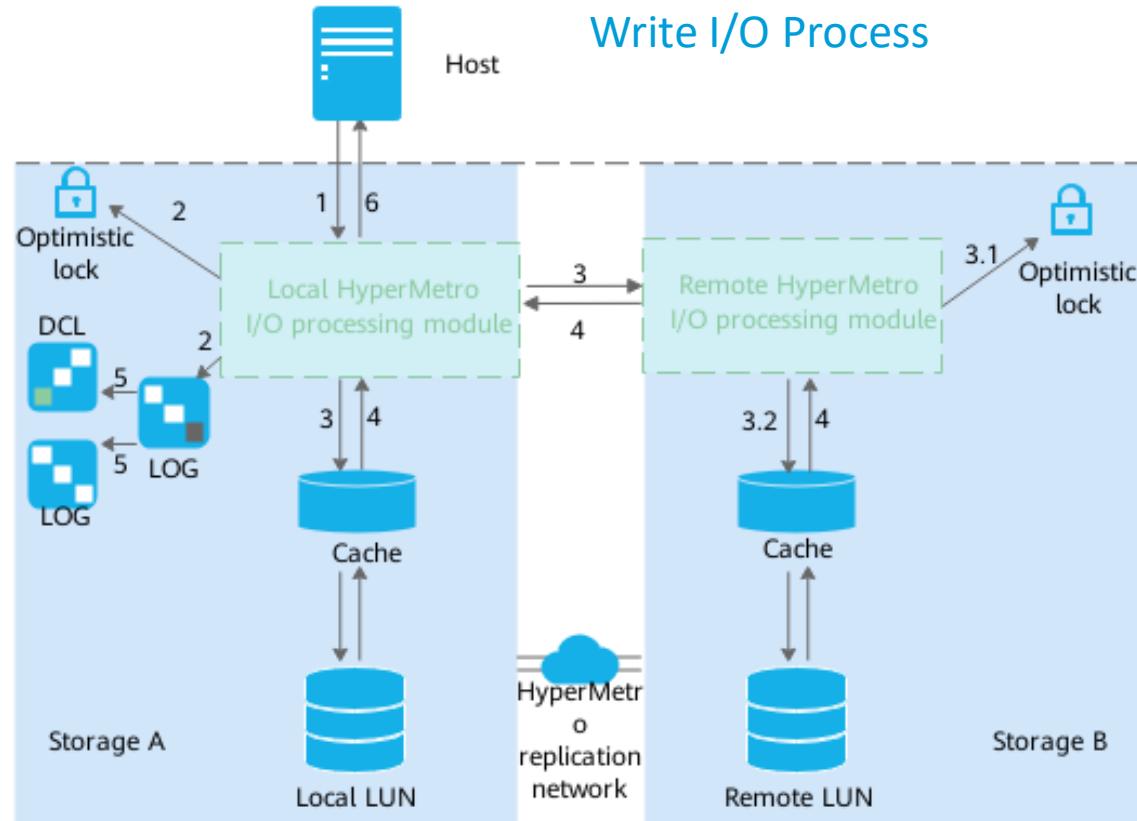
Characteristics

- Active-active, RPO = 0, RTO \approx 0 Gateway-free active-active solution. Single network between storage arrays, which is easy to deploy.
- Integrated HyperMetro for both SAN and NAS, shared arbitration

HyperMetro SAN and NAS Active-Active Design

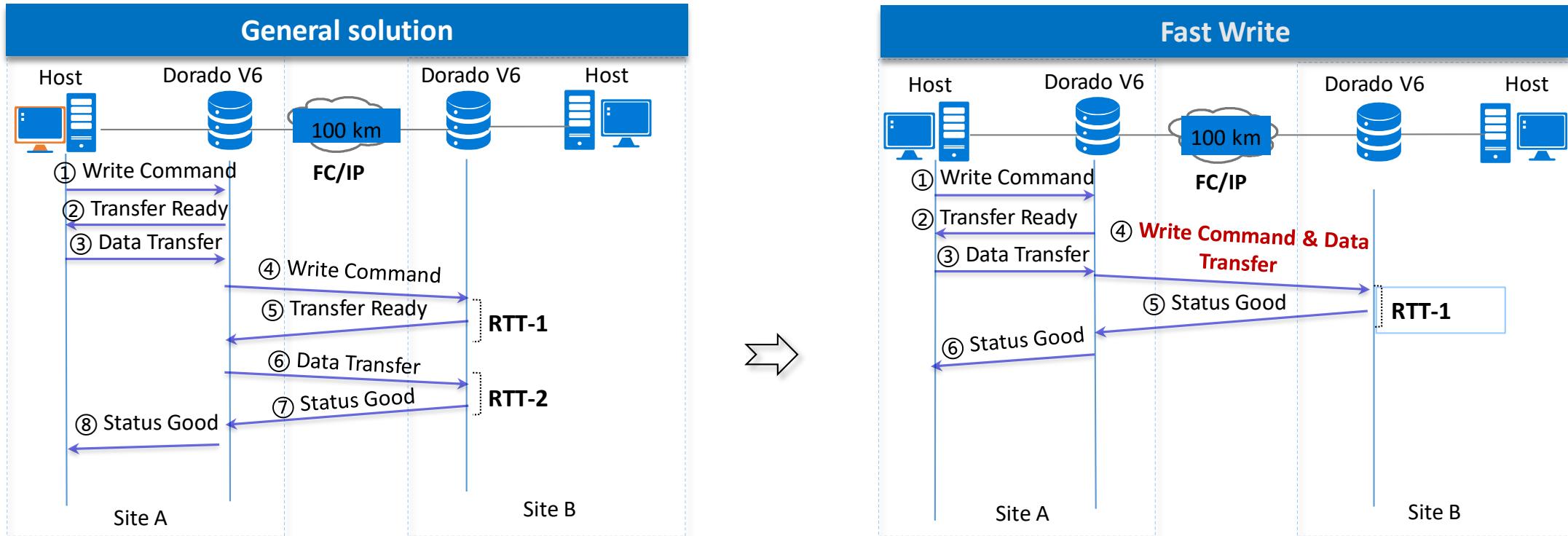
- **Architecture:** The symmetric active-active architecture across arrays supports read and write services of a single LUN or file system at both ends, and data is synchronized to the peer end in real time.
- **High reliability design:** The dual arbitration mechanism improves system reliability.
- **High-performance design:**
 - Protocol optimization for cross-site communication
 - FC: **NoF** FastWrite reduces cross-site data interactions by half.
 - IP: **RDMA** improves bandwidth and reduces latency.
 - Cross-site data synchronization: Only one synchronization
- **Easy management:**
 - Wizard-based configuration at a single site and automatic configuration synchronization
 - Simple networking, E2E all-IP configuration
 - HyperMetro can work with other Hyper and Smart series features. For details, see the feature list.

HyperMetro for SAN: I/O Processing Mechanism



HyperMetro uses **dual-write** and **data change log (DCL)** to synchronize data changes between the storage systems in two DCs, ensuring data consistency. The storage systems in both DCs provide services for hosts concurrently.

HyperMetro: Fast Write Accelerates Writes Across Sites



- General solution: Write I/Os experience two interactions at two sites: write command and data transfer.
- **Two RTTs** over 100 km transmission links.

- Fast Write: optimizes the protocol to combine write command and data transfer into one interaction, **reducing cross-site write I/O interactions by half**.
- **Only one RTT** over 100 km transmission links, improving service performance by **30%**.

HyperMetro for NAS: Active-Active Mode , RPO = 0, RTO \approx 0

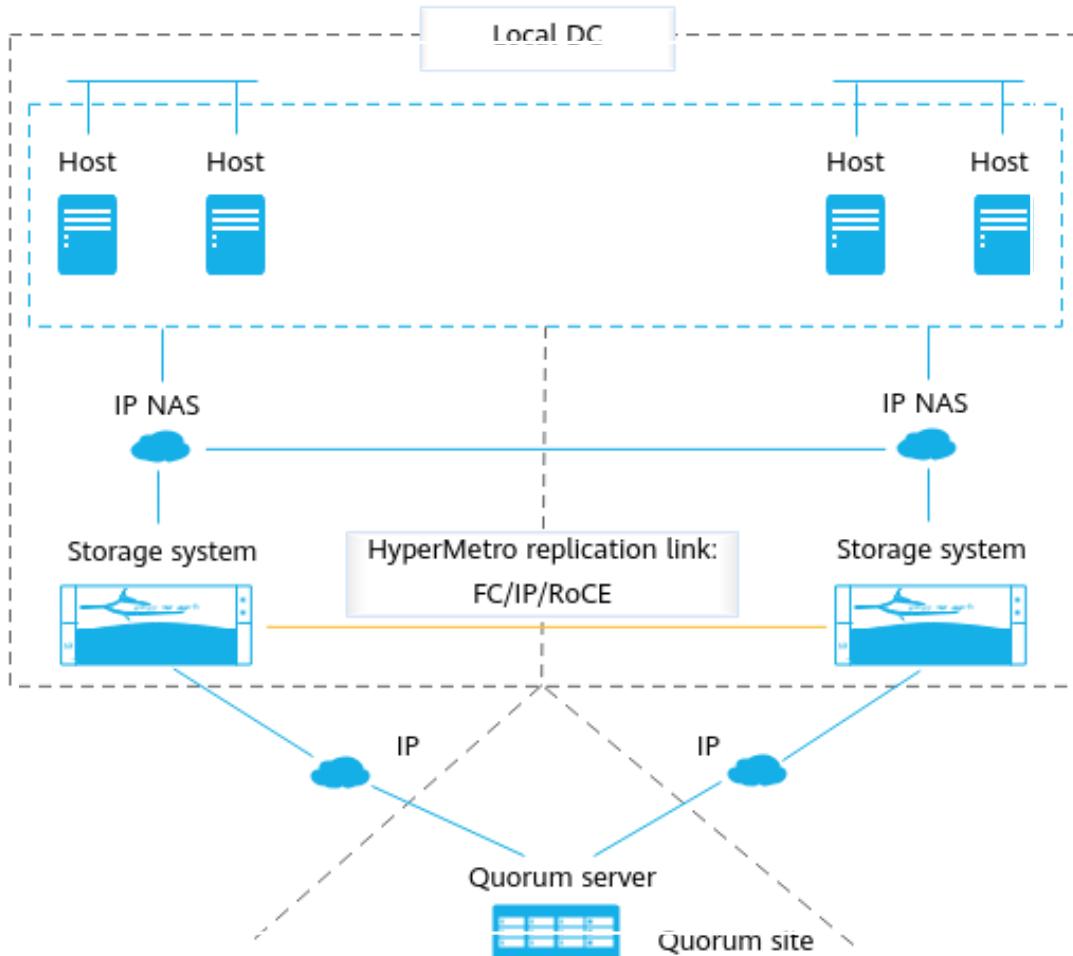


Figure 1 Single-DC deployment

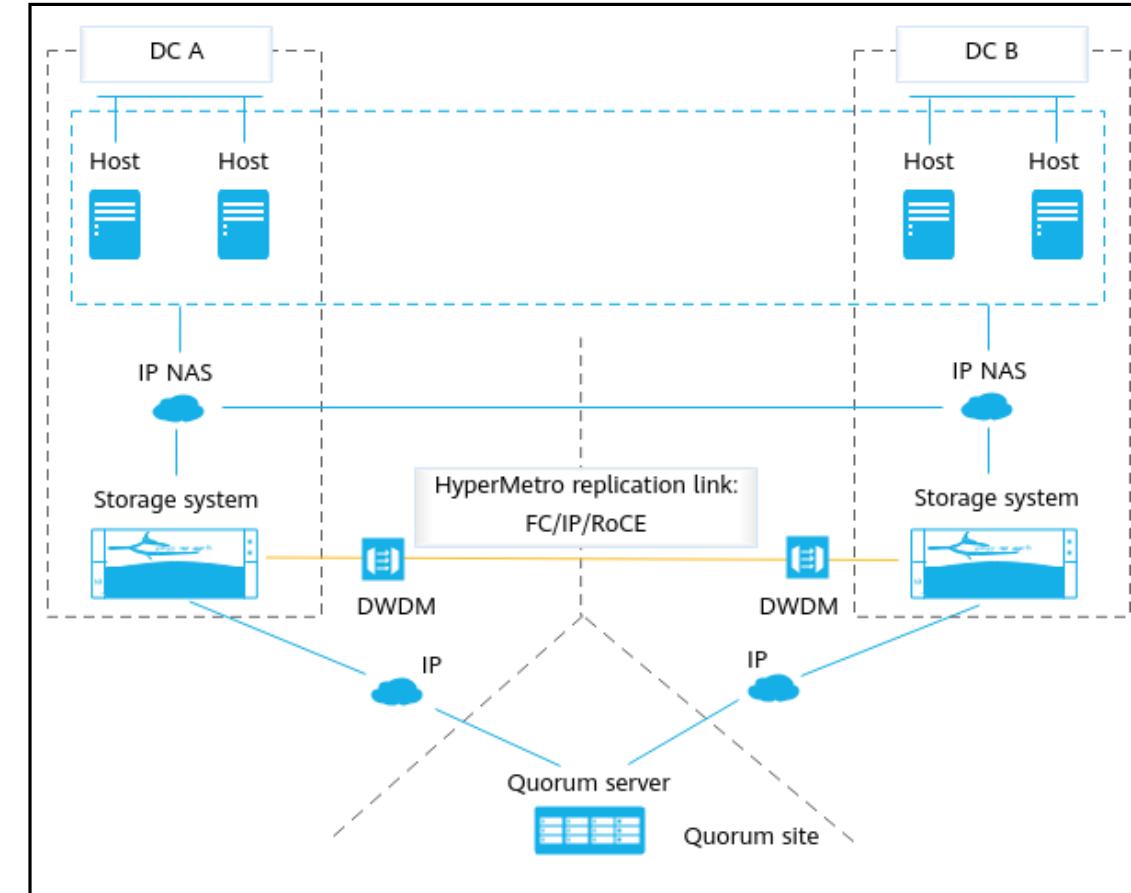


Figure 2 Cross-DC deployment

HyperMetro for NAS: Synchronous Mode , RPO = 0, RTO > 0

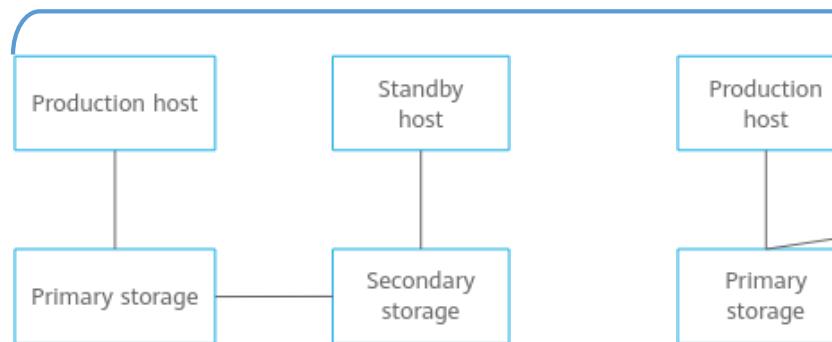
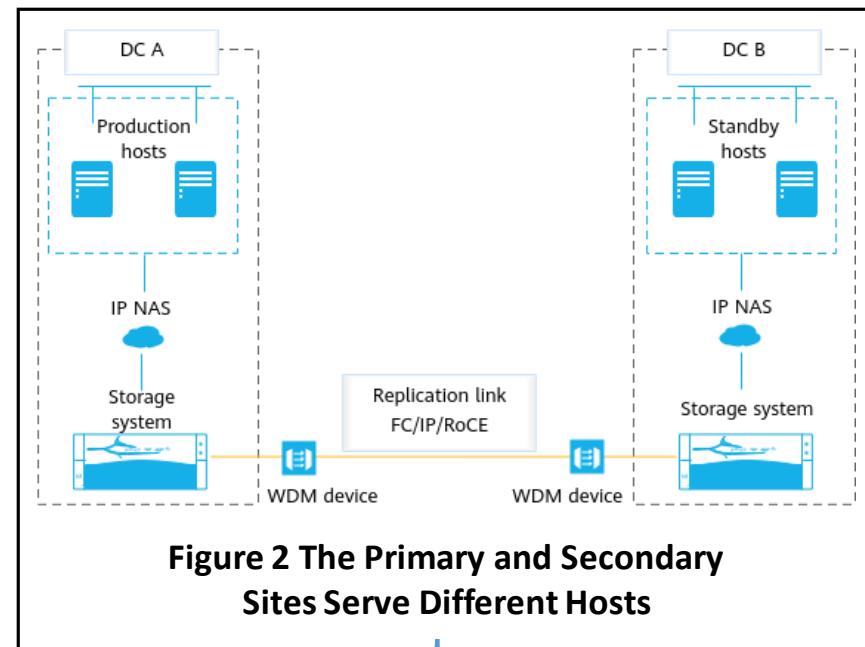
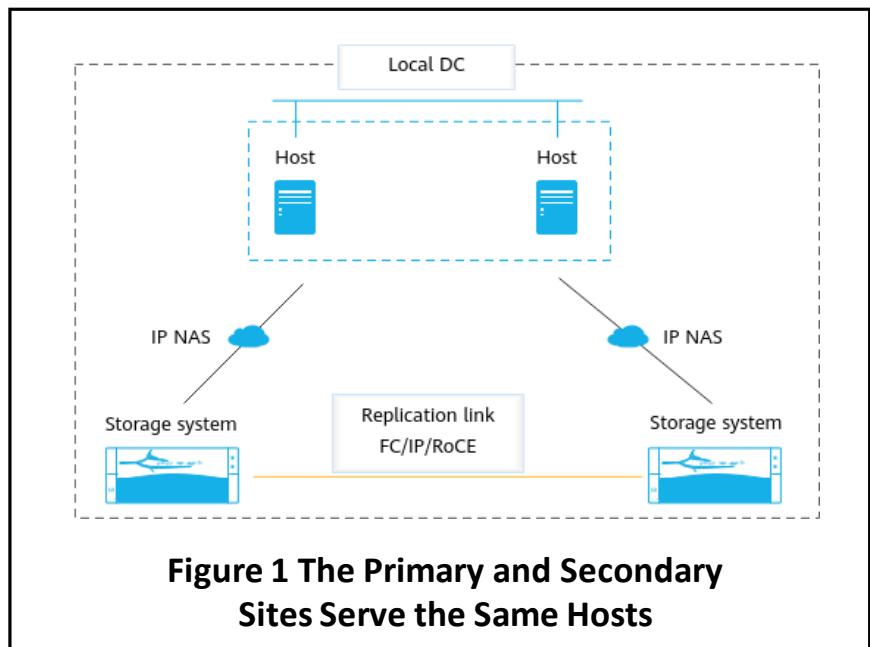


Figure 3 No domains

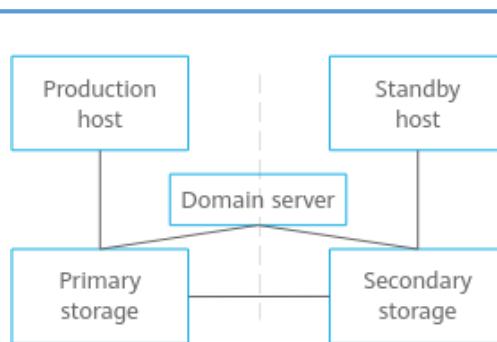


Figure 4 Same domain

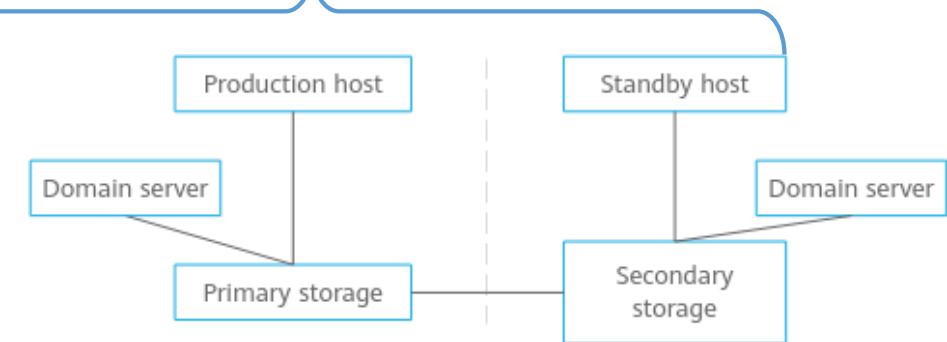
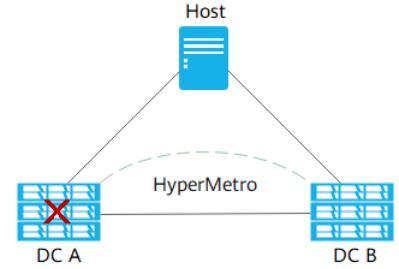
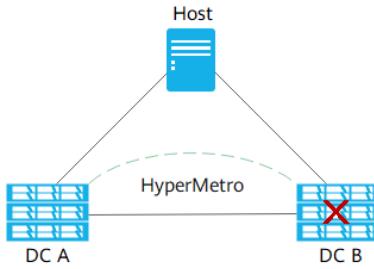
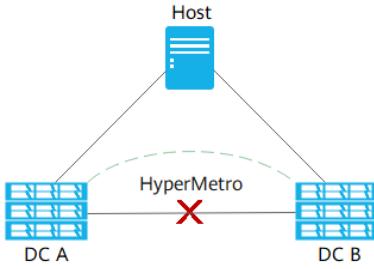


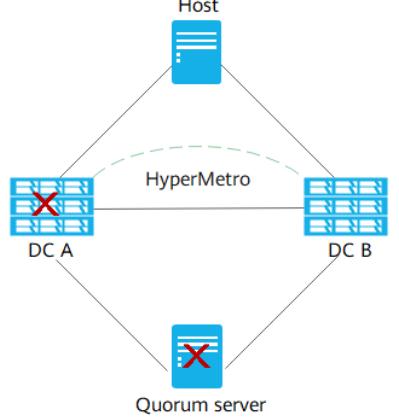
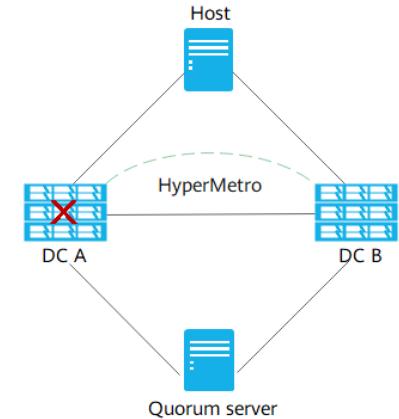
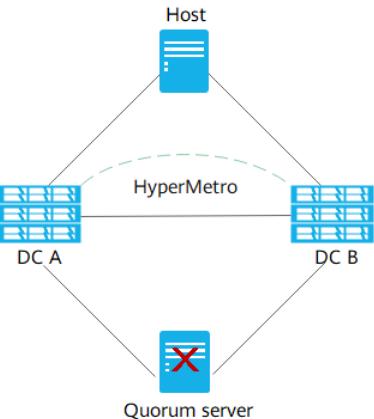
Figure 5 Different domains

HyperMetro: Arbitration Mechanism

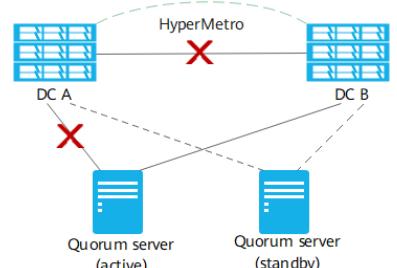
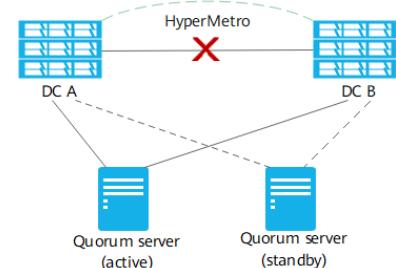
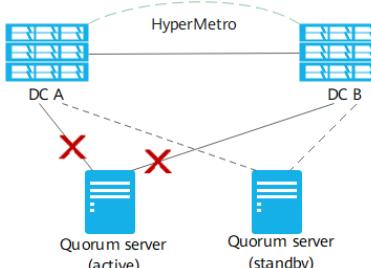
Static Priority Mode



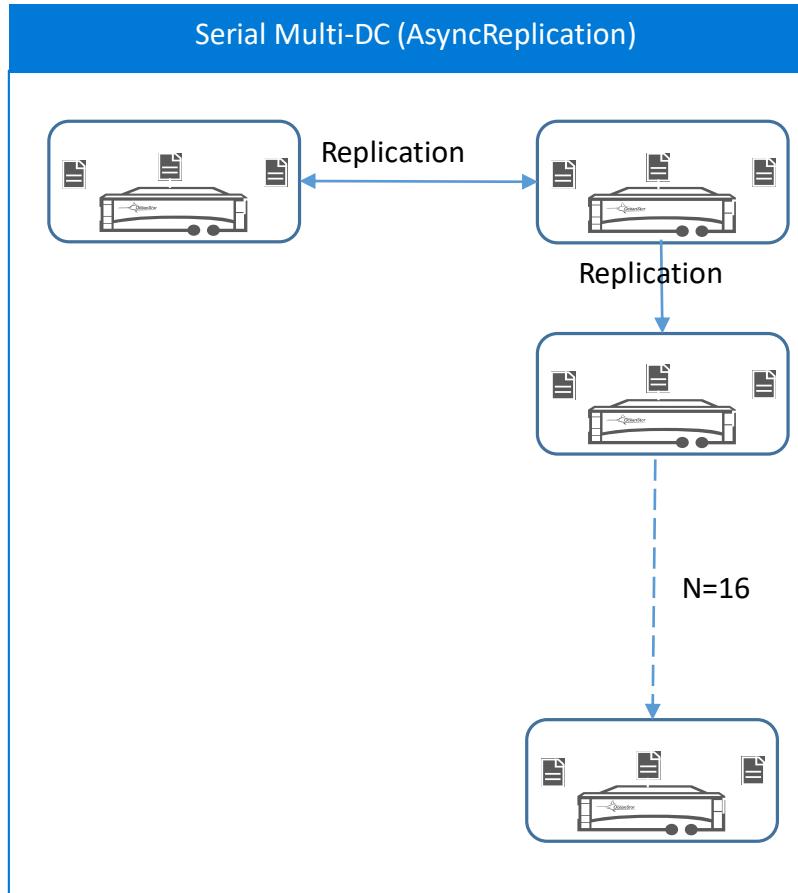
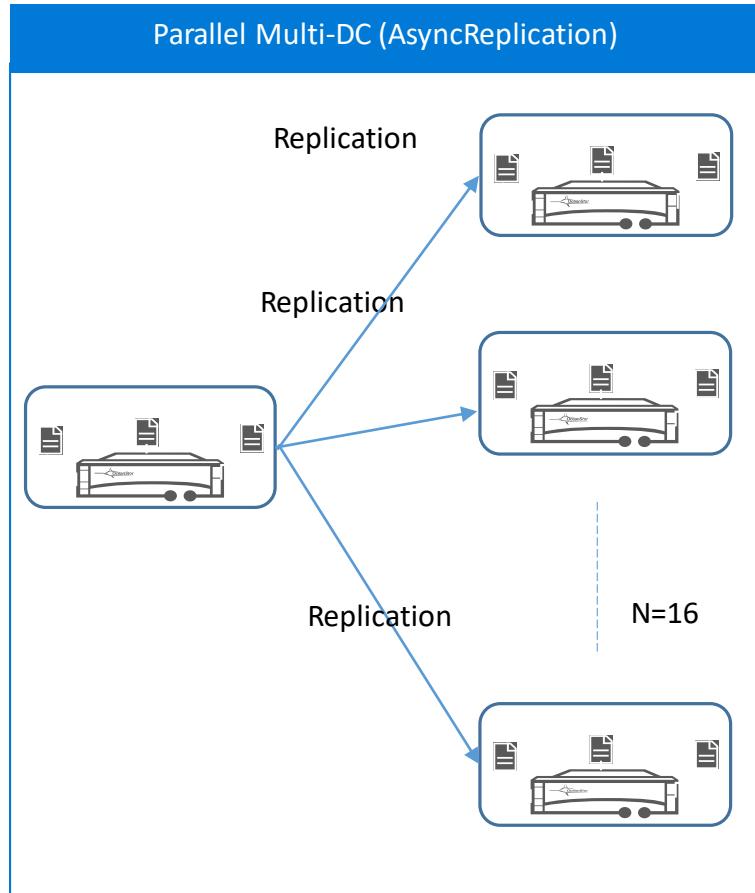
Single-Quorum-Server Mode



Dual-Quorum-Server Mode

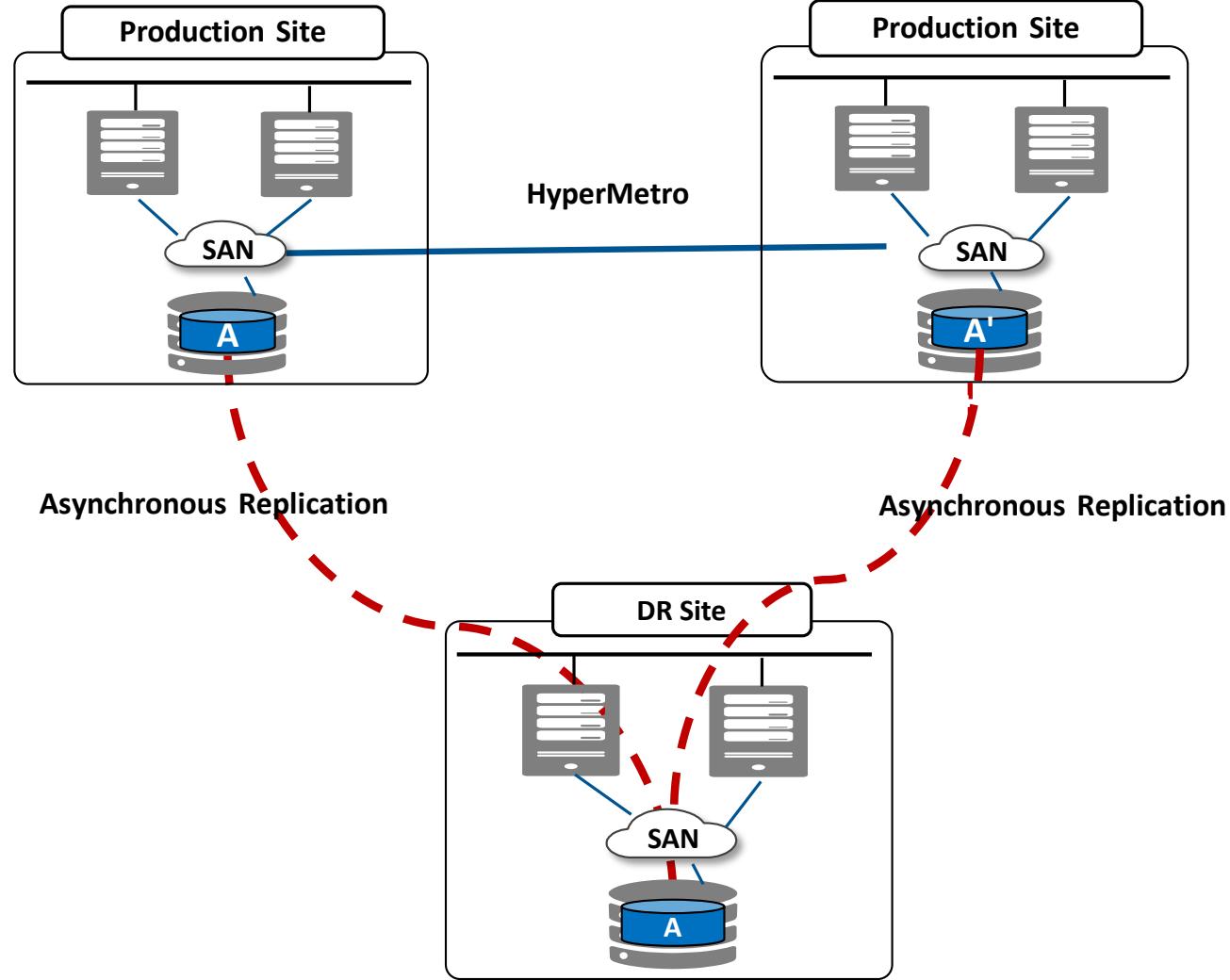


Multi-DC (NAS)

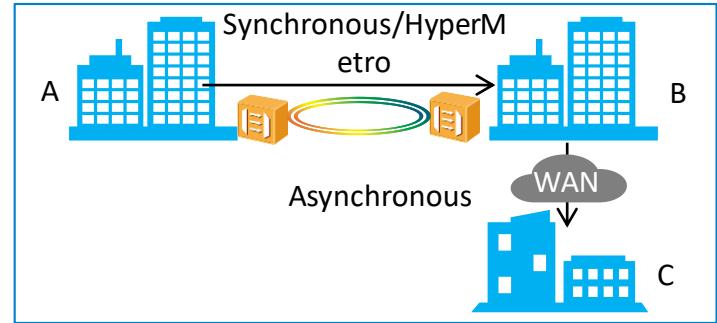


3DC for NAS		
Topology	combo	Support
Serial	Hypermetro + Async	Y
	Hypermetro + Sync	N
	Async + Async	Y
	Sync + Async	Y
Parallel	HyperMetro+Asyn c	Y
	HyperMetro+Sync	N
	Async + Async	Y
	Sync + Async	Y

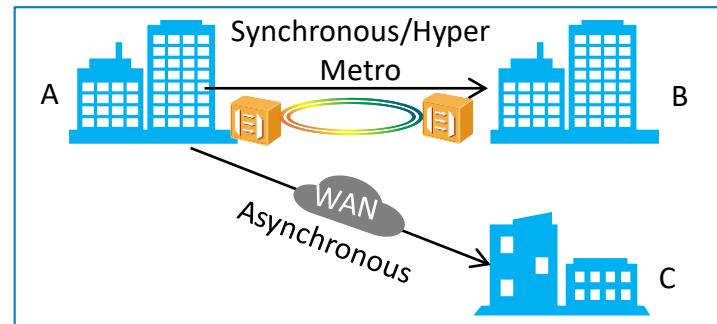
Multi-DC (SAN)



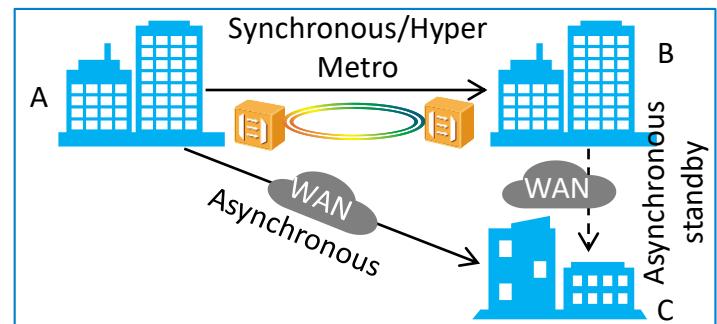
Serial



Parallel



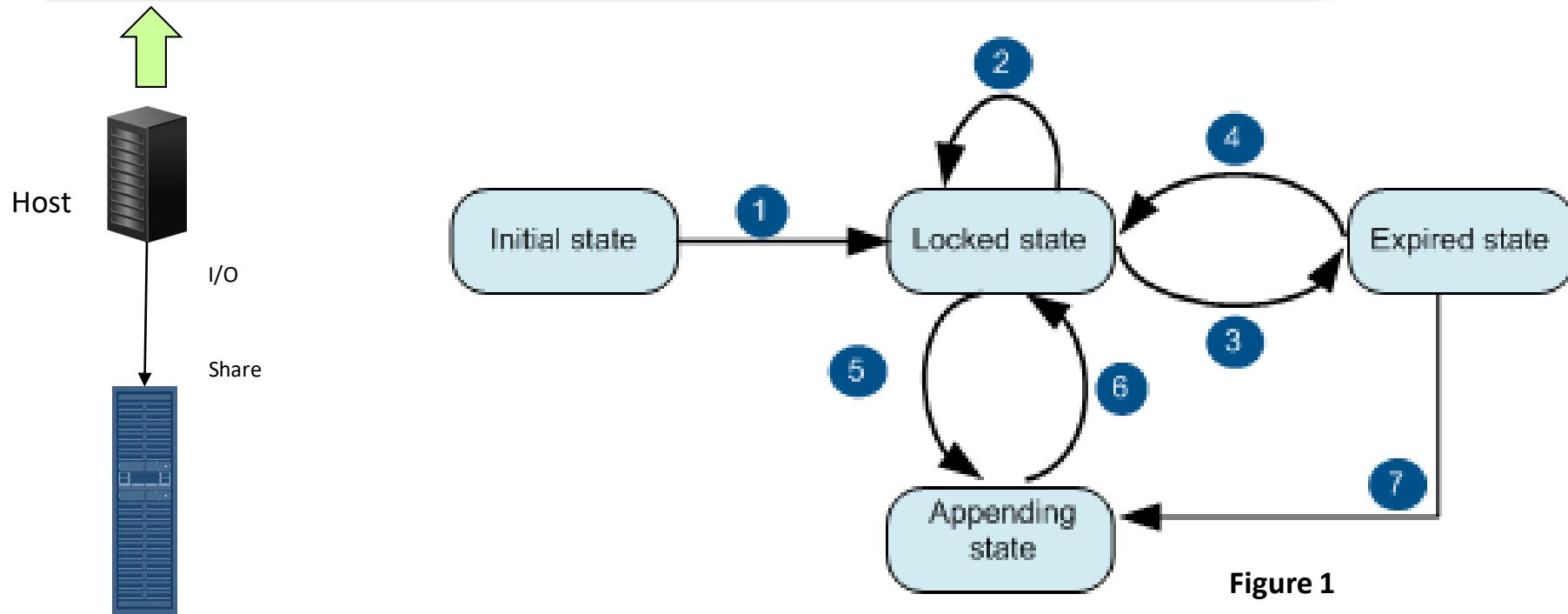
Star



HyperLock(NAS): The Write Once Read Many (WORM) feature

- Write once read many (WORM) is a technology that allows data to be read-only once being written. You cannot modify, delete, or rename a file within its protection period.

- **atime**: time when the protection period expires.
- **mode**: file status.
- **wormStatus**: WORM status, including initial, locked, appended, and expired.
- **legalhold**: litigation hold flag.



Quiz

1. (Single-choice) OceanStor Dorado HyperReplication for NAS only supports async replication, for sync replication, which feature should be used?
 - A. HyperClone
 - B. HyperLock
 - C. HyperMetro
 - D. HyperCDP
2. (Single-choice) Assume that a customer needs to generate a large number of data copies based on the scheduling policy for the purpose of recovering data when production data is deleted by mistake. In this case, which feature must be configured for the OceanStor Dorado?
 - A. HyperSnap
 - B. HyperCDP
 - C. HyperClone
 - D. HyperReplication

Contents

1. Overview
2. Hardware Architecture
3. Software Architecture
4. Smart Series Features
5. Hyper Series Features
- 6. Other Key Features**

Overview and Objectives

- This section describes the new software features of Huawei OceanStor Dorado.
- On completion of this section, you will be able to:
 - Describe the design of key technologies such as cloud backup, antivirus, log audit etc.

CloudBackup(NAS)

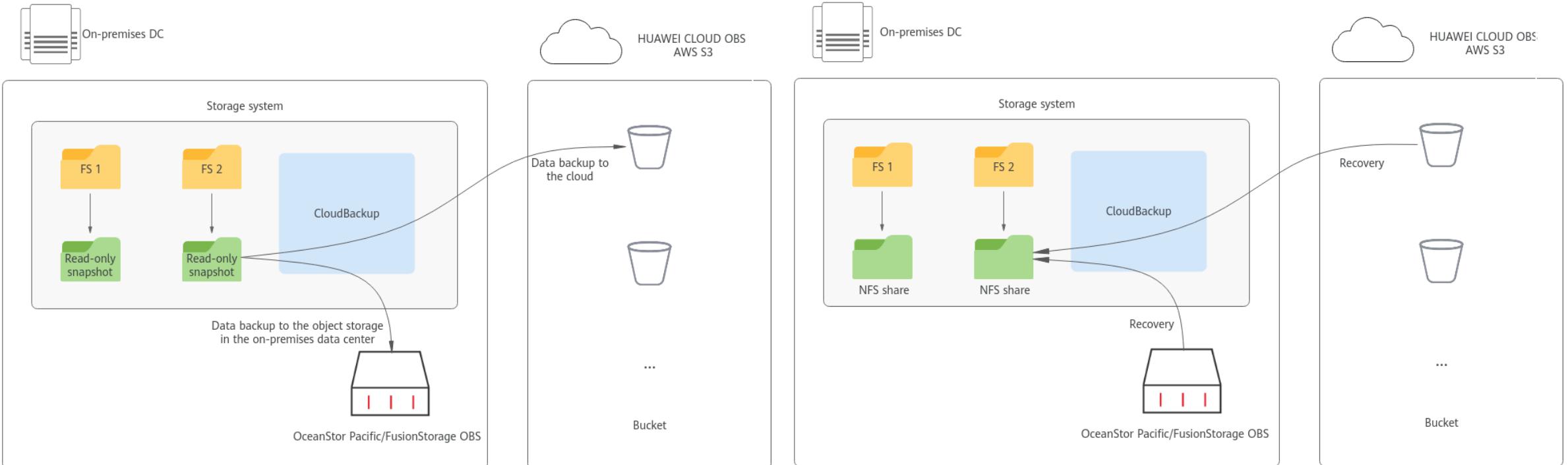
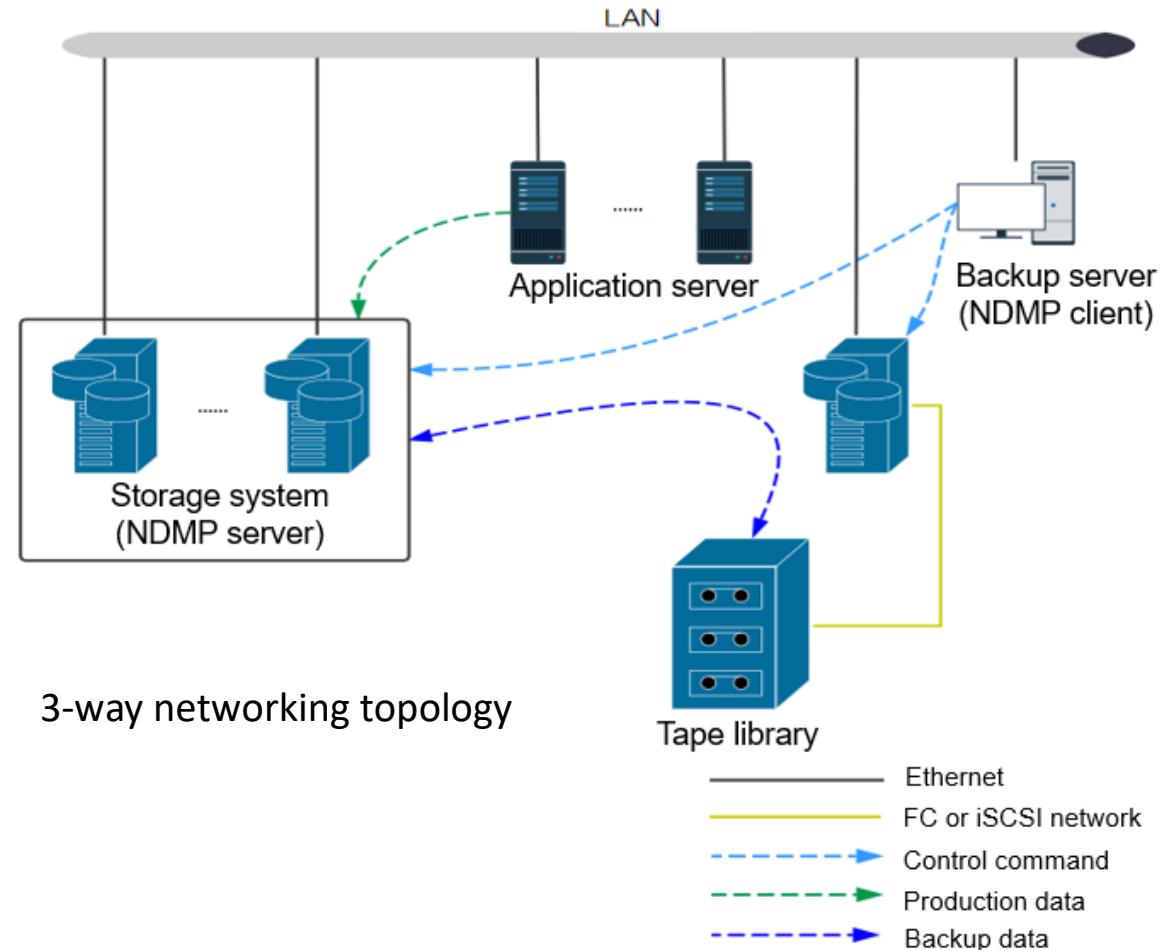
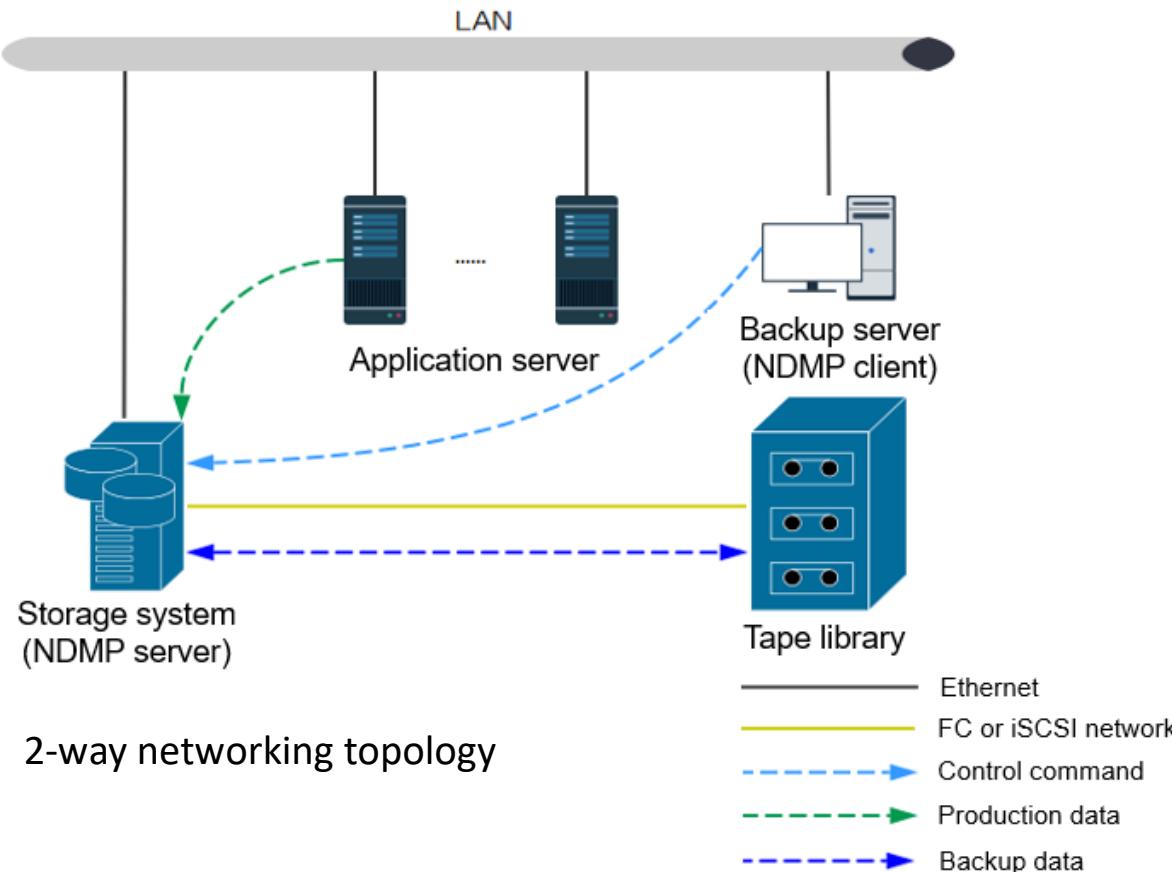


Figure1: Data backup process with CloudBackup

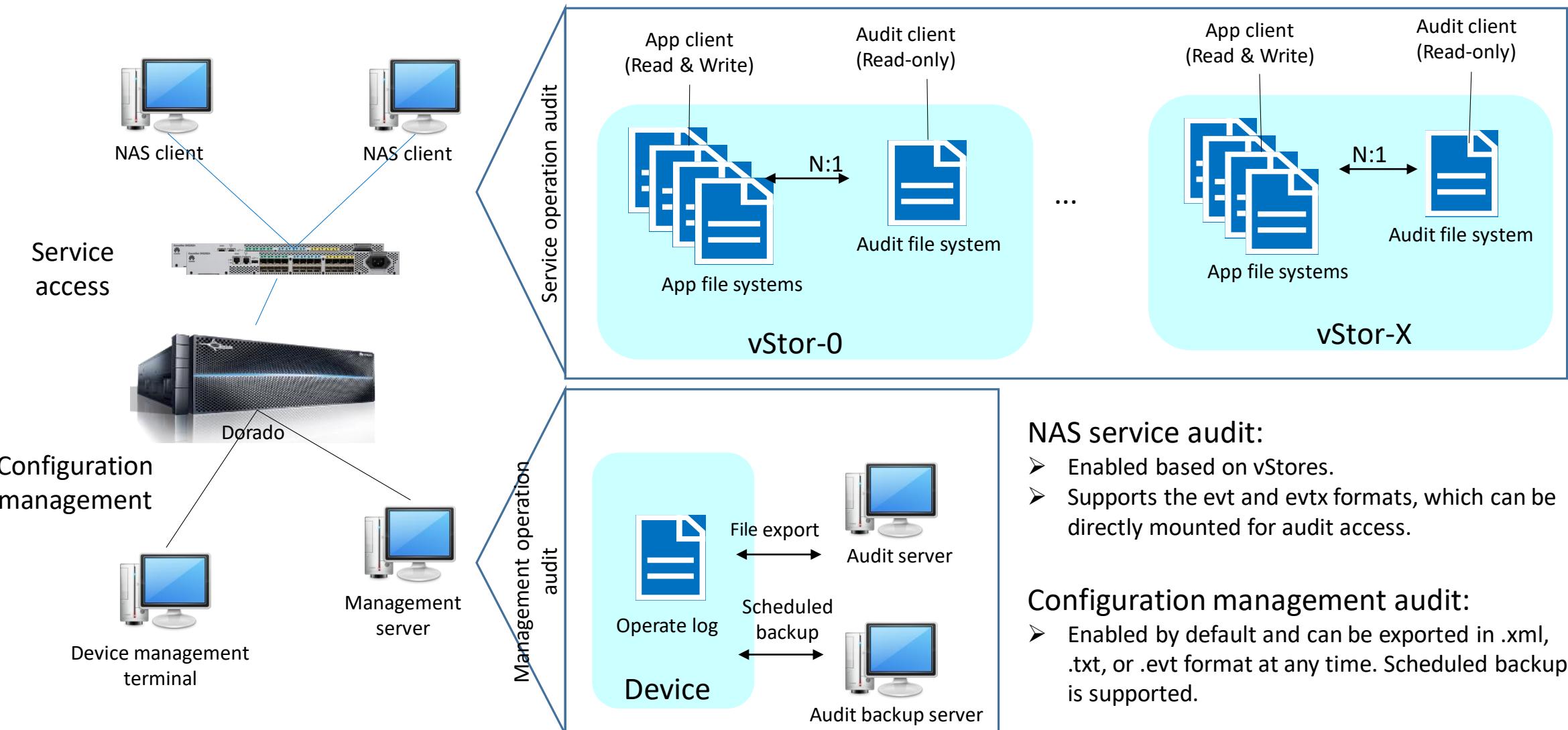
Figure2: Data Recovery process with CloudBackup

CloudBackup is a data protection technique deployed in storage system containers. It backs up **file system** data from a storage system to the object storage either in an on-premises data center or on the cloud, without the need for extra backup servers. In the event of data loss or corruption in the file system, CloudBackup can use the backup copies to restore the data to the state at the specified point in time.

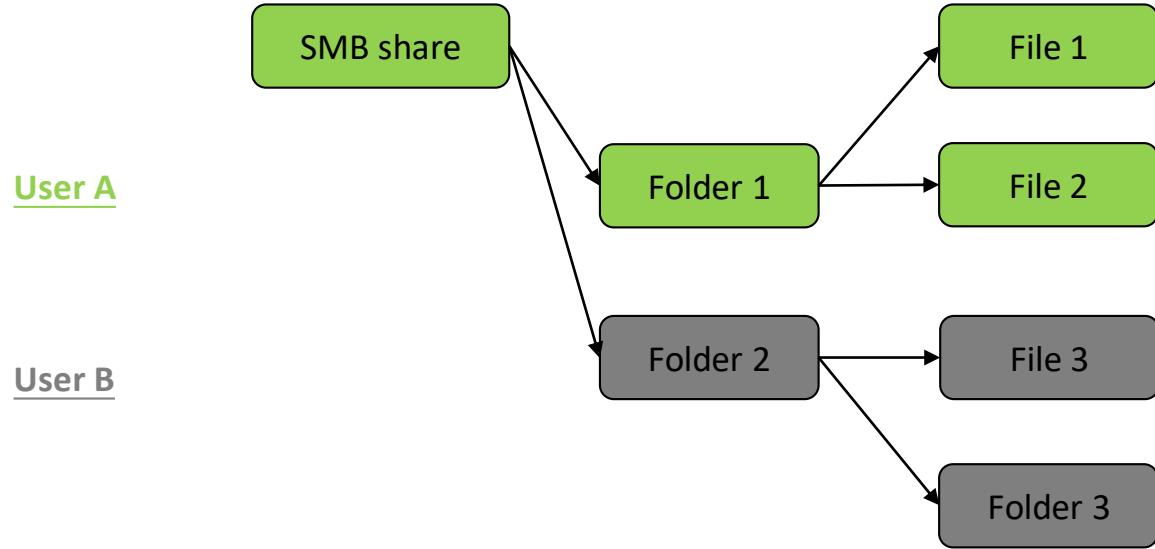
NAS Protocol Layer: NDMP Application Networking



Audit Log



SMB ABE



Through an ACL configuration:

- User A is authorized to read the files and folders in green, but user B has no permission on these files and folders.
- User B is authorized to read the files and folders in gray, but user A has no permission on these files and folders.

With the ABE function enabled:

- User A can only view or operate the files and folders in green.
- User B can only view or operate the files and folders in gray.

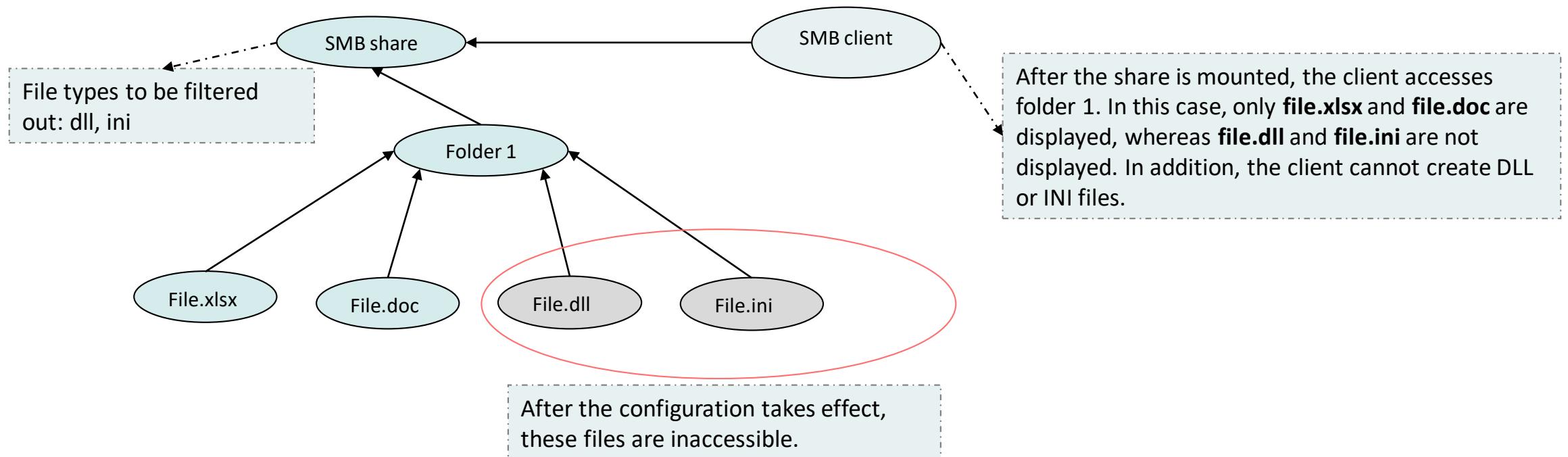
Access-based enumeration (ABE) displays only the files and folders that a user has permissions to access.

This function applies to file sharing scenarios. When a user attempts to access a share over the SMB protocol, the SMB server hides the files and folders that the user is not authorized to access. This function improves the enumeration experience, privacy, and security.

SMB File Blocking

This feature manages files based on **file types**. It allows administrators to filter out the files that are not expected to be viewed or accessed.

File filtering is configured for an SMB share and applies to all clients that access the share. After the configuration takes effect, access to the specified types of files will be denied, and these files will be filtered out in later queries.



NAS Antivirus

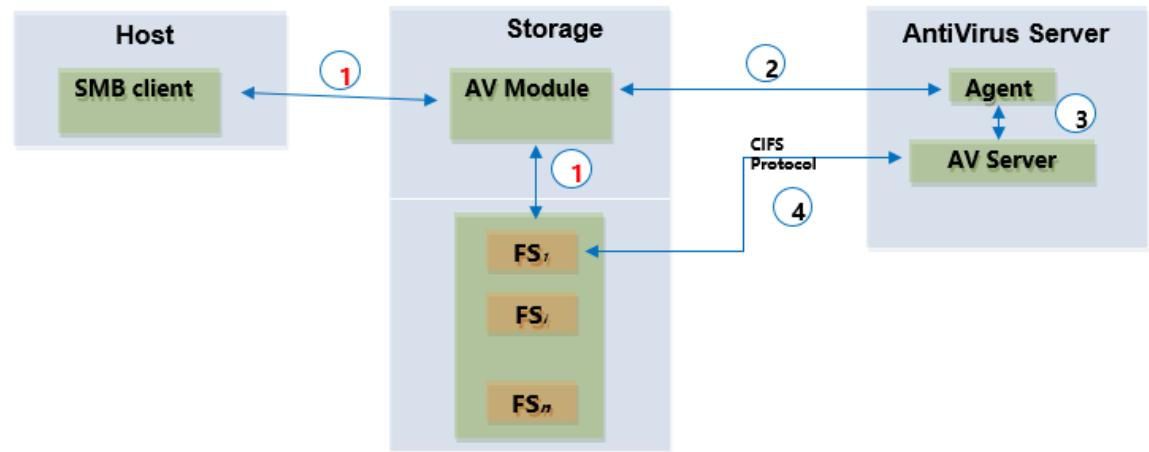
Overview

Antivirus scanning is an important measure to protect storage security. It works with a third-party antivirus engine to thoroughly protect data security.

Functions

- The main functions of the NAS antivirus feature are as follows:
 - On-access scanning:** When a file is accessed (open/close), virus scanning is performed in real time.
 - On-demand scanning:** After a scan policy and a scan server are configured, scanning is performed based on the specified policy and cycle.
 - Antivirus server management:** An antivirus server can be added or deleted.
- Antivirus logs can be exported.
- The antivirus feature does not isolate files. Instead, the isolation function provided by the antivirus software is used.

Basic Principles



On the antivirus server, the Agent module can use Share Open and ICAP to enable the AV Server to scan for viruses.

Section Summary

- This chapter describes the hardware architecture, software architecture, Value added features and principles of the OceanStor Dorado all-flash storage system. It also describes the requirements and advantages of the OceanStor Dorado all-flash storage system.

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 New-Gen OceanStor Hybrid Flash Storage Key Technologies



Foreword

- The New-Gen OceanStor Hybrid Flash Storage is newly developed hybrid flash product, it inherits the architecture and features of OceanStor Dorado, provides cost-effective and high performance solution for customer.
- The Chapter will describe the architecture of New-Gen OceanStor Hybrid Flash Storage, especially SmartAcceleration feature, also it will describe the main sales target of new hybrid flash storage.

Contents

- 1. Product Overview**
2. Hardware & Software Architecture
3. Target Sales Scenario

Overview and Objectives

- This section describes highlights of new-gen OceanStor Hybrid Flash Storage and differences between OceanStor V5 and new storage.
- On completion of this section, you will be able to:
 - Describe the highlights of new hybrid flash storage
 - Describe the main improvement of new hybrid flash storage.

Future-Oriented Design: Continuous Evolution and Cost Effectiveness



OceanStor
5310/5510/5610/6810/18510/18810



Continuous Evolution, Oriented to Multi-Workloads

- Supports multi-workloads, such as blocks, files, virtualization, and containers
- Cloud backup and private cloud interconnection to support cloud-based evolution
- Anti-ransomware and secure snapshots to protect against ransom issues



Full Upgrade for Flash-Like Experience

- SmartAcceleration in all scenarios and convergence of cache and tier
- SmartMatrix A-A fully load balancing architecture maximizes the performance



Reduced TCO

- No SAS HDD required reduces construction costs; supports reuse of cross-gen devices
- Intelligent and automatic O&M greatly reduces O&M costs

Overview of Product Portfolio

Mid-Range

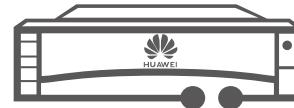
	Middle Level			High Level		
Type	OceanStor 5310	OceanStor 5510	OceanStor 5610	OceanStor 6810	OceanStor 18510	OceanStor 18810
Height / Controllers per Engine	2U/2C	2U/2C	2U/2C	4U/4C	4U/4C	4U/4C
Controller Expansion	2-16	2-16	2-16	2-16	2-32	2-32
Maximum Disks	1200	1600	2000	3200	6400	9600
Maximum Capacity per Engine (TiB)	512 (64G/C) 1024 (128G/C)	2048 (192G/C) 2048 (256G/C)	4096 (384G/C) 4096 (512G/C)	4096 (256G/C) 8192 (512G/C) 16384 (1024G/C)	4096 (256G/C) 8192 (512G/C) 16384 (1024G/C)	8192 (512G/C) 16384 (1024G/C)
Front-end Ports	8/16/32G FC/NOF(NVMe over Fabric); 10/25/40/100G Ethernet, 25/100G NOF (NVMe over RoCE), 10GE/1GE electrical interface					
Back-end Ports	SAS 3.0/100G Ethernet					

Note: For detailed specifications, please refer to the product specification list.

New Architecture in New-Gen OceanStor Hybrid Flash Storage



**New-Gen OceanStor Hybrid
Flash Storage**



OceanStor V5

Category		New-Gen OceanStor Hybrid Flash Storage	OceanStor V5 Kunpeng
Architecture	Architecture	LUN: Fully A-A controllers File system: Distributed file system	LUN: ALUA A-A File system: A-P architecture
	Reliability	High-end models tolerate 7/8 controller failures	High-end models tolerate 3/4 controller failures
	Upgrade	NDU supported. No restart required in 95% of scenarios	Controller restart required
Hardware	Host protocol	Supports FC-NVMe and NVMe over RoCE (NoF+)	NoF is not supported.
	Drives	SSD +NL_SAS, SSD is mandatory	SSD,SAS and NLSAS
Functions	Snapshot	ROW (Redirection on Write)-based lossless mode: Supports high-density and cascaded snapshots.	COW(Copy on Write) mode: snapshot performance deteriorates significantly. High-density snapshots and cascaded snapshots are not supported.
	Metro clustering	Industry's only A-A solution for SAN and NAS	A-A solution for SAN A-P solution for NAS
	RAID	Supports RAID 5, RAID 6, RAID-TP, and RAID zoom (dynamic RAID reconstruction). Tolerance for 3 disk failures.	Supports RAID 0, 1, 3, 5, and 6. Tolerance for up to 2 disk failures
	SSD Acceleration	SmartAcceleration , including SSD cache and SSD Tier	Optional. SmartCache and Smart Tier are separated licenses.
	Deduplication & Compression	HDD pool supports compression SSD pool supports both with less performance impact	Supported but not recommended as performance downgrade.

Quiz

1. (True or False) Inside of SmartAcceleration in new hybrid storage, there are two separated software licenses: SmartTier and SmartCache.
A: T
B: F

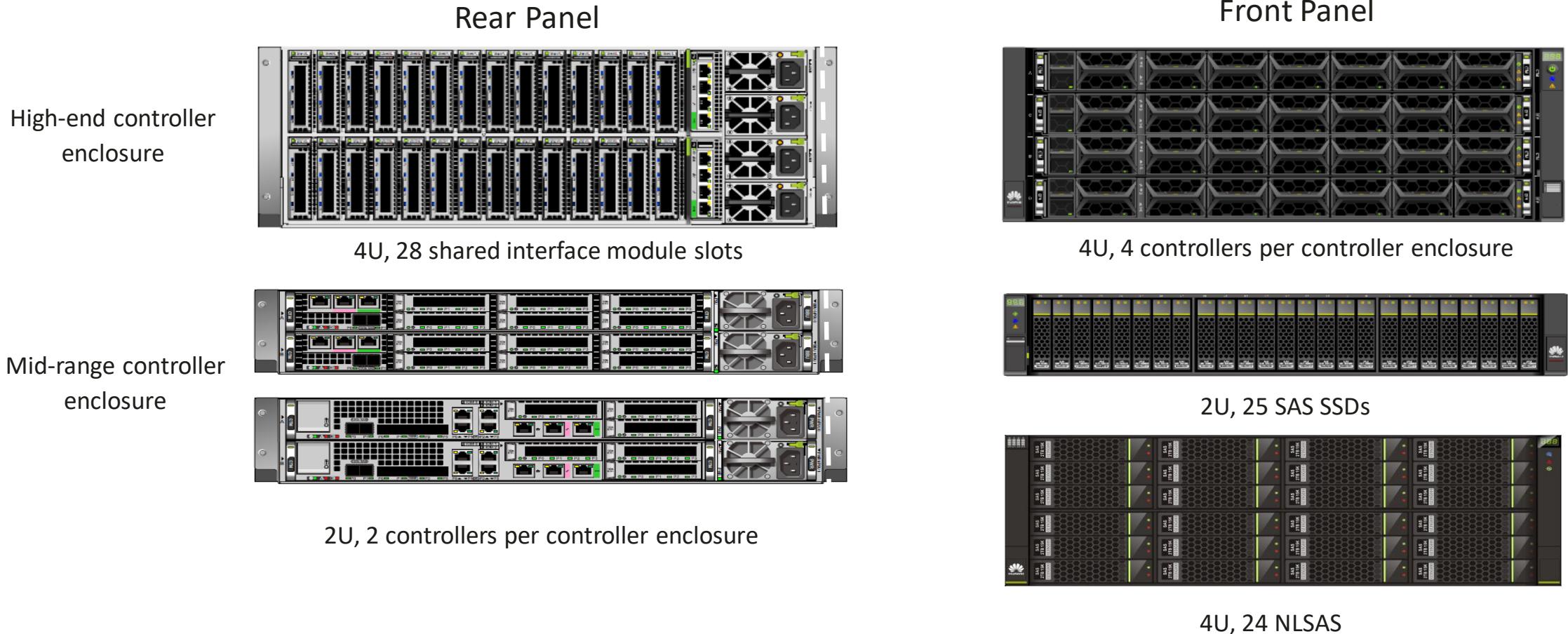
Contents

1. Product Overview
- 2. Hardware & Software Architecture**
3. Target Sales Scenario

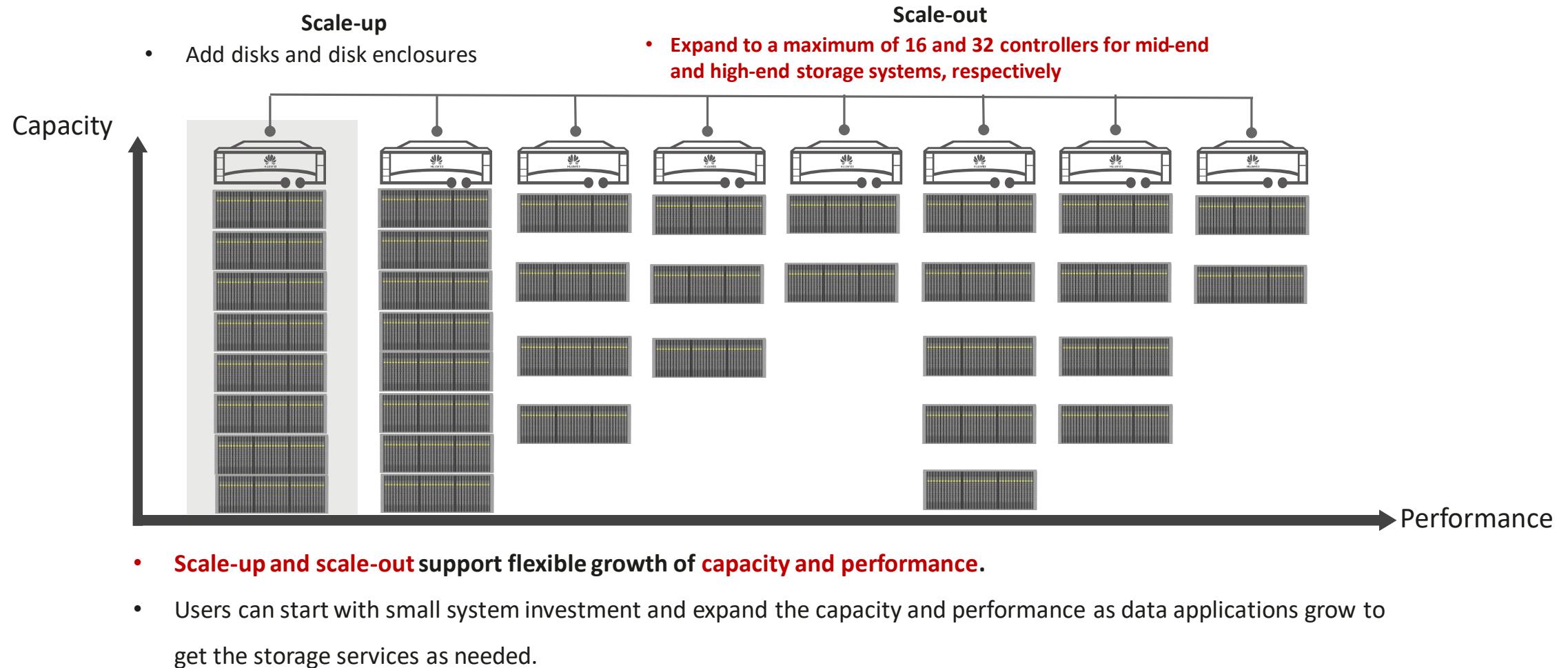
Overview and Objectives

- This section describes hardware and software architecture of new-gen OceanStor Hybrid Flash Storage.
- On completion of this section, you will be able to:
 - Describe hardware and software features
 - Describe highlights of SmartAcceleration feature.

New Hybrid Flash Storage Hardware Overview

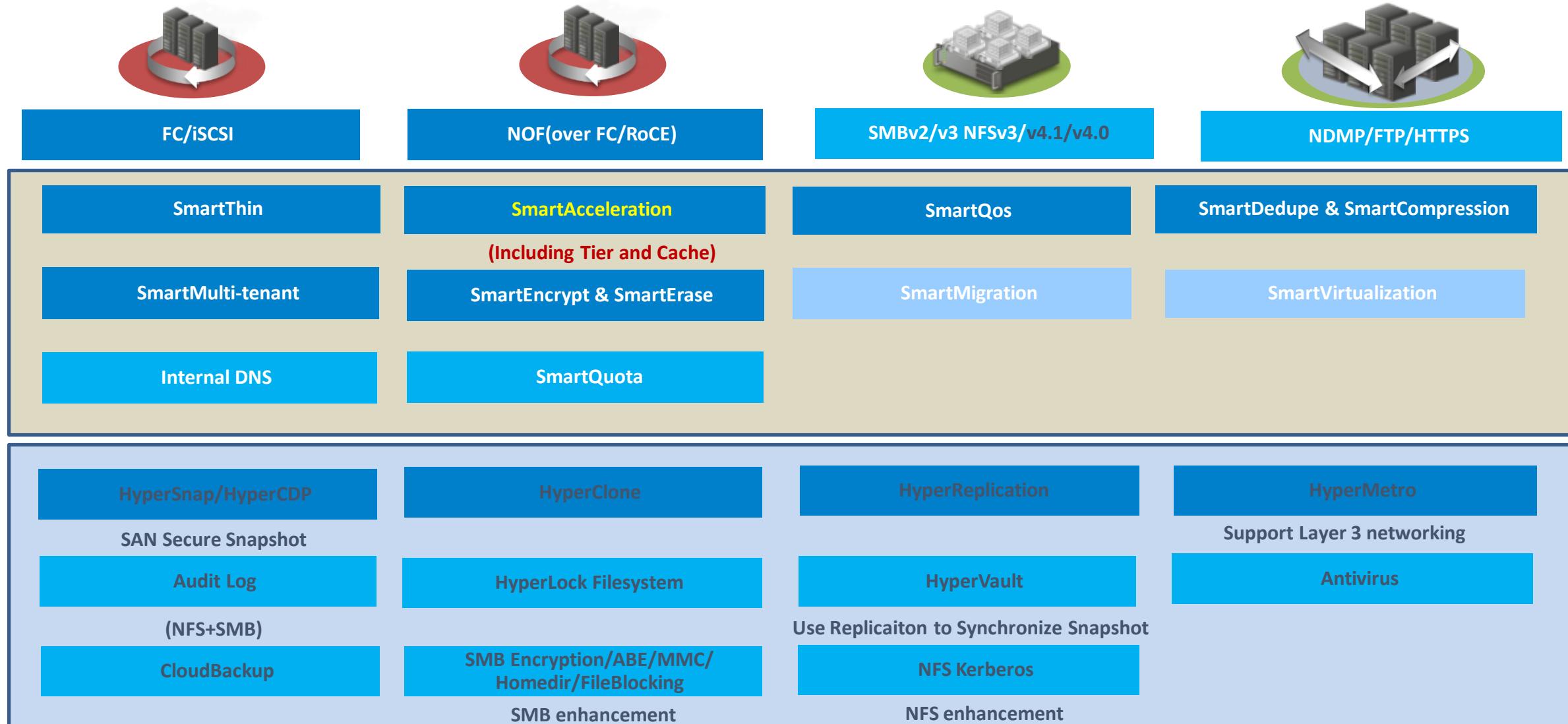


Flexible Expansion for Increasing Requirement



Key feature of OceanStor New Hybrid Flash Storage

SAN only NAS only SAN&NAS



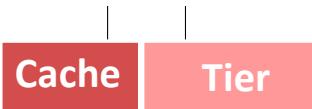
SmartAcceleration: Innovative Data Layout Algorithm Accelerates All Hot Data and Delivers Flash-Like Experience

SmartAcceleration



ROW-based large block sequential writes

Global data write optimization



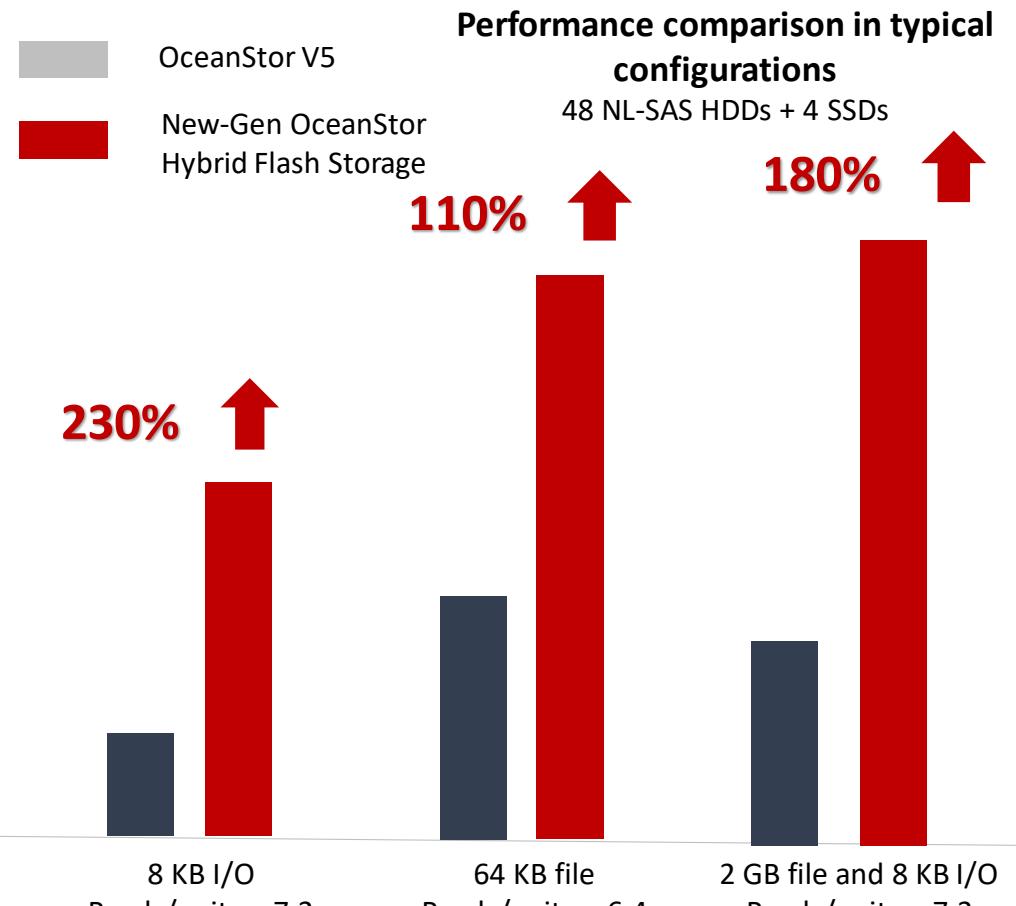
Flexible convergence of heterogeneous media

Optimal cost-performance balance



Global cold/hot data perception

Multidimensional feature learning of neural networks

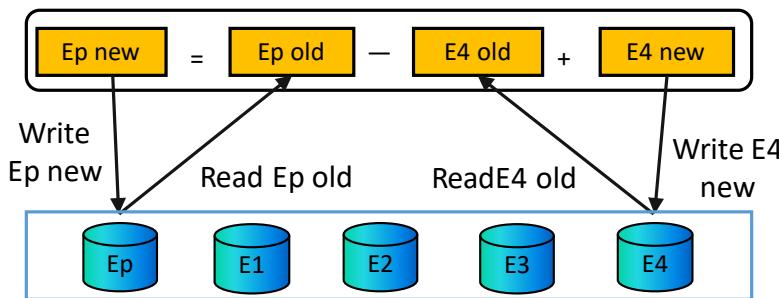


*Test results from Huawei Lab

SmartAcceleration: Breaks the Random Performance Bottleneck of Traditional HDDs Based on ROW Large-Block Sequential Writes

Traditional mode (CoW)

RAID 5, the IO process is :

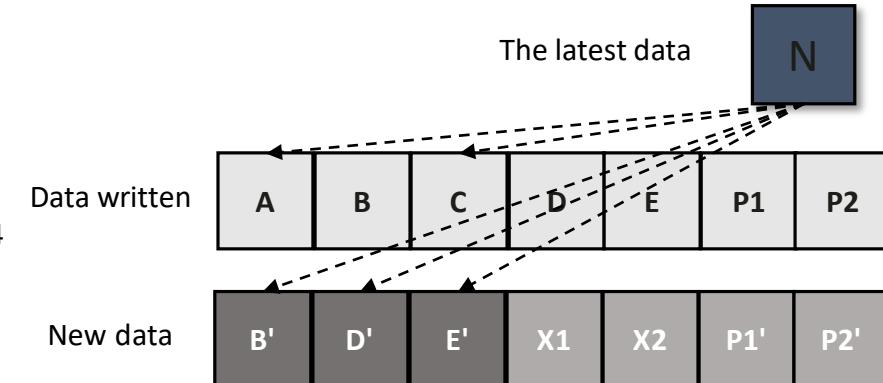


E_p: Parity data

E₄: Data to be changed

One Write operation required 2 extra Read and one extra Write. So the penalty is 4 for RAID 5 in CoW mode.

OceanStor New Hybrid Flash Storage(ROW):



BDE is the data to be changed

- Small I/Os are aggregated into large I/Os for sequential writes, eliminating HDD bottlenecks.
- New data makes a RAID group alone and no write penalty is generated.

Traditional mode				OceanStor mode			
Configuration	Reads	Writes	Penalty	Configuration	Reads	Writes	I/Os
RAID 5	2	2	4	RAID 5	0	0	1
RAID 6	3	3	6	RAID 6	0	0	1
RAID-TP	4	4	8	RAID-TP	0	0	1

RAID 5 RAID 6 RAID-TP



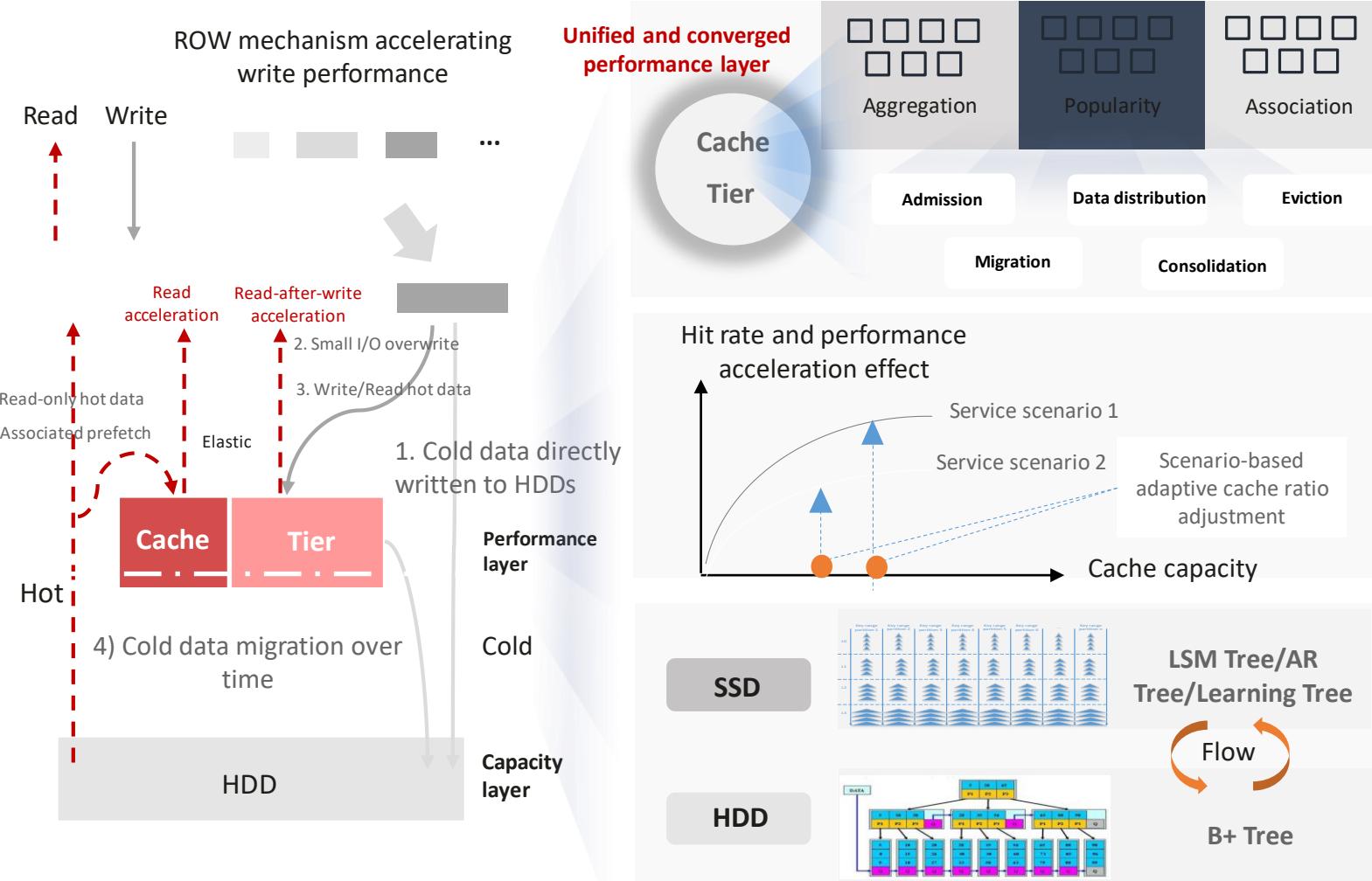
OceanStor New-Gen Hybrid Flash storage

RAID 5



RAID 6 RAID-TP

SmartAcceleration: Elastic Converged Performance Layer with All-Scenario Adaptive Tier+Cache



1 Global cold and hot data sensing and data collaboration algorithms, breaking the boundaries of caches and tiers, providing optimal data layout and simplified configuration

- Unified information collection points for the cache and tier, avoiding repeated collection
- Globally unified algorithms for the cache and tier, avoiding repeated memory computing overhead
- Cache and tier's hot data sensed at the semantic layer, providing higher accuracy

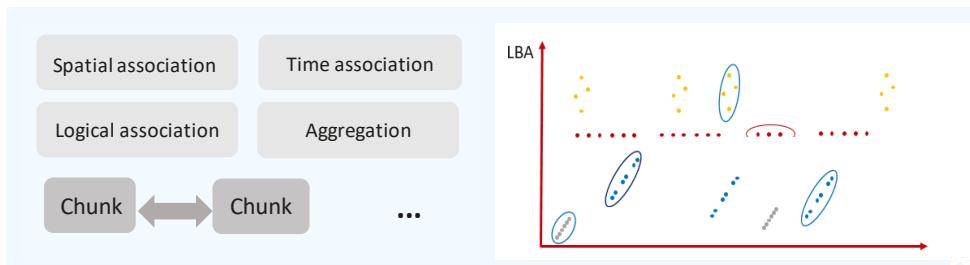
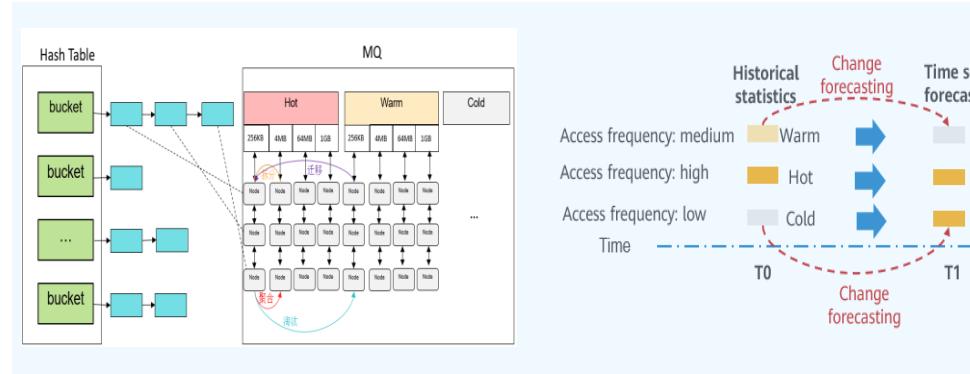
2 Scenario-based adaptive elastic cache adjustment, achieving optimal balance between performance and capacity

- Flexible and dynamic adjustment between the cache and tier, going for optimal cost-effectiveness
- No need to configure caches and tiers (physical) separately, simplifying configuration

3 Performance layer and capacity layer using the multiple index technologies to adapt to different data layouts, achieving optimal efficiency in various mixed models

- Performance layer using PelagoDB, small- and medium-capacity, fast data update
- Capacity layer using KVDB, medium- and large-capacity, balanced performance and capacity.

SmartAcceleration: All-Scenario Global Cold & Hot Data Sensing and Data Collaboration Algorithms Based on Self-learning



Large I/O model + Small I/O model + Sequence and randomness of I/Os
Multi-granularity learning statistical structure

Historical statistics + Time series forecasting algorithm of machine learning
Cold and hot data sensing algorithm with multi-dimensional feature convergence

Sequence + Aggregation + Spatial and time association
Data association prefetch algorithm with multi-dimensional feature convergence

Workload + System configuration + Adaptive algorithm resource adjustment
Global data flow and collaboration algorithm

Adaptable to flexible service models

Adaptable to variable cold and hot changes

Adaptable to various data associations

Optimal data flow and placement

Quiz

1. (Multiple Choice) Which of following description of SmartAcceleration are True?
 - A: SmartAcceleration can reduce the write penalty of small and random IO.
 - B: SmartAcceleration can reduce the amount of data to be written.
 - C: In SmartAcceleration, the cache and tier are converged and adjustable.
 - D. You need to set the SSD for cache or tier when doing the setup.

Contents

1. Product Overview
2. Hardware & Software Architecture
- 3. Target Sales Scenario**

Overview and Objectives

- This section describes main application where the new hybrid flash storage can be applied.
- On completion of this section, you will be able to:
 - Find the right sales target of new hybrid flash storage.

Six Enterprise IT Systems for All E2E Operations

Covers R&D, Sales, Executive Support, Onsite Management and Knowledge Sharing, Each with Different Storage Requirements.

Core System: Performs daily operations, including PSI and payment collection during producing flow, and requires high I/O processing capacity, low delay, no interruption or no data loss.

Transaction Processing System

Support System: Supports office work, collaboration, and enquiries into potential customers/clients. Its huge amounts of data are infrequently accessed, experiencing tidal phenomenon.

Office Automation System

Core System: Different types of enterprises use different systems, such as CAX for manufacturers, EDA for electronics companies, and IDE for software companies.

R&D Support System

Core System: Requires high-speed, stable date throughput, and uninterrupted analysis.

Decision Support System

Executive Support System

Support System: IT dept. customizes functions needed for procedures, requiring storage to quickly prepare dataset for dev & testing, and provide raw data isolation.

Support System: Stores quality documents and training course materials created by employees, partners, and certified engineers. The data access frequency gradually declines as the amount of data increases over time.

Knowledge Management System

Hybrid-Flash Storage Deployment

Smart Choice for Large-Enterprise OA and KMS Systems that Feature Mass Data, Hot and Cold Data Layering, and Data Value Decrease as Time Goes By



OA & KMS System

For enterprises of all sizes, their applications, such as **mail systems, knowledge sharing platforms, conference and IM, and enterprise portals**, hold hot and cold data:

- Long-term stability of business support systems: Improved operation efficiency required; but do not need the same reliability as core systems.
- Diverse data types and access protocols to add more applications: Infrequent data usage makes it uneconomical to keep data on the high-priced media.
- Peak performance: Large size of user groups increases data access, which impacts performance and causes tidal phenomenon during rush hours.

Core Requirements

- Cost-effective media tiering
- Automatic load balancing and quick access of hot data to handle burst peak traffic
- Long-term uninterrupted and stable operations
- Continuous evolution and containerized deployment of new applications

Overall Recommendation

All-Flash Storage for Most Enterprise Systems and Hybrid Flash Storage for Data Layering Environments, Such as OA and KMS Applications

Targeted Market Sectors

Finance: banking, insurance

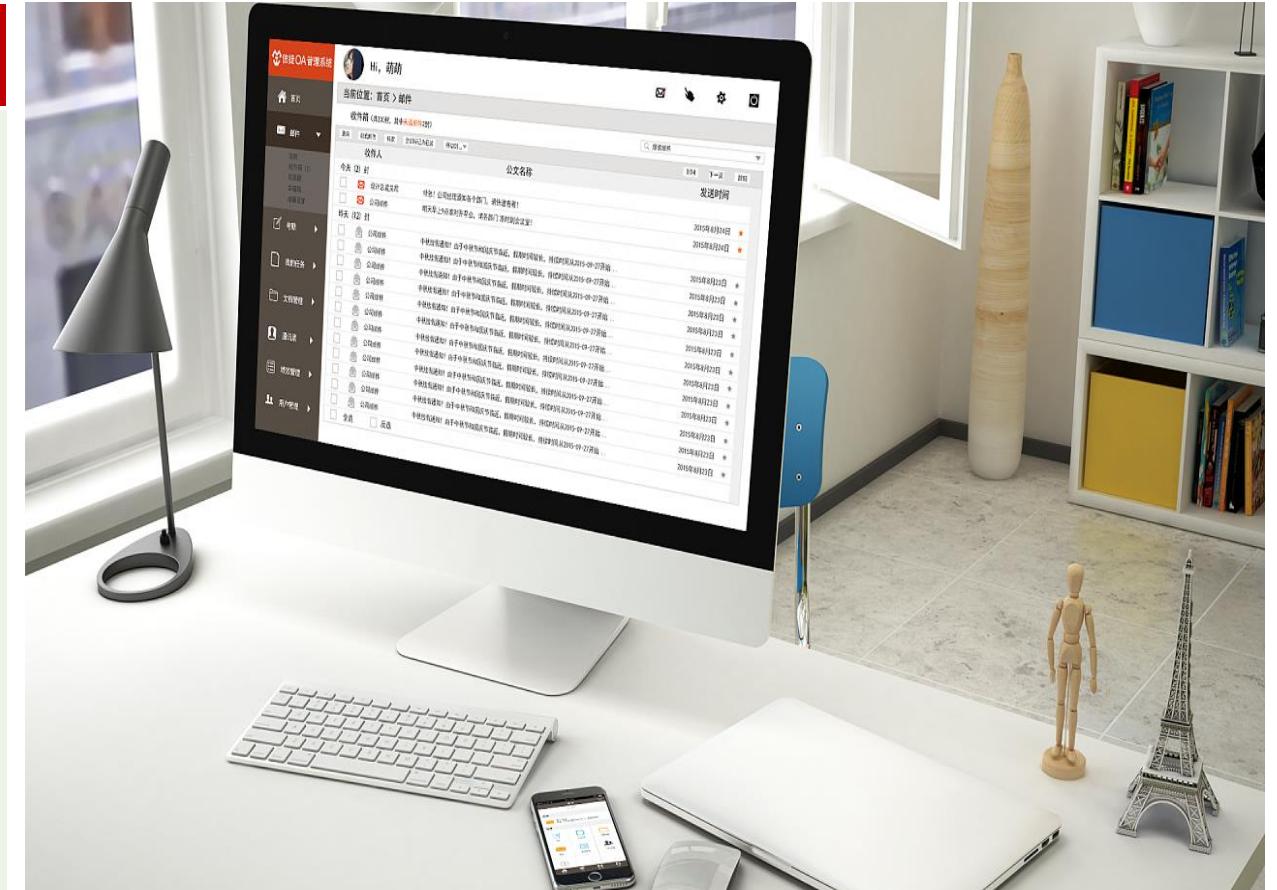
Government: courtrooms, finance & taxation, HR & social security, archives

Manufacturing: automobiles, electronics, heavy machinery and so on

Energy: oil and gas

Health: hospitals, pharmaceuticals, genomic research

Education: higher education institutions, scientific research institutions



Note: Hybrid flash storage is recommended for customers working on a budget or negotiating on price.

Summary

Key Technologies of OceanStor New Hybrid Flash Storage

- The New Hybrid Flash Storage has the same architecture with OceanStor Dorado series. Both performance and reliability are improved greatly from OceanStor V5.
- SmartAcceleration can improve the performance greatly and make the system design much easier.
- The new hybrid flash storage is fit for cold-hot data layering scenarios.

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 OceanStor Storage Reliability and Availability Technologies



Foreword

- Reliability and availability of storage product are key features customer cares. Huawei OceanStor flash storage products (OceanStor Dorado and New Hybrid Flash Storage) have multiple layers of technologies to ensure the customer data reliable and available in case of any failure.
- This chapter describes the reliability and available features and solutions provided by OceanStor Flash Storage, including OceanStor Dorado and New Hybrid Flash Storage.

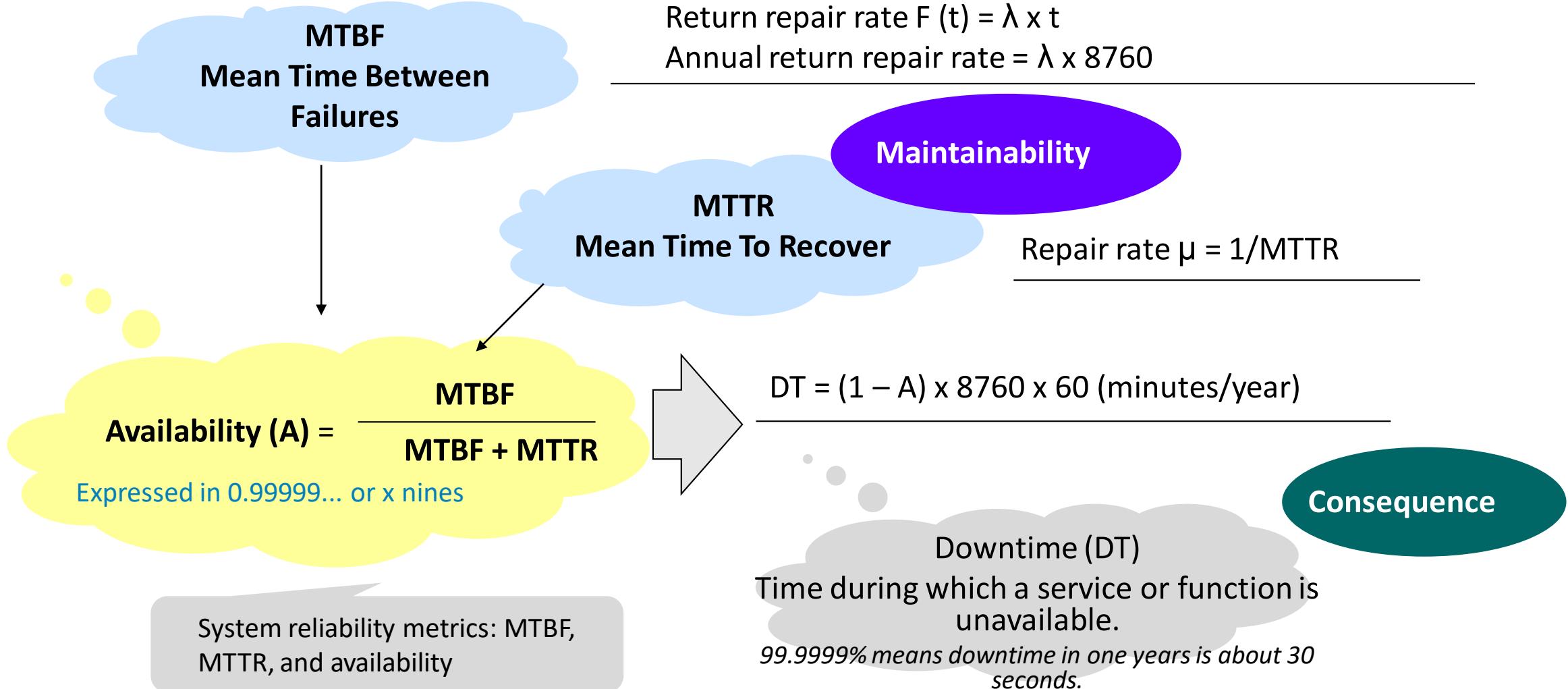
Contents

- 1. Storage Reliability Metrics**
2. Module-Level Reliability
3. System-Level Reliability
4. DC-Level Reliability
5. O&M Reliability
6. Reliability Tests and Certifications

Overview and Objectives

- This section describes the system reliability and availability metrics.
- On completion of this section, you will be able to:
 - Calculate the availability
 - Know the main reliability and availability features.

Storage Reliability Metrics



Overview of Storage System Reliability

Level 3: DC-level

Data protection and disaster recovery (DR) solutions provide system-level data protection and DR.

Service availability

HyperMetro (active-active)

3DC (geo-redundancy)

HyperReplication (remote replication)

O&M reliability

Fast upgrade

Intelligent prediction

Level 2: System level

System-level reliability design enables fault self-healing and data integrity protection for a system.

Multiple cache copies

Reconstruction offloading

I/O data protection

RAID 2.0+

Dynamic reconstruction

HyperSnap (snapshot)

Switchover within seconds

HyperMetro-Inner (high-end models)

Continuous mirroring

Overload control

High disk fault tolerance

Wear leveling

Bad block/sector scanning

Quick response to slow I/Os

Online diagnosis

Isolation of slow disks

Level 1: Module level

Lean manufacturing and processing ensure the yield rate.

Hardware reliability

Component

Device

Environment

Production

Disk

Quiz

1. (Single-choice) if the availability of one system is 99.999%, it means the downtime per year is less than:
 - A. 12 mins
 - B. 5 mins
 - C. 30 seconds

Contents

1. Storage Reliability Metrics
- 2. Module-Level Reliability**
3. System-Level Reliability
4. DC-Level Reliability
5. O&M Reliability
6. Reliability Tests and Certifications

Overview and Objectives

- This section describes the module reliability.
- On completion of this section, you will be able to:
 - Describe the technologies applied to ensure reliability of HSSD.

Module-Level Reliability Overview

Module Level

System Level

DC Level

Component	Component reliability	Disk reliability	Data integrity	Planned activity
	<ul style="list-style-type: none"> Optical module Clock component Storage component Connector Dedicated IC 	<ul style="list-style-type: none"> Fault forecast Production-phase filtering Anti-vibration design ERT Heat dissipation design 	<ul style="list-style-type: none"> T10 PI Chip ECC/CRC Soft failure prevention Protocol data integrity Data storage redundancy 	<ul style="list-style-type: none"> Error prevention Replacement of faulty components Reliable upgrade Reliable expansion
Media	Reliability of Huawei-developed chips	HSSD reliability	Manufacturing reliability	Running fault detection and self-healing
Board	Board reliability	<ul style="list-style-type: none"> Backup power reliability RAID Load balancing algorithm Error correction algorithm Bad block management 	<ul style="list-style-type: none"> Key hardware signal Low-speed management bus Board temperature Storage component Clock signal Board power supply Board process/material BIST SI/PI 	<ul style="list-style-type: none"> Fault diagnosis and locating Fault prediction Power on self-test Environment check Component check and self-healing Functional module check and self-healing Software check and self-healing
Device	Redundancy design	System power supply/backup power	System cooling	
Environment	<ul style="list-style-type: none"> FRU/Network redundancy 	<ul style="list-style-type: none"> PDU/BBU/CBU/Power supply reliability 	<ul style="list-style-type: none"> Fan reliability 	
	Power supply	Temperature	Anti-vibration	Anti-corrosion
				Altitude
				Dust proof
				Moisture proof
				EMC
				Security standards compliance

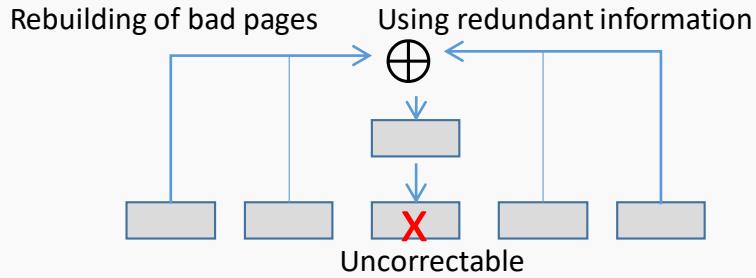
Module-Level Reliability: HSSD Reliability

Module Level

System Level

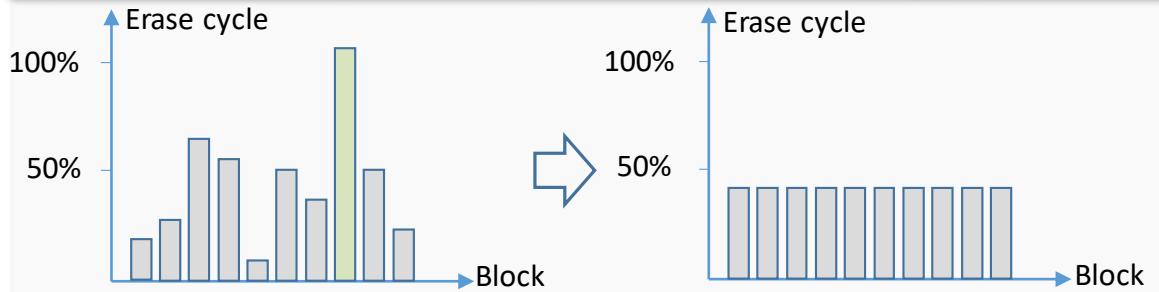
DC Level

Data redundancy



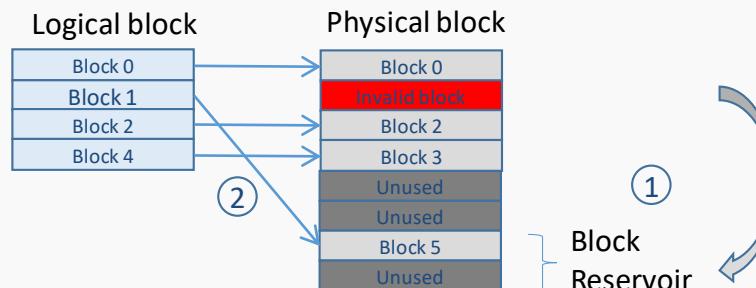
- **Die-level Multi-copy & RAID:** metadata (multiple copies) and user data (RAID)
- **Data restoration:** LDPC, read retry, and intra-disk XOR that enable data restoration using redundancy

Wear leveling



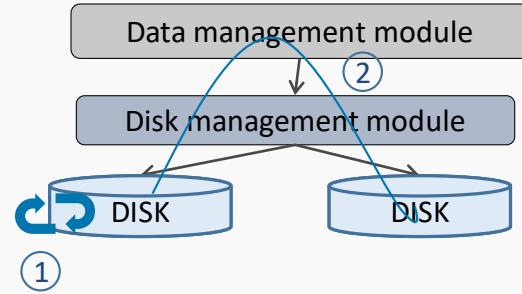
Wear leveling: periodically moves data blocks so that the data blocks with less wear can be used again.

Bad block management and background inspection



1. **Background inspection:** combines read inspection and write inspection and proactively reports bad blocks detected during inspection.
2. **Bad block isolation:** detects, migrates, and isolates bad blocks.

Advanced management



1. **Online self-healing:** restores a disk to its factory settings online.
2. **Die failure:** active reporting and capacity reduction
3. **Power failure protection:** flush dirty data when power failure using capacitor

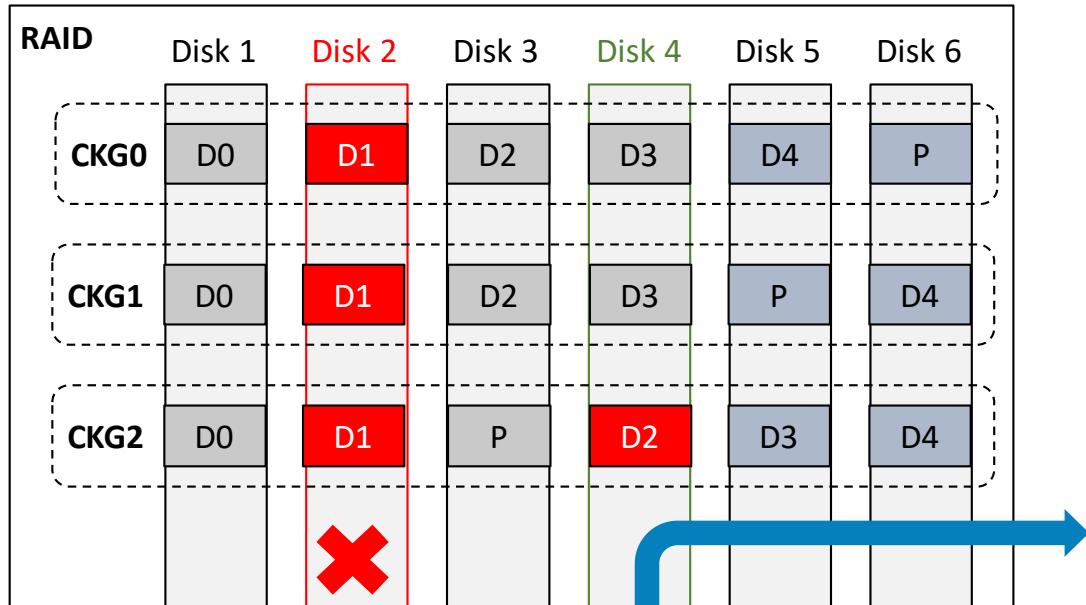
Module-Level Reliability: HSSD RAID

Module Level

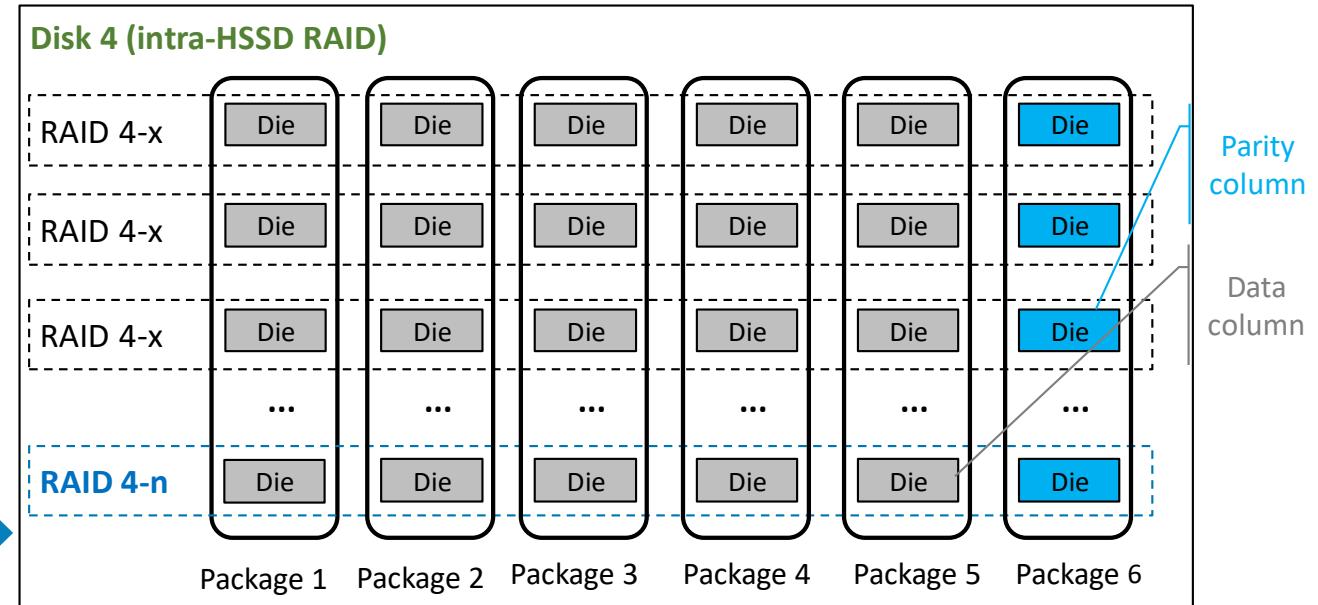
System Level

DC Level

Conventional SSD: silent data corruption (D3) + single-disk fault → data loss



HSSD: bad block (die) self-recovery + single-disk fault → no data loss



- **Without RAID:** Silent data corruption may occur on disks. (HDDs may have bad sectors and SSDs may have bad blocks, on which data is unavailable.) If such bad sectors or blocks are seldom accessed, corruption of data in them cannot be detected or rectified in time. Once a disk fails, data in these bad sectors or blocks cannot participate in reconstruction, resulting in loss of user data.
- **With RAID:** HSSDs periodically scan data blocks and restore detected bad blocks using intra-disk RAID. In addition, data in bad blocks can be restored in real time using inter-disk RAID when these blocks are accessed by a host or participate in data reconstruction. In this way, data will not be lost.

Quiz

1. (Single-choice) If finding that one SSD in storage system is running heavy workload, and its life time is much shorter than others, which following technology can avoid this problem?
 - A: Garbage collection
 - B: Wear leveling
 - C: Bad block isolation
 - D: Inner disk RAID

Contents

1. Storage Reliability Metrics
2. Module-Level Reliability
- 3. System-Level Reliability**
4. DC-Level Reliability
5. O&M Reliability
6. Reliability Tests and Certifications

Overview and Objectives

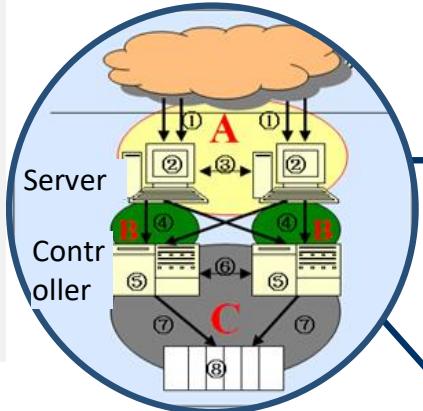
- This section describes the system reliability technologies.
- On completion of this section, you will be able to:
 - Describe the technologies applied to ensure high service availability.
 - Describe the technologies applied to ensure solid data reliability.
 - Describe the technologies applied to ensure high disk fault tolerance.

System-Level Reliability

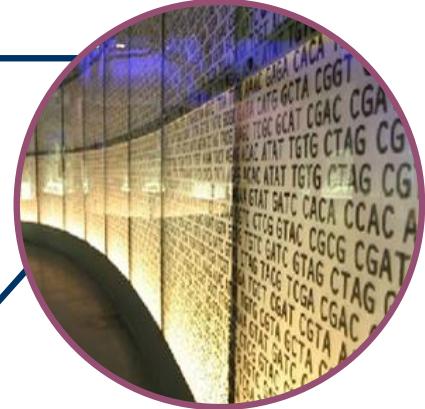
Module Level

System Level

DC Level



- High disk fault tolerance**
- Wear leveling/Anti-wear leveling
 - Bad sector/block scanning and repair
 - Online diagnosis
 - Quick response to slow I/Os
 - Isolation of low disks

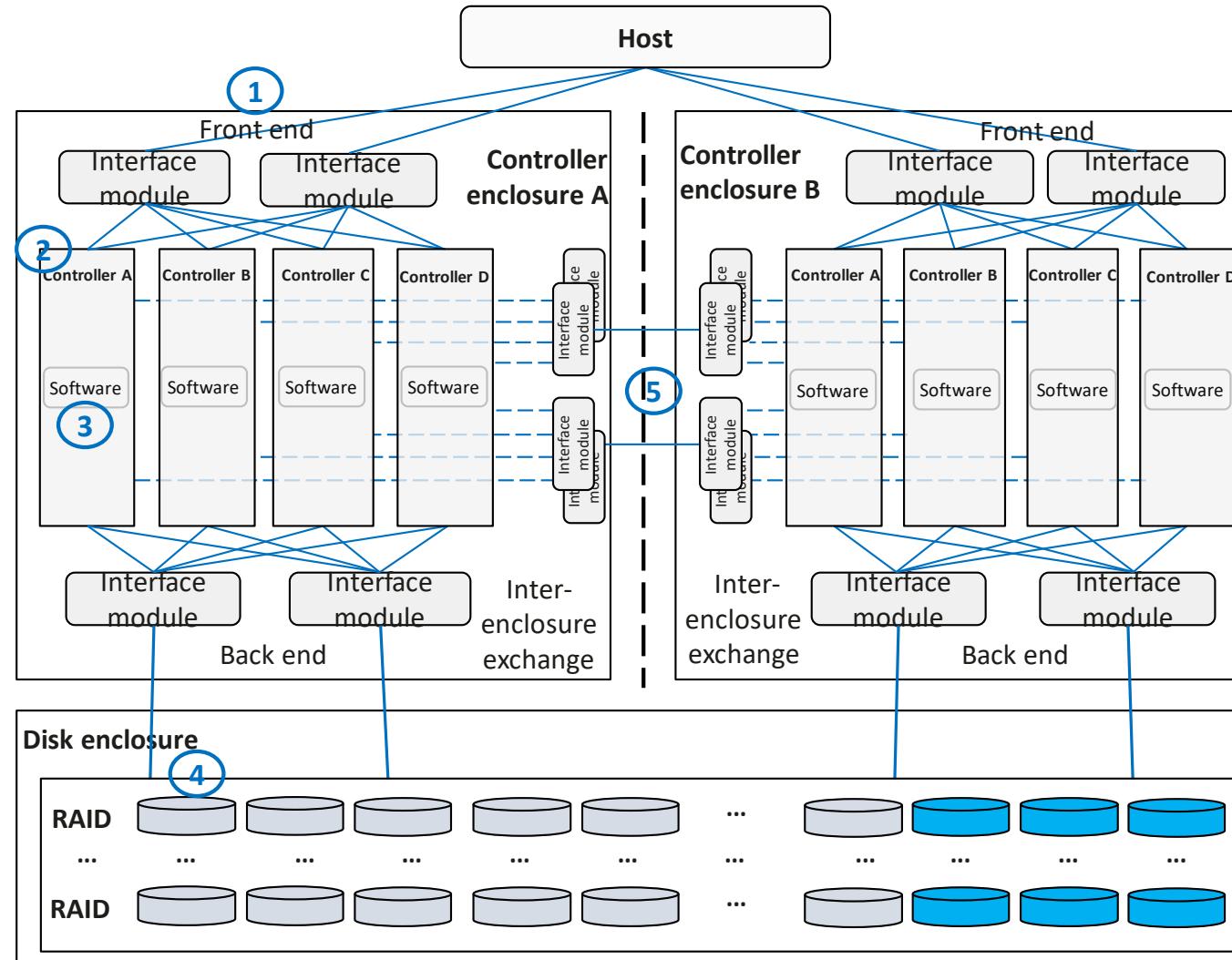


High Service Availability

Module Level

System Level

DC Level



E2E Redundancy Design

1. **Controller switchover is transparent to services:** front-end interconnect I/O modules (FIMs), protocol offloading, and controller failover within seconds
2. **Services are not interrupted if multiple controllers are faulty (HyperMetro-Inner, for high-end):** three cache copies, continuous mirroring, cross-engine mirroring, and full-mesh back end.
3. **Services are not affected in the case of a software fault:** process availability detection, startup of processes within seconds upon a fault, and intermittent isolation of frequently abnormal background tasks.
4. **Services are not interrupted if multiple disks are faulty:** EC-2/EC-3 for user data
5. **Controllers do not reset if interface modules for inter-enclosure exchanging are faulty:** multiple interface modules for redundancy on high-end storage, and TCP forwarding for mid-range and entry-level storage with a single interface module

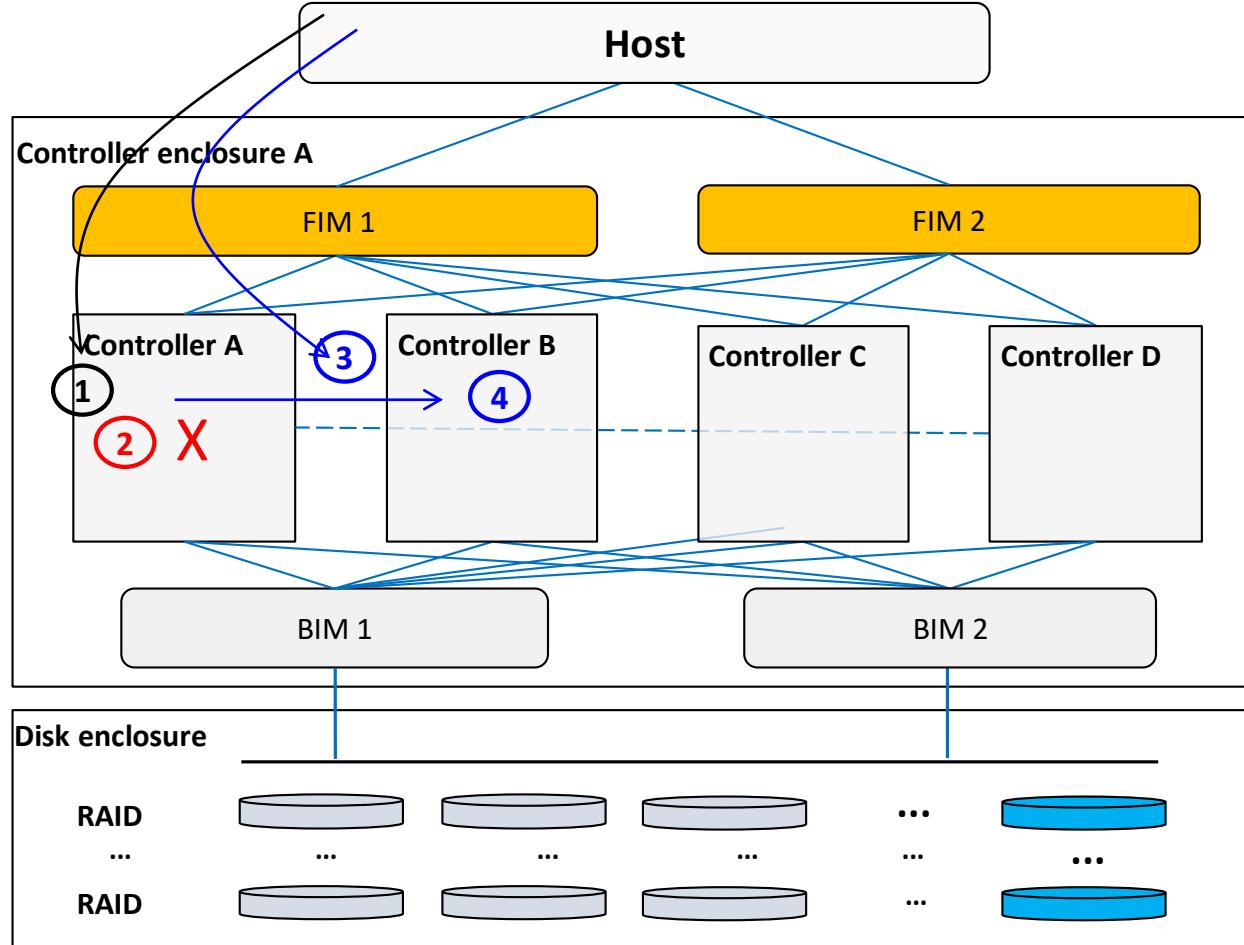
High Service Availability: Controller Failover Within Seconds

Module Level

System Level

DC Level

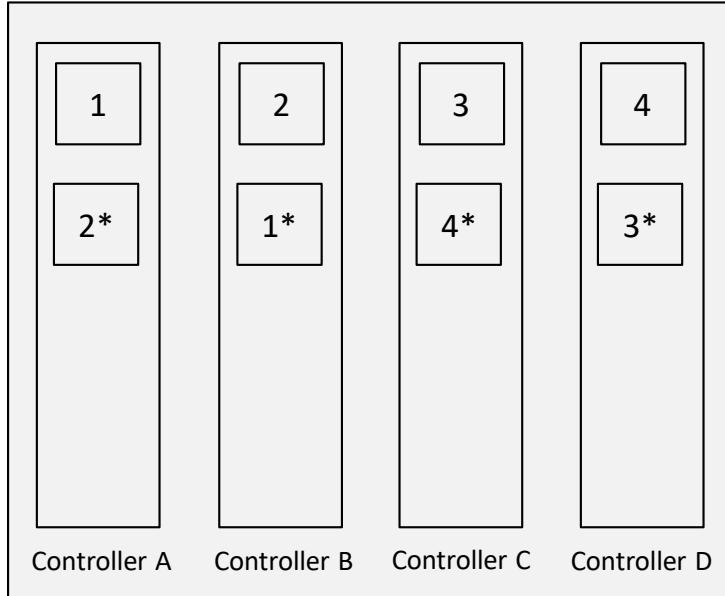
Quick Controller Failover



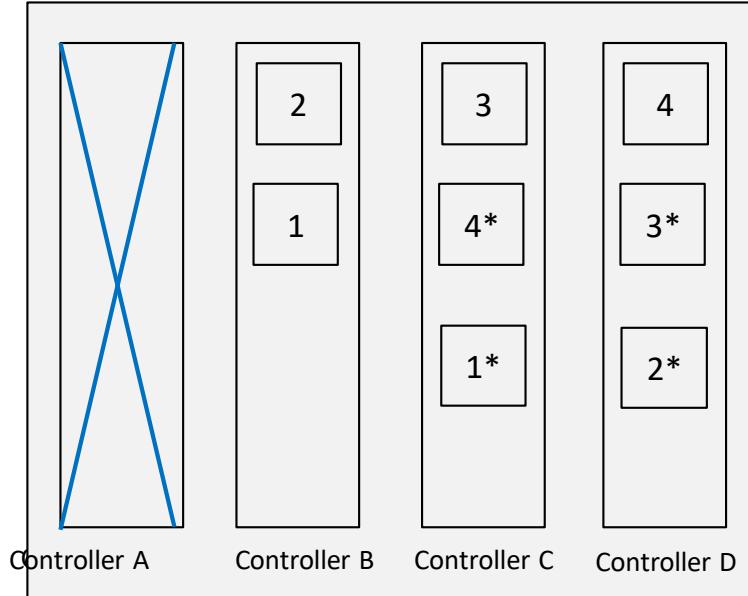
1. **The host delivers I/Os to controller A:** If all controllers are normal, I/Os are delivered to controller A through FIM 1.
2. **Controller A is faulty:** FIM 1 and controller B detect that controller A is unavailable by means of interrupts.
3. **Service switchover:** Services are quickly (within 1s) switched to controller B that has data copies of controller A, by switching the vNode. Then FIMs are instructed to refresh the distribution view.
4. **I/O path switchover:** FIM 1 returns **BUSY** for the I/Os that have been delivered to controller A. The retried and new I/Os delivered by the host are delivered to controller B based on the new view.

High Service Availability: Continuous Mirroring

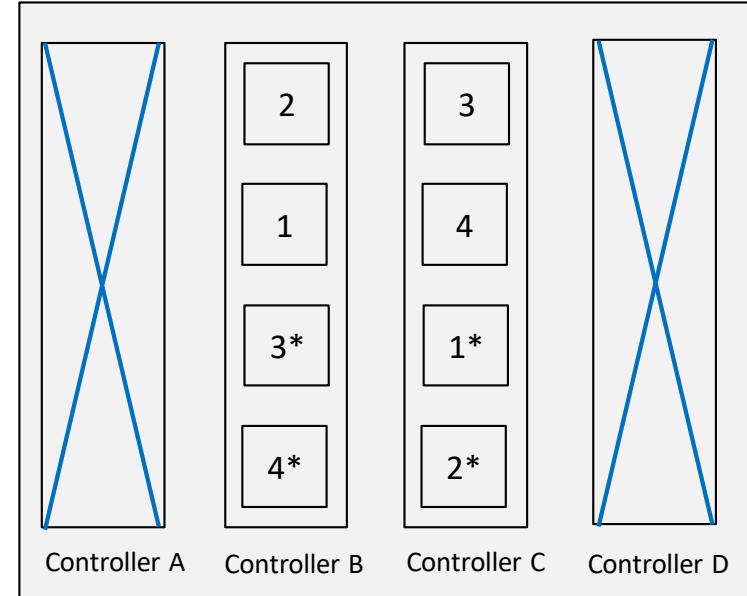
Module Level System Level DC Level



Normal



Failure of one controller (controller A)



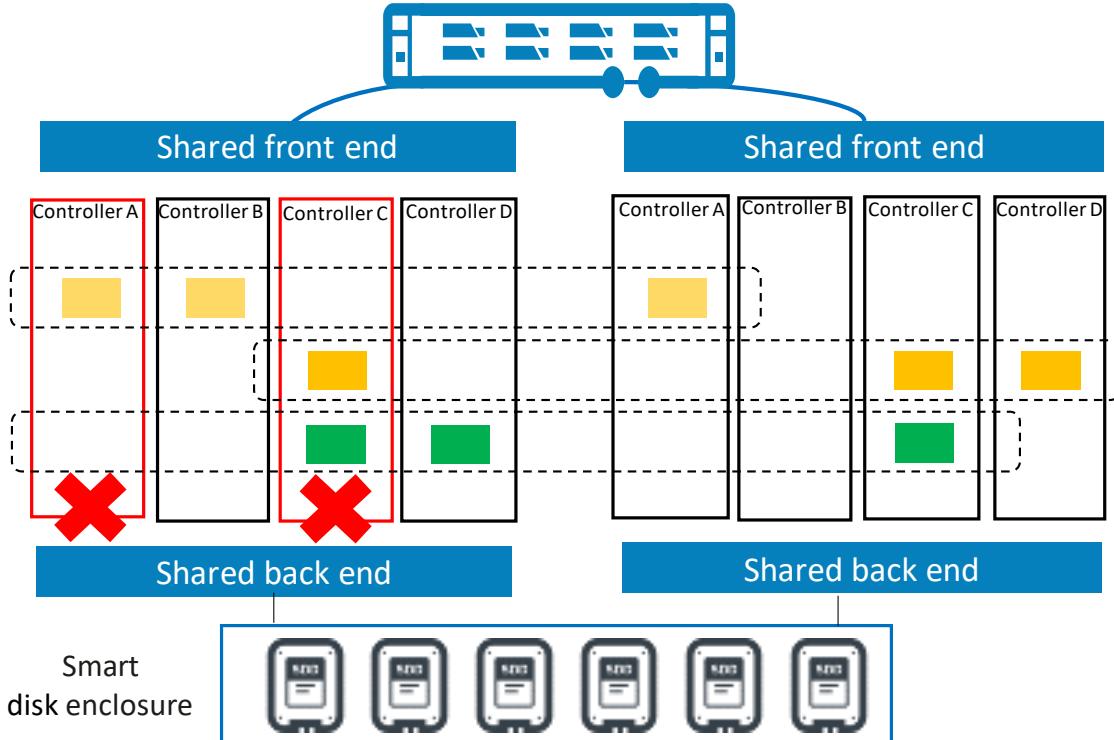
Failure of one more controller (controller D)

- **Continuous mirroring (ensuring service continuity even when seven out of eight controllers are faulty):** If controller A is faulty, controller B selects controller C or D as the cache mirror. If controller D fails at the moment, cache mirroring is implemented between controller B and controller C to ensure dual-copy redundancy.
- **Service continuity:** If a controller fails, its mirror controller establishes a mirror relationship with another functional controller within 5 minutes. This design increases the service availability by at one nine and ensures service continuity in the event that multiple controllers fail successively.

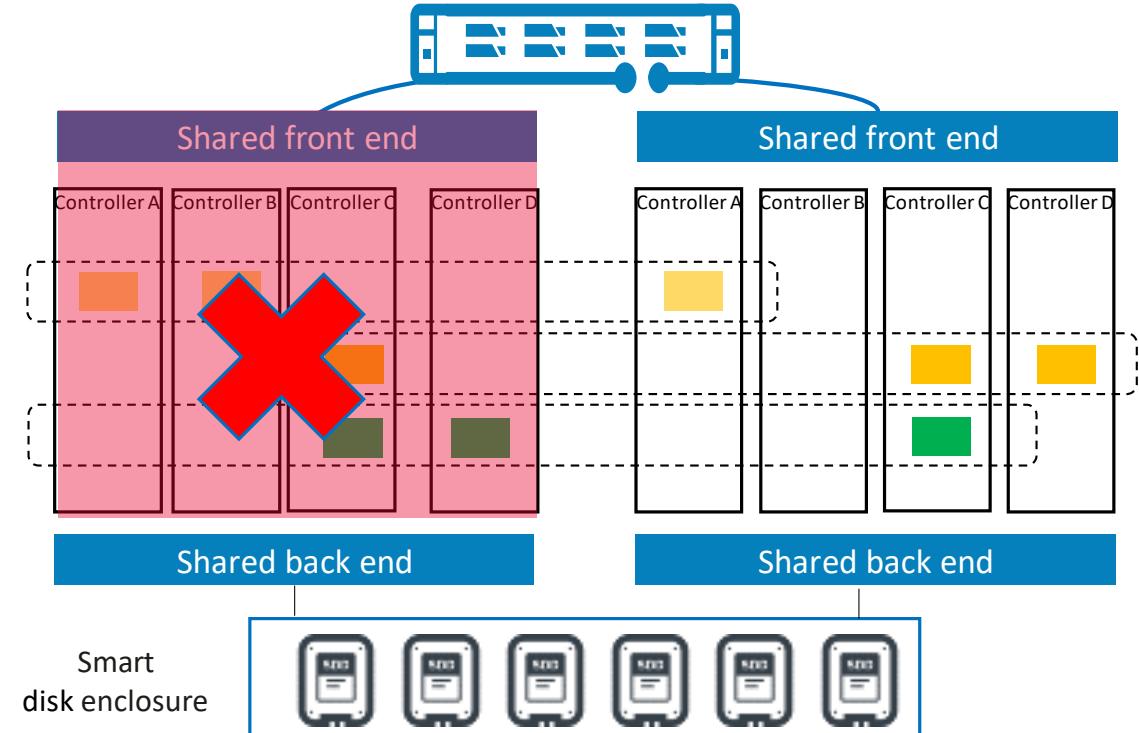
High Service Availability: HyperMetro-Inner for High-End Storage

Module Level System Level DC Level

Tolerating simultaneous failure of two controllers



Tolerating failure of a single controller enclosure

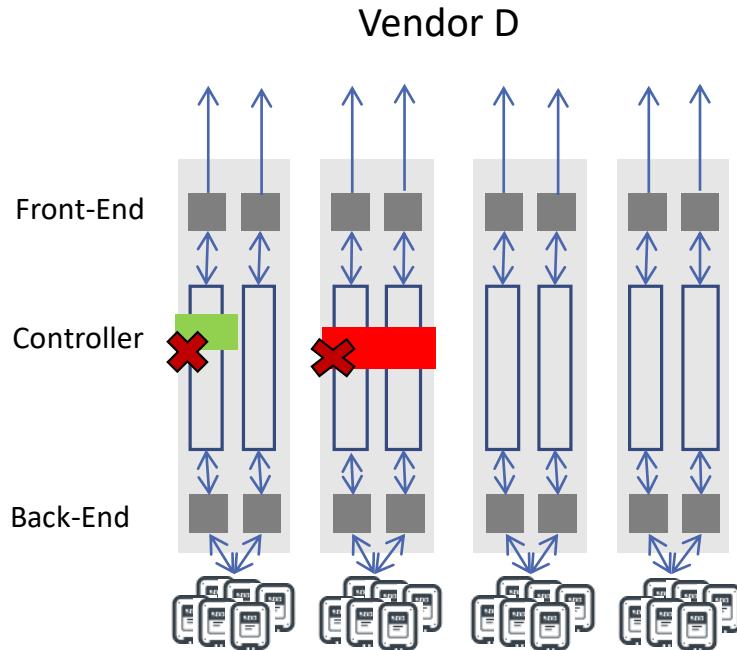


- The global cache provides three cache copies across controller enclosures.
- If two controllers fail simultaneously, at least one cache copy is available.
- A single controller enclosure can tolerate simultaneous failure of two controllers with the three-copy technology.

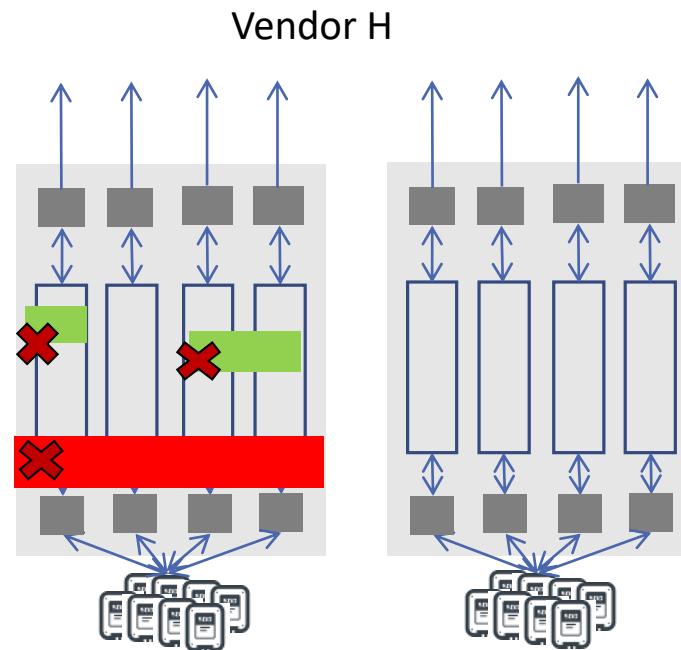
- The global cache provides three cache copies across controller enclosures.
- A smart disk enclosure connects to 8 controllers (in 2 controller enclosures) through BIMs.
- If a controller enclosure fails, at least one cache copy is available.

SmartMatrix: Industry's Unique Architecture, The Only Storage Which Tolerant 7-out-of-8 Controller Failure

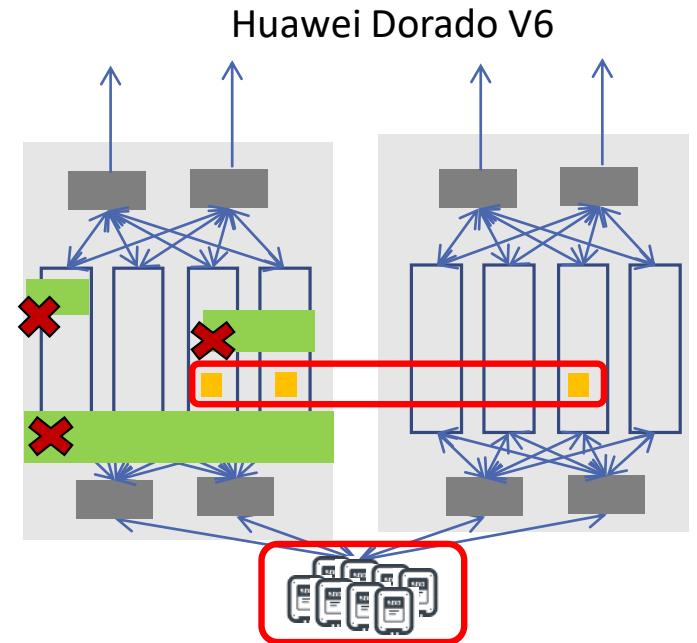
Module Level System Level DC Level



- Disk enclosure shared with dual-controller
- **dual-controller**(one engine) failure causes service interruption.



- Disk enclosure shared with four-controller
- **4 controller** failure(one engine) causes service interruption.



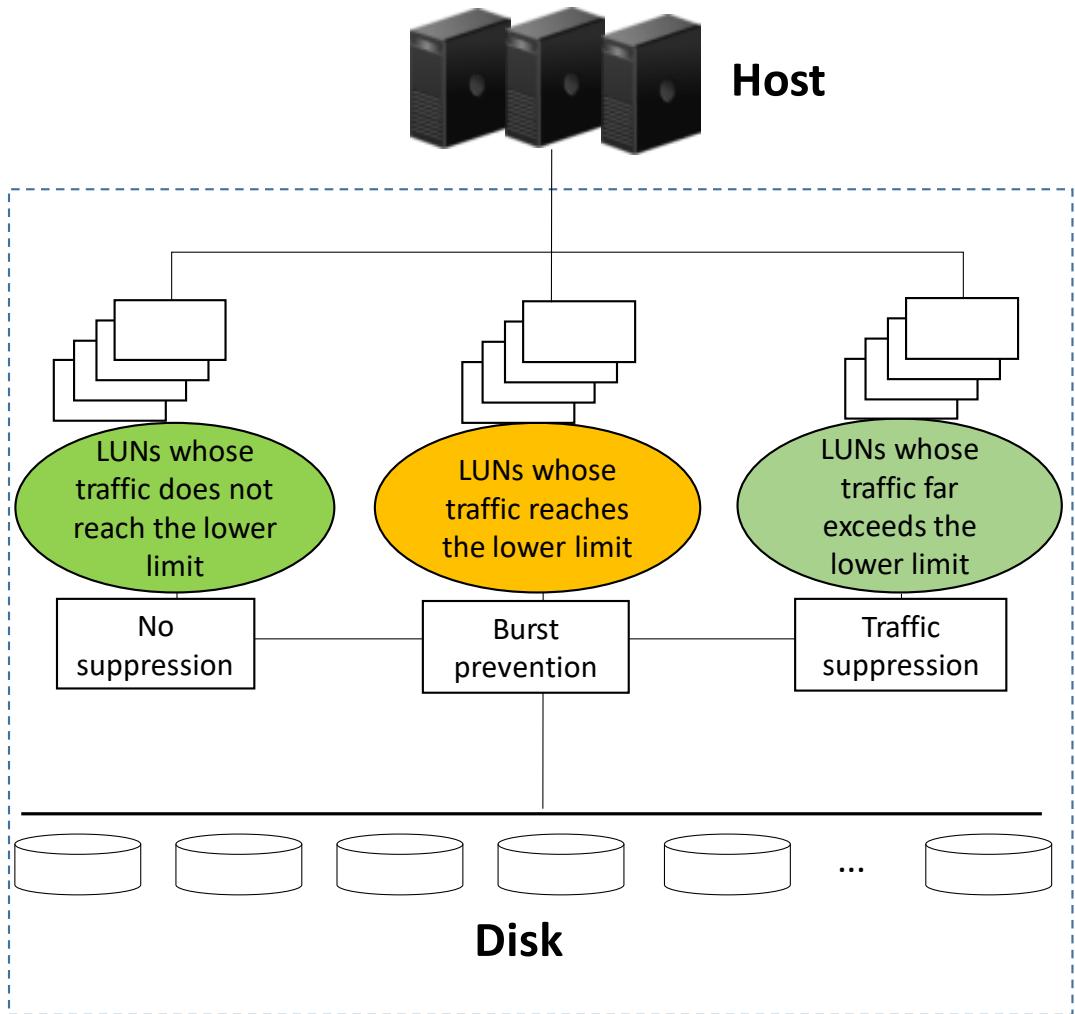
- Disk enclosure shared with 8 controllers.
- **7 controllers** failure one by one of 8 controllers(2 engines)

High Service Availability: SmartQoS

Module Level

System Level

DC Level



Upper Limit Control

- **Priority setting:** The storage system converts the traffic control objective into a number of tokens. You can set upper limit objectives for low-priority LUNs or snapshots to guarantee sufficient resources for high-priority LUNs or snapshots.
- **Token application:** The storage system processes the dequeued LUN or snapshot I/Os by tokens. I/Os can be dequeued and processed only when sufficient tokens are obtained.

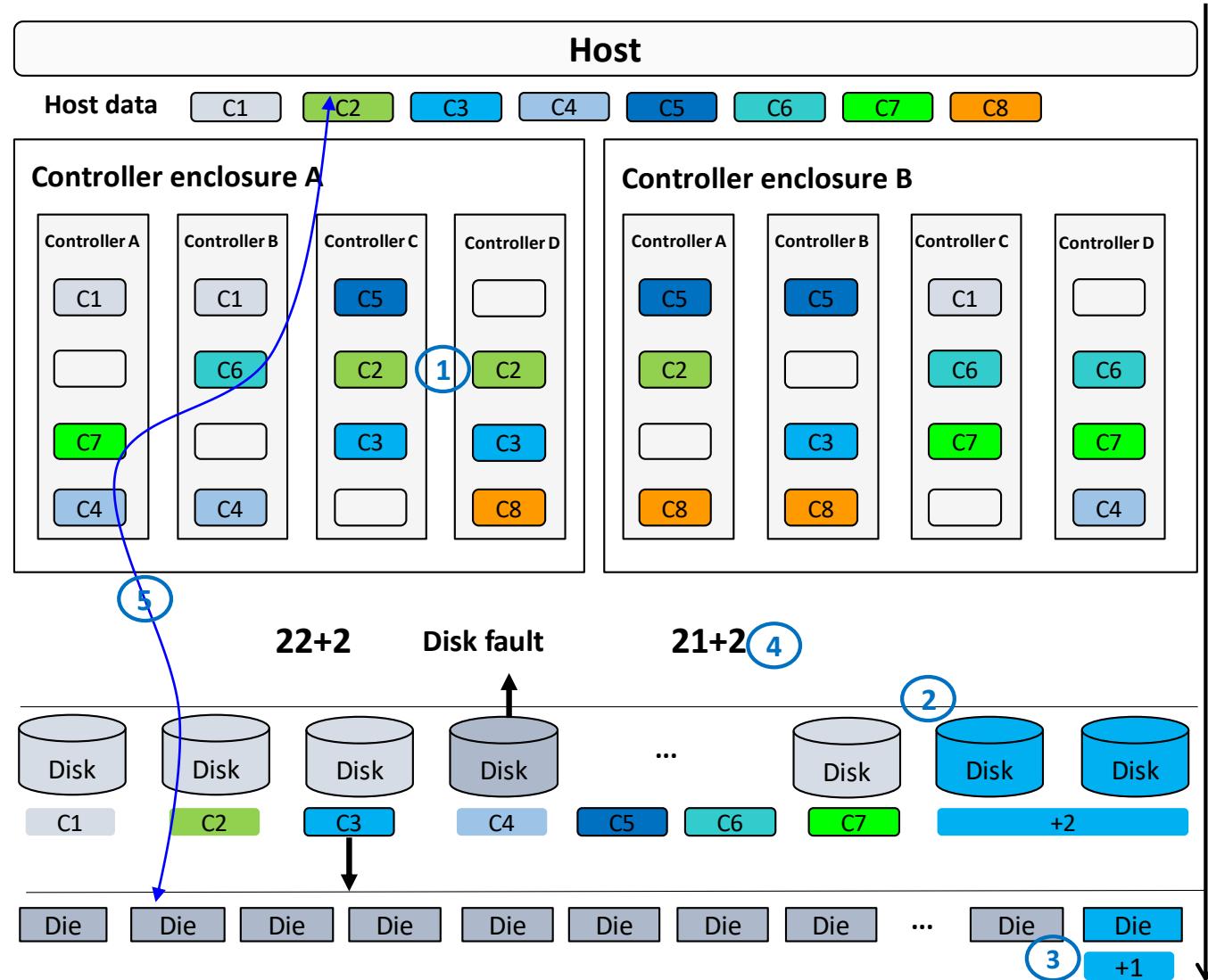
Burst Quota

- **Token accumulation:** If the performance of a LUN, snapshot, LUN group, or host is lower than the upper threshold within a second, a one-second burst duration is accumulated. When the service pressure suddenly increases, the performance can exceed the upper limit and reach the burst traffic. The accumulated tokens are used by the current objects and last the configured duration. In this way, the system can respond to the burst traffic in time.

Lower Limit Guarantee

- **Minimum traffic:** If each LUN is configured with the minimum traffic (IOPS/bandwidth) by default, the minimum traffic must be ensured when the LUN is overloaded.
- **Traffic suppression for high-load LUNs:** When the system is overloaded, if the traffic of some LUNs does not reach the lower limit, the system performs load rating on all LUNs. The system provides a loose traffic condition for medium- and low-load LUNs based on the load status. The system suppresses the traffic of high-load LUNs until the system releases sufficient resources to enable the traffic of all LUNs to reach the lower limit.

Solid Data Reliability



Module Level

System Level

DC Level

E2E Data Protection

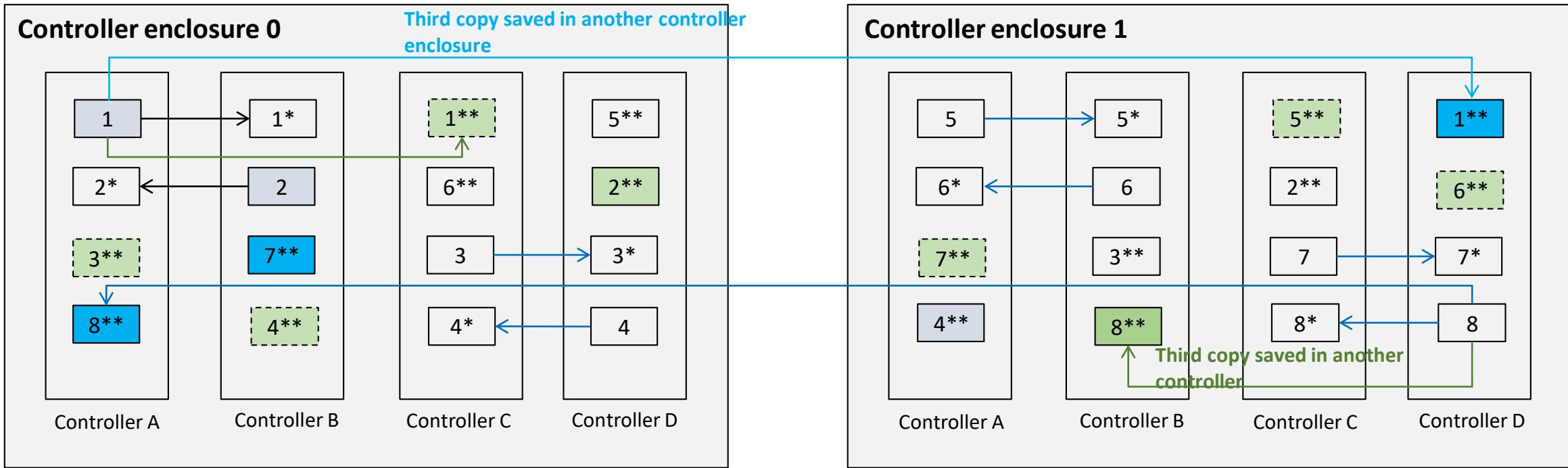
- Cache data redundancy:** Two or three copies of cache data ensure no data loss in the case that multiple controllers or a single controller enclosure is faulty.
- Disk data redundancy:** RAID 2.0+ ensures that user data on disks will not be lost in the case that multiple disks are consecutively or simultaneously faulty. Data reconstruction is offloaded to smart disk enclosures, further ensuring data reliability.
- Intra-disk data redundancy:** RAID 4 ensures die-level redundancy within a disk, preventing user data loss in the case of bad blocks or die failure.
- Maintained data redundancy even when RAID disks are insufficient:** Dynamic reconstruction, in which fewer data disks are involved, is used to maintain redundancy if the number of member disks does not meet RAID requirements.
- E2E data consistency:** E2E PI and parent-child hierarchy verification ensure that data on I/O paths will not be damaged.
- Snapshot technology **HyperSnap** and Clone technology **HyperClone** provide multiple copies in system for quick data recovery.

Solid Data Reliability: Multiple Cache Copies

Module Level

System Level

DC Level



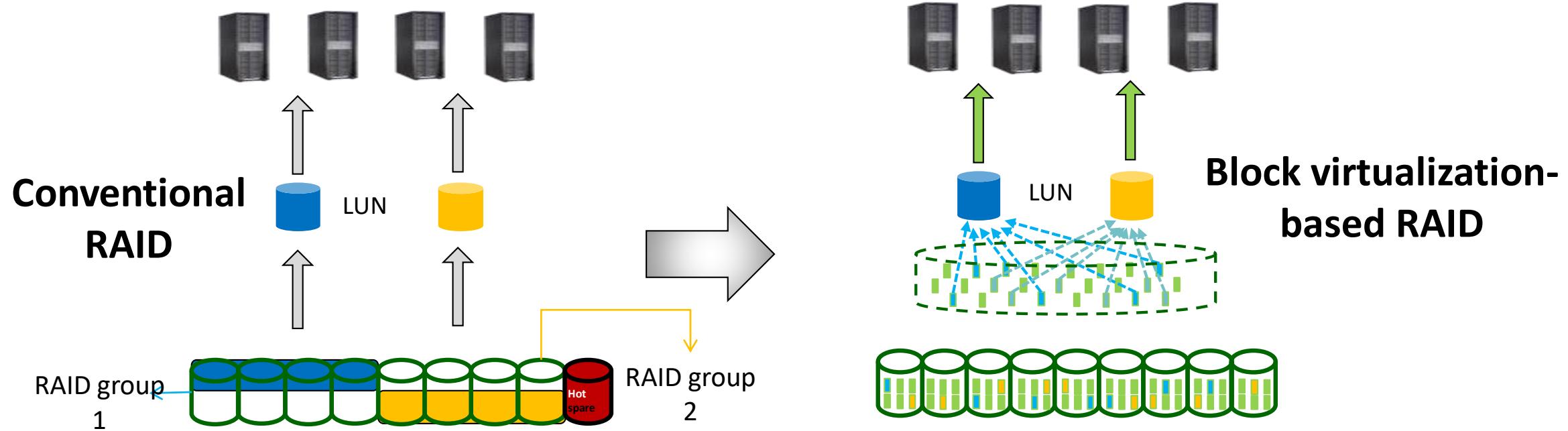
- **Data will not be lost if two controllers are faulty:** Three copies of cache data are supported. For host data with the same LBA, the system creates a pair of cache data copies on two controllers and the third copy on another controller.
- **Data will not be lost if a controller enclosure is faulty:** When the system has two or more controller enclosures, three copies of cache data are saved on controllers in different controller enclosures. This ensures that cache data will not be lost in the event that a controller enclosure (containing four controllers) is faulty.

Solid Data Reliability: RAID 2.0+

Module Level

System Level

DC Level



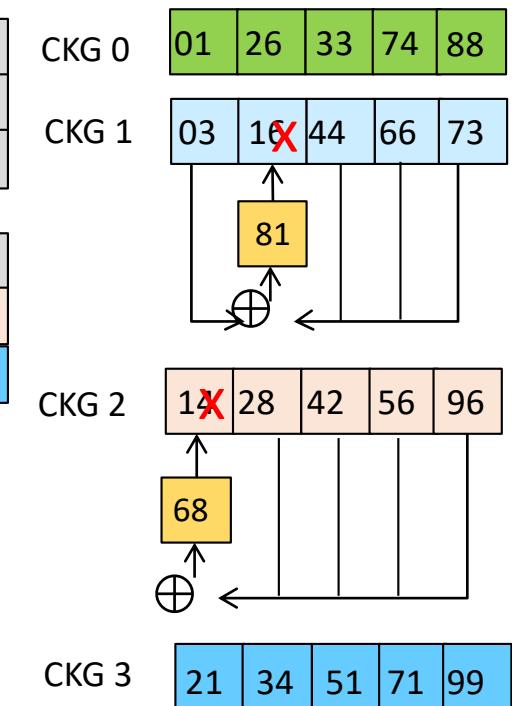
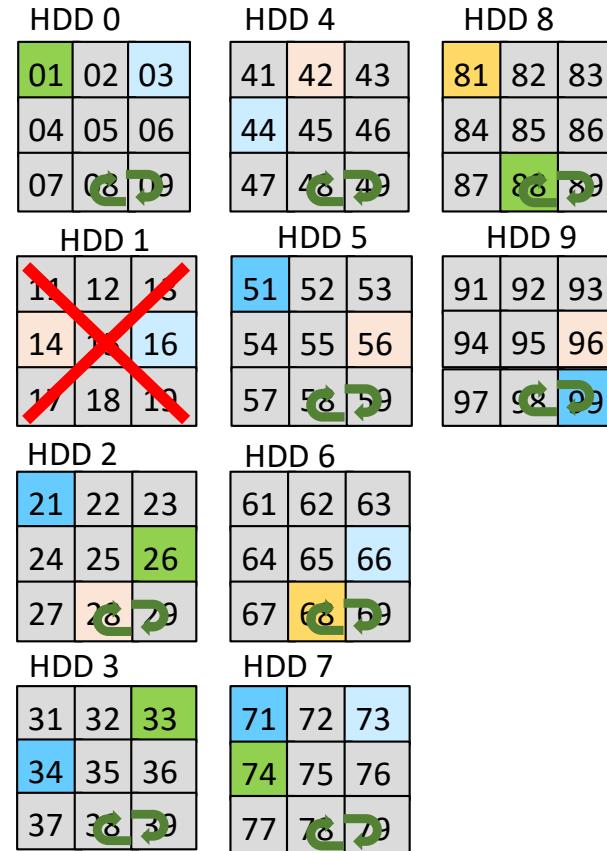
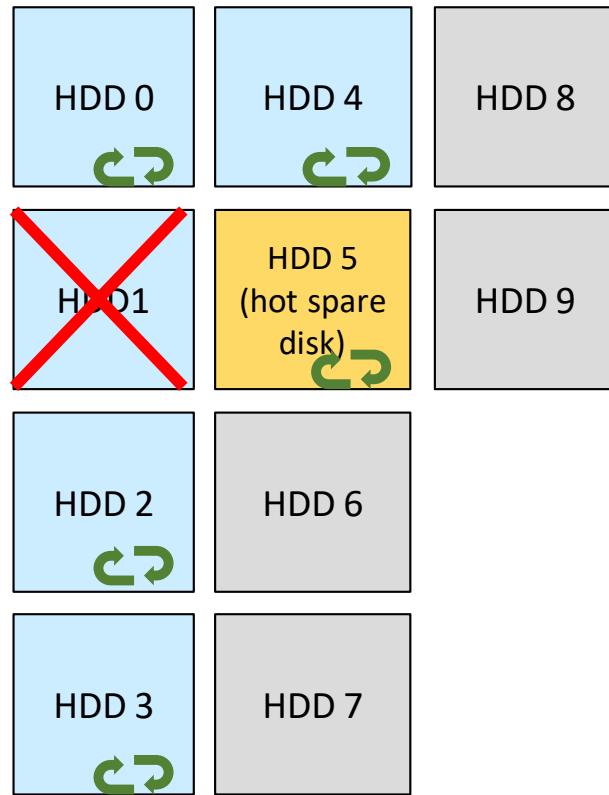
Conventional RAID	Block Virtualization-based RAID
Resource management based on disks	Resource management based on data blocks
I/Os of a LUN are processed by limited disks in a RAID group.	I/Os to each LUN are evenly distributed to all disks, balancing performance.
Slow reconstruction: If a single disk is faulty, only a limited number of disks in the RAID group participate in reconstruction.	Fast reconstruction: If a single disk is faulty, all disks participate in reconstruction.
A hot spare disk must be specified. Once the hot spare disk is faulty, it must be replaced in time.	Reconstruction can be performed as long as there is free space, independent of specific hot spare disks.

Solid Data Reliability: Fast Reconstruction

Module Level

System Level

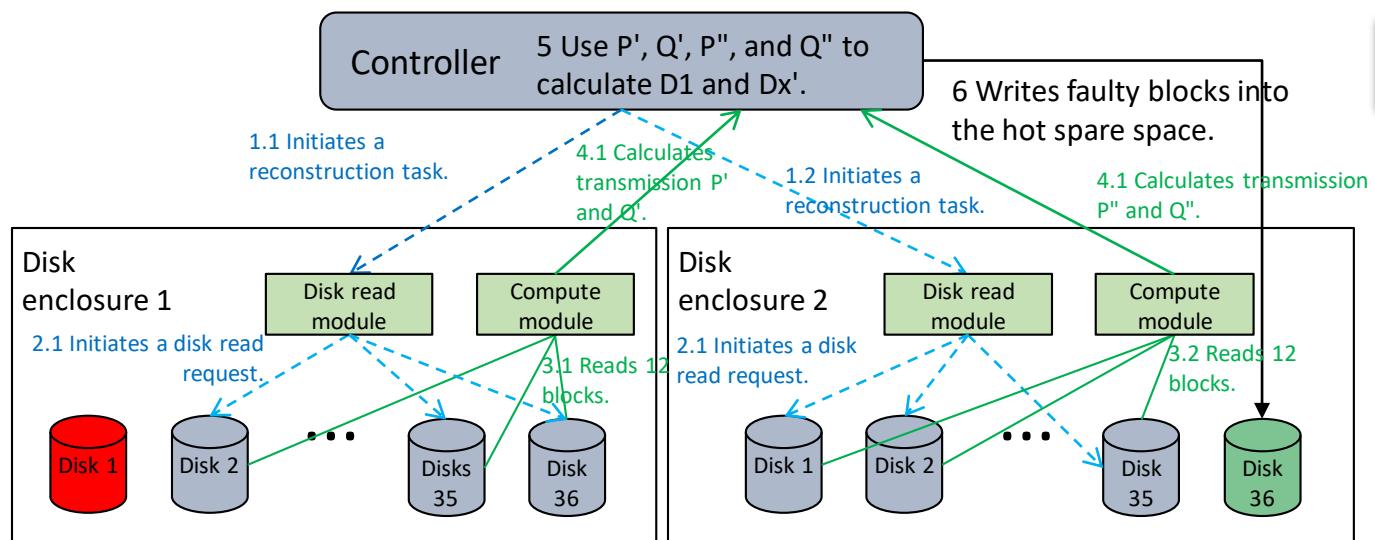
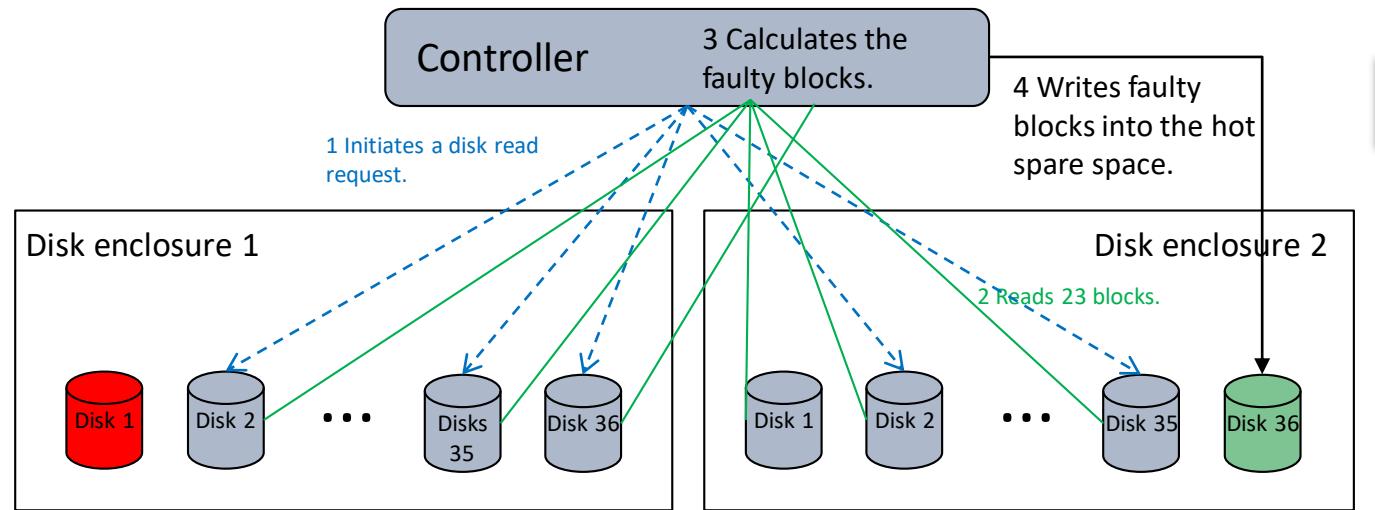
DC Level



Reconstruction using conventional RAID: During reconstruction, data is read from the other functional disks and reconstructed. Then, reconstructed data is written to a hot spare disk or a new disk. The write performance of a single disk restricts the reconstruction. Therefore, the reconstruction takes a long time.

Reconstruction using RAID 2.0+: RAID 2.0+ supports dozens of member disks. When a disk fails, other disks participate in reconstruction reads and writes, greatly shortening the reconstruction time. As more disks share reconstruction loads, the load on each disk significantly decreases.

Solid Data Reliability: Reconstruction Offloading



Module Level

System Level

DC Level

Controller-based Reconstruction

- Reconstruction occupies controller computing resources (CPU resources):** If a single or multiple disks are faulty, all data is computed on the controller, causing the controller CPU to be overloaded. The host I/O processing capabilities are adversely affected.
- Reconstruction occupies massive data write bandwidth:** All data on the disks in the RAID group is read to the controller for computing, occupying the data write bandwidth. As a result, the host I/O write bandwidth is affected.

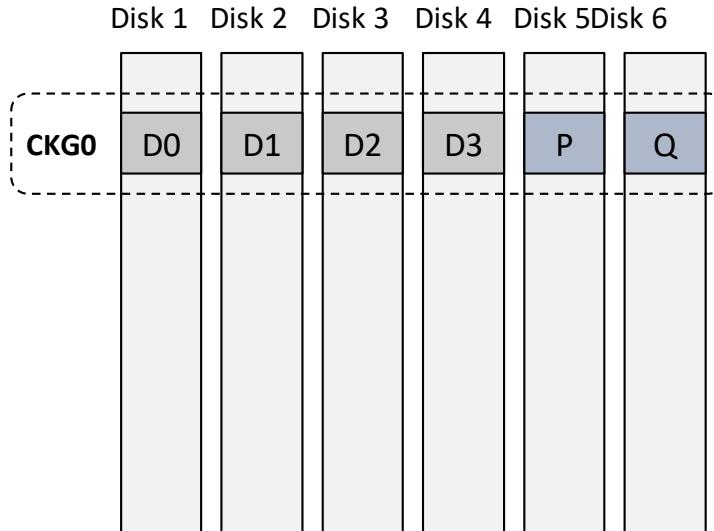
Reconstruction Offloading (2x Better Performance)

- The computing of RAID member disks is offloaded to smart disk enclosures:** Smart disk enclosures have idle CPU resources. The disk recovery data read by the IP enclosure is calculated in the enclosure by using P' and Q'.
- Reconstruction occupies a small amount of data write bandwidth:** When RAID data recovery is involved, data on the disk to be recovered is computed in the smart disk enclosure and does not need to be transmitted to the controller, reducing back-end bandwidth occupation and reducing the impact of reconstruction on system performance.

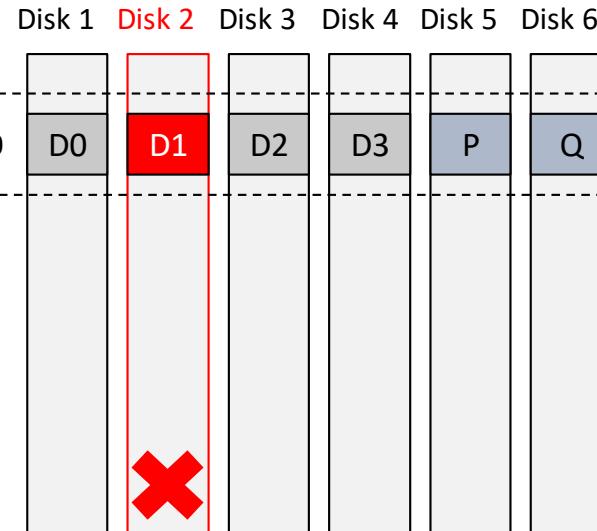
Solid Data Reliability: Dynamic Reconstruction

Module Level System Level DC Level

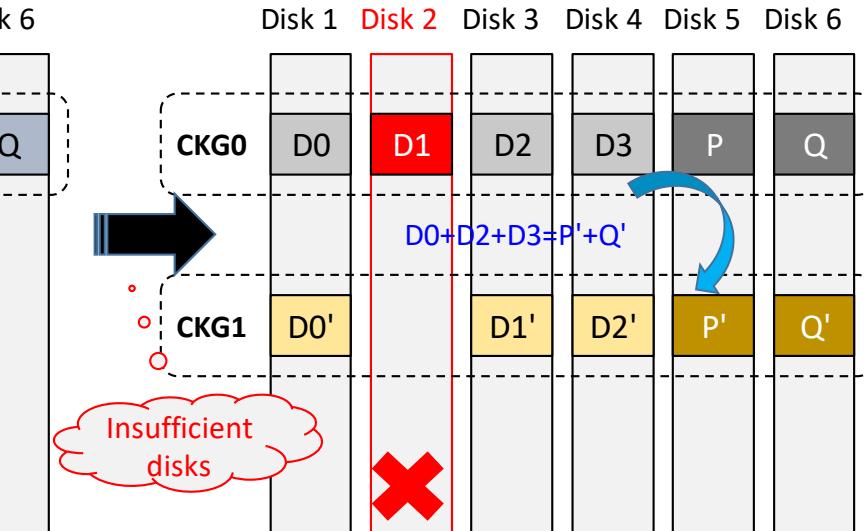
Normal: RAID (4+2)



A RAID member disk is faulty.



Dynamic reconstruction: RAID (3+2)



For a RAID group has $M+N$ members (M indicates data columns and N indicates parity columns):

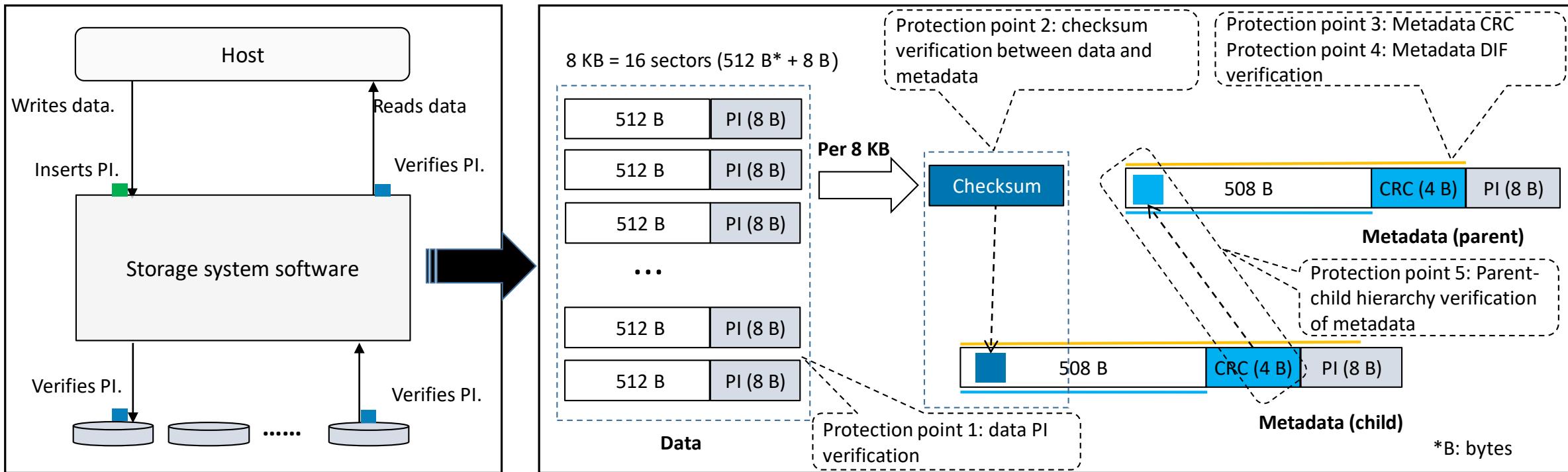
- **Common reconstruction:** When a disk is faulty, the system uses an idle CK to replace the faulty one and restores data to the idle CK. If the number of member disks in the disk domain is fewer than $M+N$, two CKs reside on the same disk, decreasing the RAID redundancy level.
- **Dynamic reconstruction:** If the number of member disks in the disk domain is fewer than $M+N$, the system reduces the number of data columns (M) and retains the number of parity columns (N) during reconstruction. This method retains the RAID level, ensuring system reliability.

Solid Data Reliability: E2E Data Protection

Module Level

System Level

DC Level



Data protection at hardware boundaries: Data verification is performed at multiple key nodes on I/O paths within a storage system, including front-end chips, controller software front end, controller software back end, and back-end chips.

Multi-level software based data protection: For each 512-byte data, in addition to 8-byte PI (two bytes of which are CRC bytes), the system extracts the CRC bytes in 16 PI sectors to form the checksum and stores the checksum in the metadata node. If skew occurs in a single or multiple pieces of data (512+8), the checksum is also changed and becomes inconsistent with that saved in the metadata node. When the system reads the data and detects the inconsistency, it uses the RAID redundancy data on other disks to recover the error data, preventing data loss.

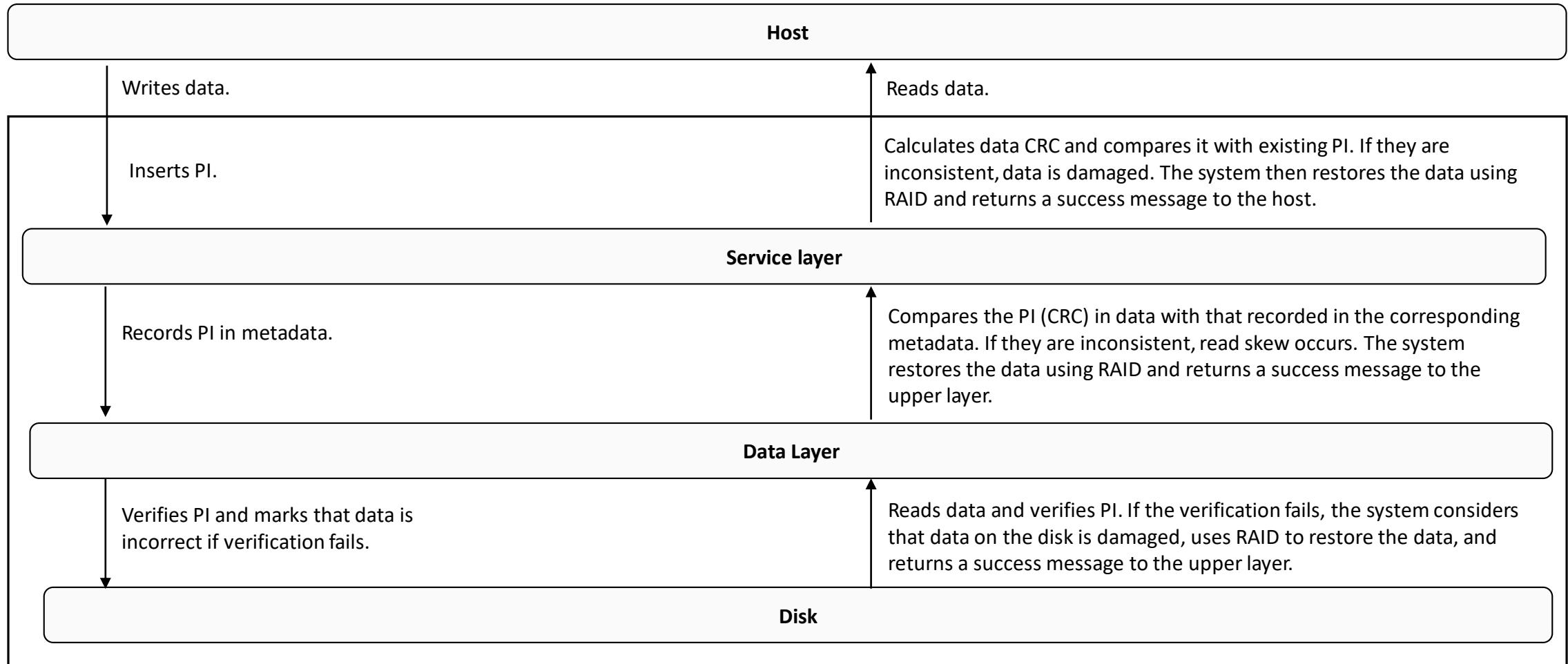
Metadata protection: Metadata is organized in a tree structure, and a parent node of metadata stores the CRC values of its child nodes, which is similar to the relationship between data and metadata. Once the metadata is damaged, it can be verified and restored using the parent and child nodes.

Solid Data Reliability: E2E Data Protection (Example)

Module Level

System Level

DC Level

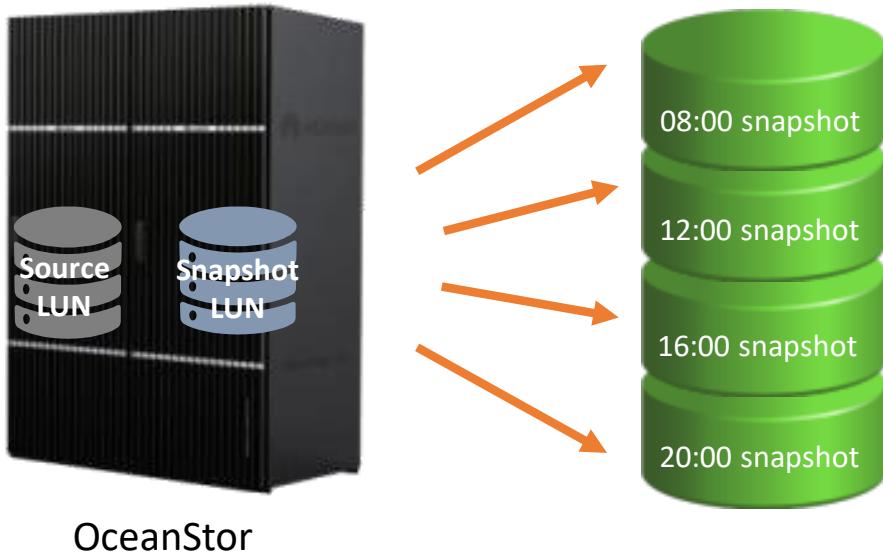


Solid Data Reliability: HyperSnap

Module Level

System Level

DC Level



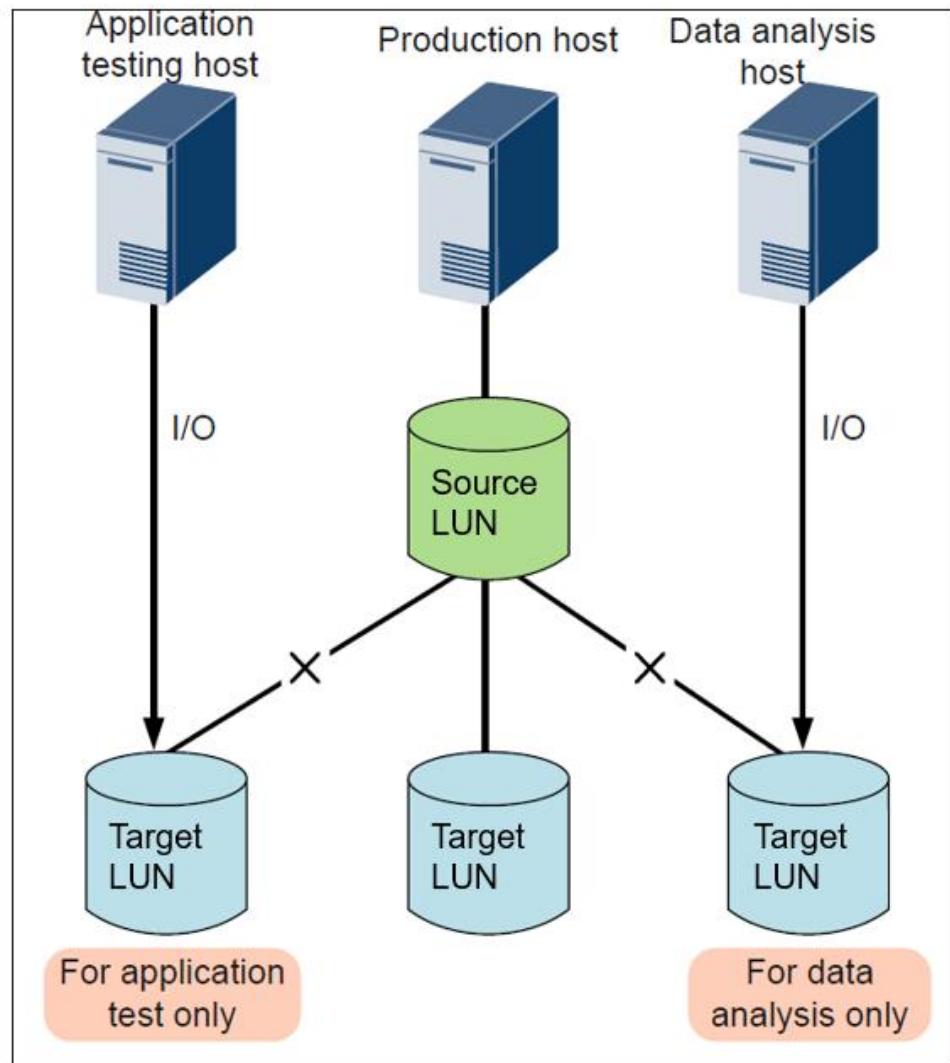
Overview

- HyperSnap quickly captures online data and generates snapshots for source data at specified points in time without interrupting system services, preventing data loss caused by viruses or misoperations. Those snapshots can be used for backup and testing.

Highlights

- The innovative [multi-time-segment cache technology](#) can continuously activate snapshots at an interval of several seconds. Activating snapshots does not block host I/Os, and host services can be quickly responded.
- Based on the RAID 2.0+ virtualization architecture, the system flexibly allocates storage space for snapshots, making resource pools unnecessary.

Solid Data Reliability: HyperClone



Module Level

System Level

DC Level

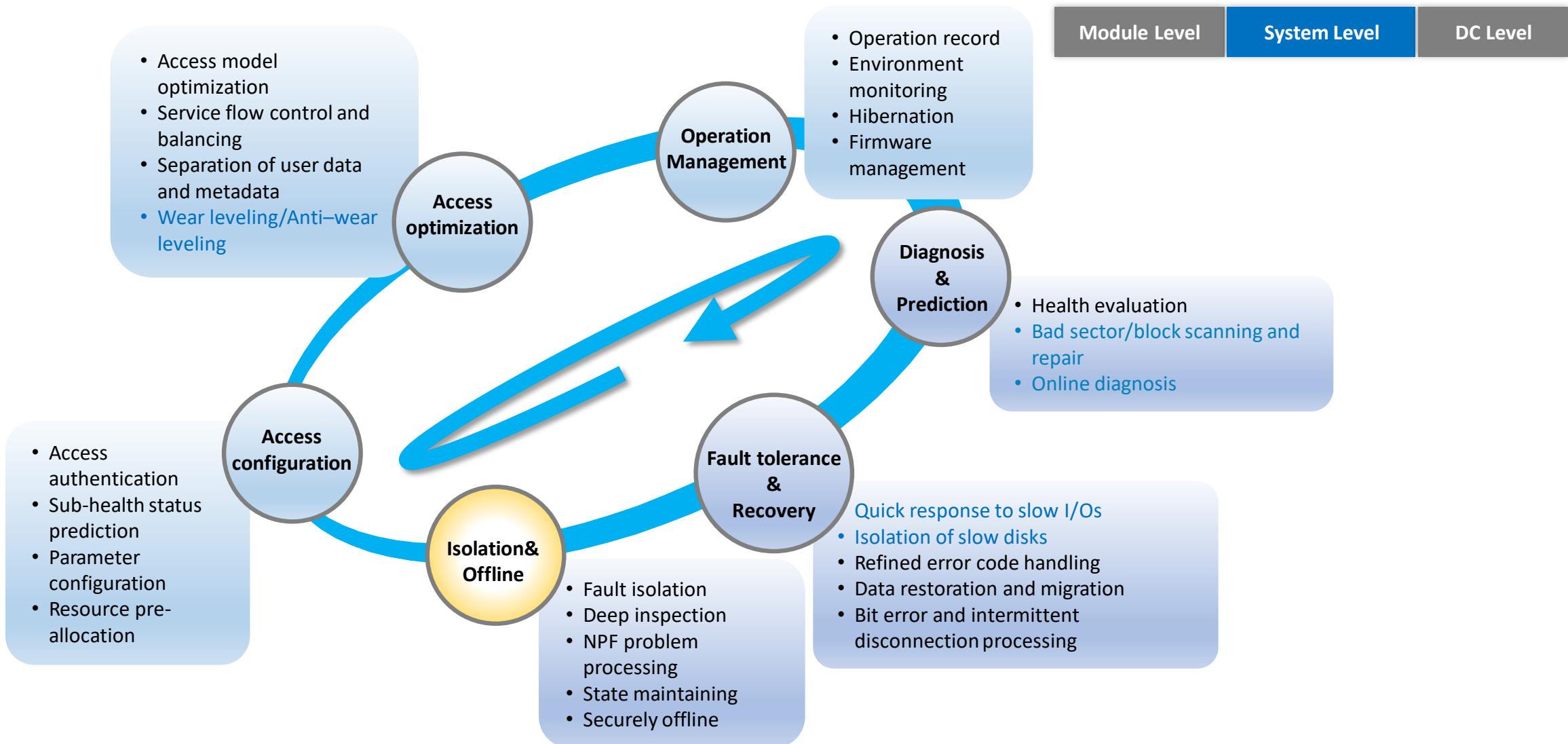
Overview

- HyperClone generates a complete physical copy (target LUN) of the production LUN (source LUN) at a point in time. The copy can be used for backup, testing, and data analysis.

Highlights

- A complete, consistent physical copy
- Isolation of source and target LUNs, eliminating mutual impact on performance
- Consistency groups, enabling the consistent splitting of multiple LUNs
- Incremental synchronization
- Reverse incremental synchronization (from the target LUN to the source LUN)

High Disk Fault Tolerance



Solid Disk Reliability: Wear Leveling and Anti-Wear Leveling

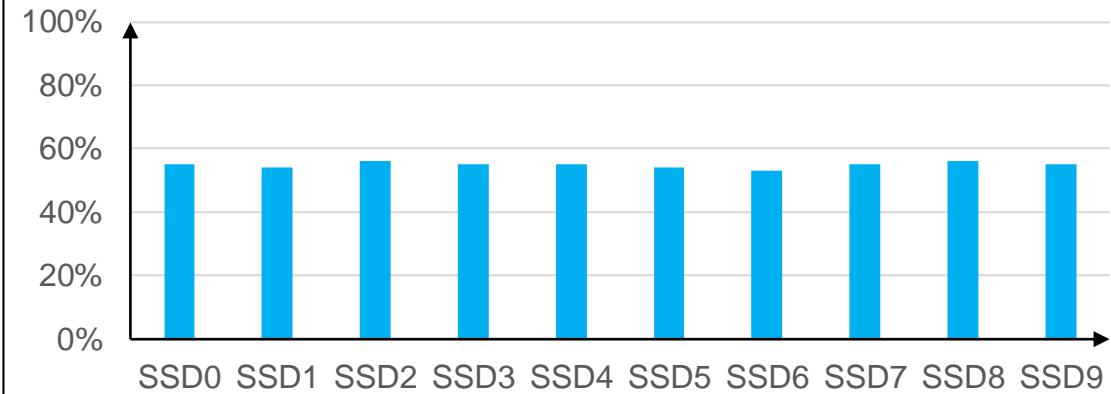
Module Level

System Level

DC Level

Global wear leveling and anti-wear leveling

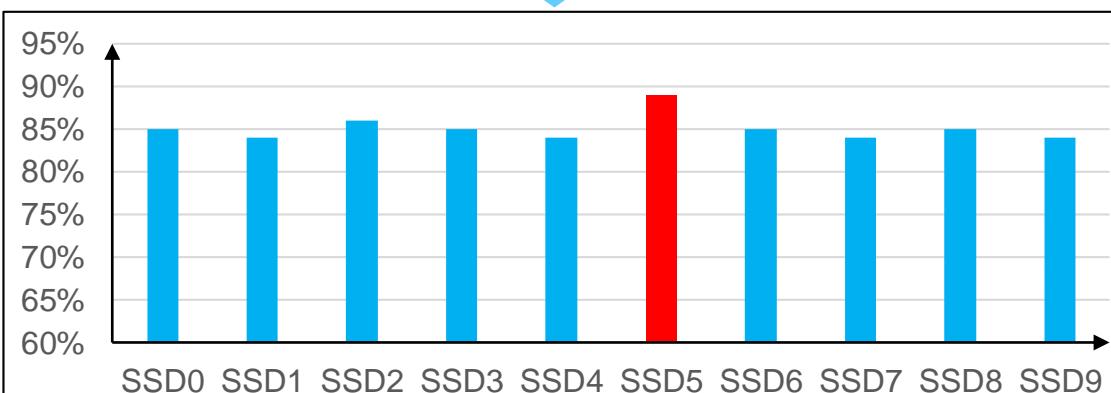
Wear leveling



Wear leveling

SSDs can only withstand a limited number of read and write operations. The system evenly distributes workloads to each SSD, preventing some disks from failing due to continuous frequent access.

Anti-wear leveling



Anti-wear leveling

To prevent simultaneous multi-SSD failures, the system starts anti-global wear leveling when detecting that the SSD wear has reached the threshold. Unbalanced data distribution on the SSDs makes their wear degrees differ by at least 2%.

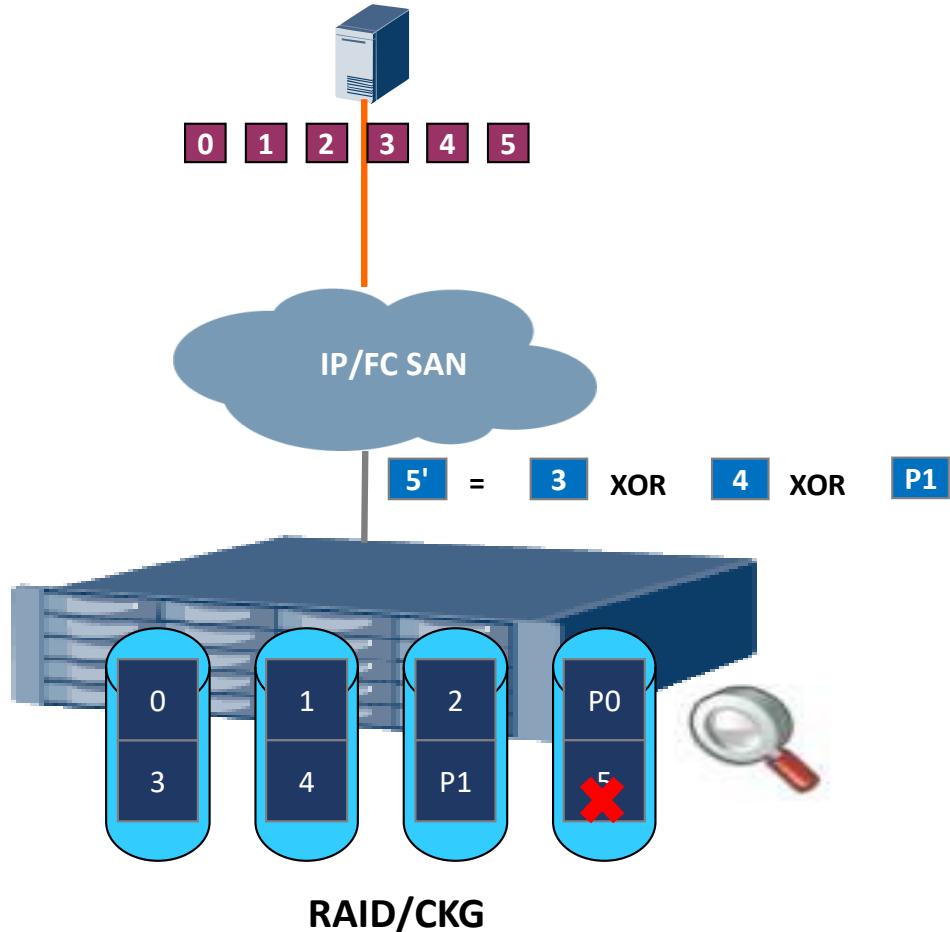
Solid Disk Reliability: Bad Sector/Block Scanning and Repair

Module Level

System Level

DC Level

Bad Sector Repair



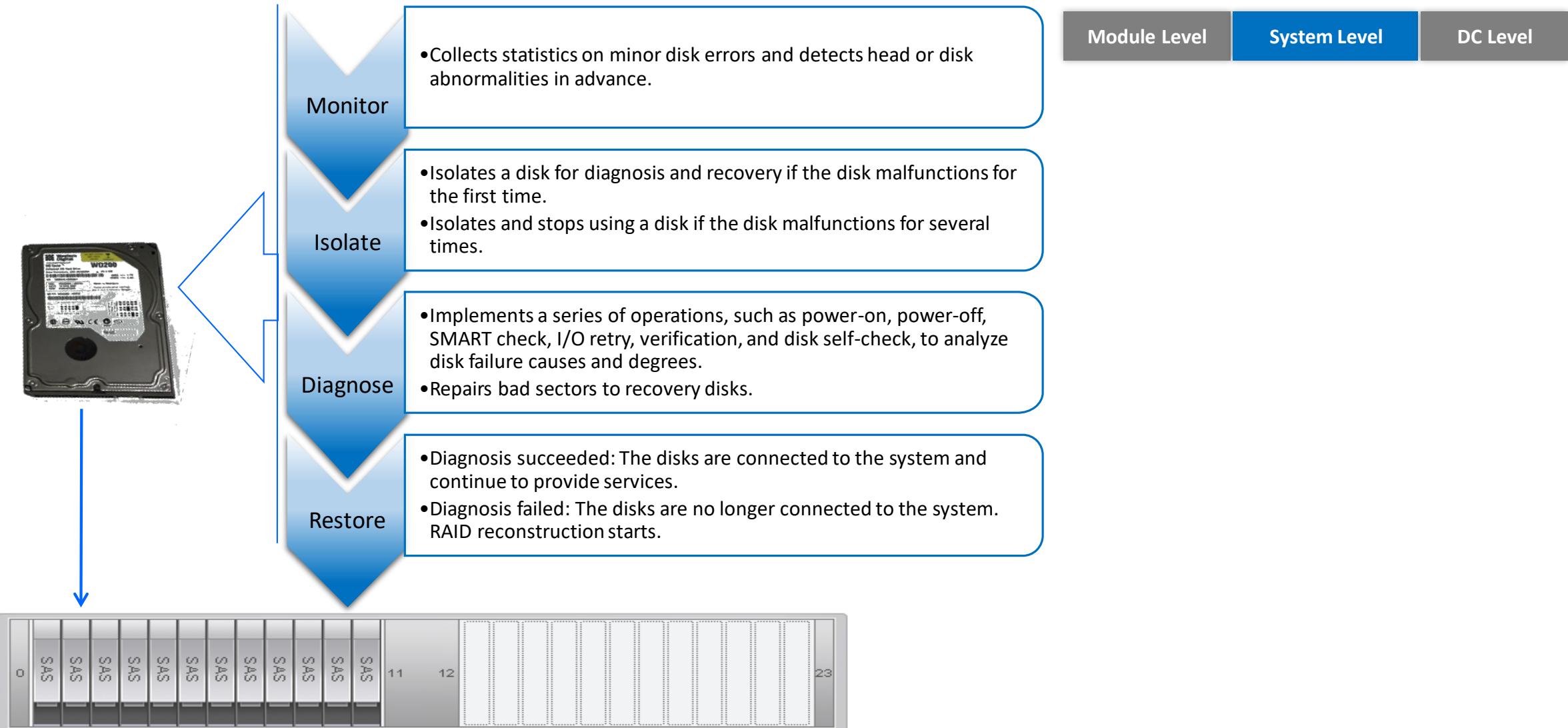
Technical Principles

- Storage systems periodically scan disks for potential faults.
- Scan algorithm: cross scan, dynamic rate adjustment
- When detecting a bad sector, the storage system employs RAID group redundancy to recover data on the bad sector and writes the recovered data to disk (disk remapping will isolate the faulty block and map the block to the internal reserved space).

Technical Highlights

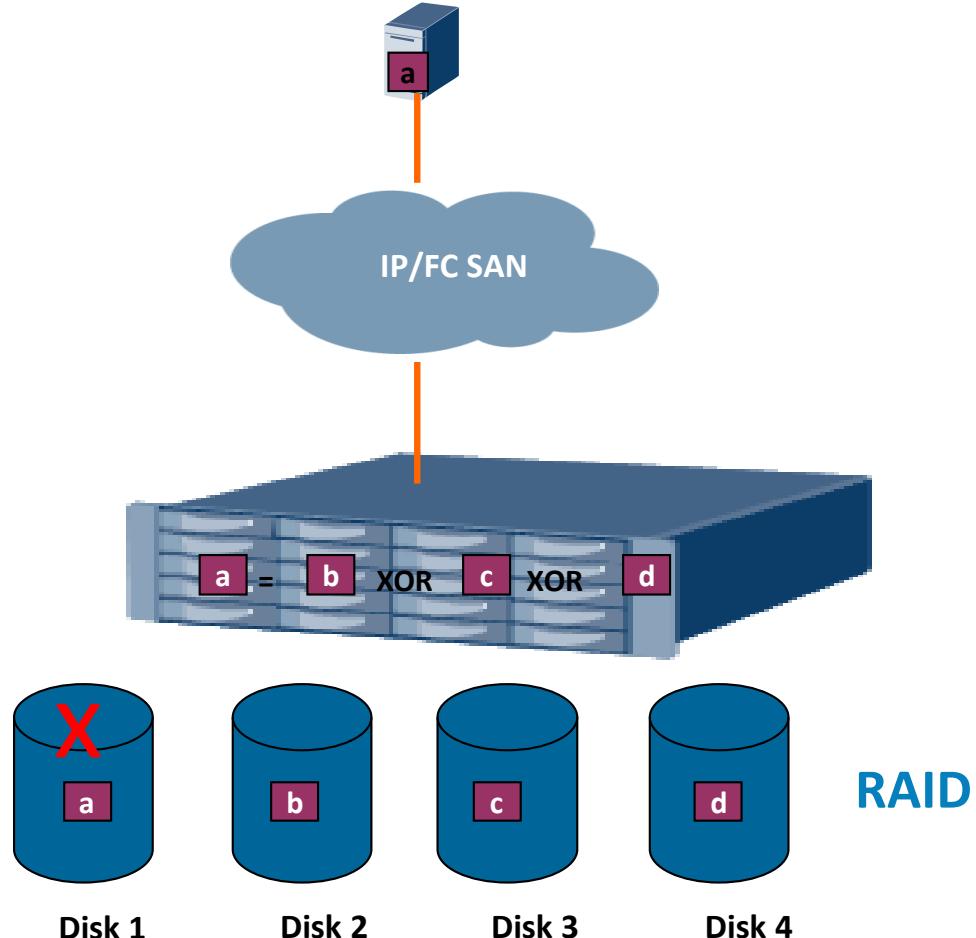
- The disk vulnerability can be detected and removed as early as possible to minimize the system risks due to disk failure.

Solid Disk Reliability: Online Diagnosis



Solid Disk Reliability: Quick Response to Slow I/Os

Quick Response to Slow I/Os



Module Level

System Level

DC Level

Technical Principles

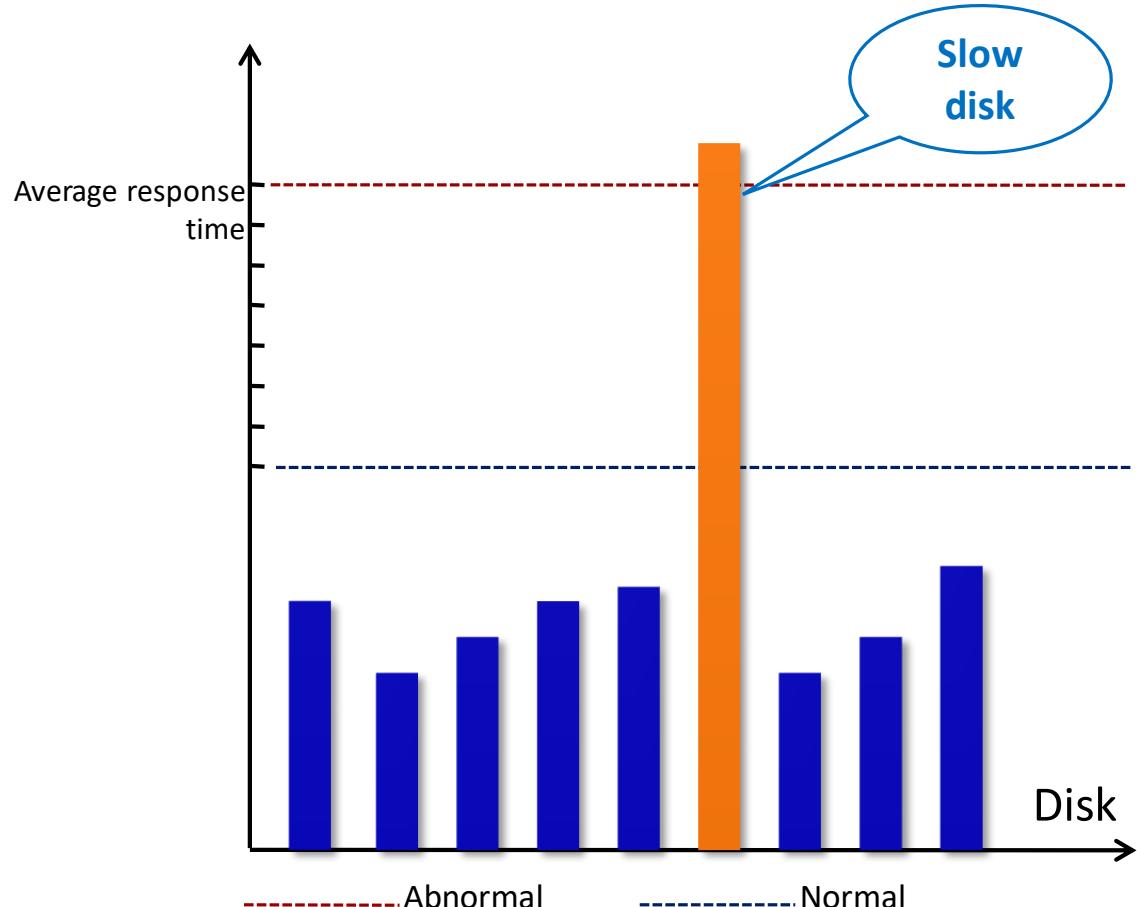
- When a read error occurs due to physical bad sectors, head problems, vibration, or electron escape, the system will handle the error by adding voltage, re-reading data, or re-mapping. As a result, the I/O response time may be prolonged.
- The system monitors the response to I/Os delivered to disks. If the response time of I/Os to a disk exceeds the upper threshold, the system stops accessing the disk and responds to host requests quickly using data restored on other RAID member disks. This disk can be accessed if no fault is found or after it recovers from a fault.

Technical Highlights

The adverse impact brought by slow disks or slow disk response within a short period can be eliminated before disks are isolated. I/O requests can be responded in a timely manner, preventing services from being affected by long retry and recovery policies.

Solid Disk Reliability: Slow Disk Detection and Isolation

Slow Disk Detection and Isolation



Module Level

System Level

DC Level

Technical Principles

- Disks are classified into domains based on disk characteristics and array attributes (such as the disk rational speed, interface type, and medium type). The average response time of all disks in a domain is used as a baseline to find out the disks that are relatively slower.
- If a disk is identified to be slower than other disks in multiple periods, the system determines that this disk is a slow disk and returns a message indicating that the disk is temporarily isolated for diagnosis. If the disk cannot be recovered, it will be permanently isolated.

Technical Highlights

- When all disks in the system become slow, the response time of a single disk cannot be much greater than that of other disks. In this case, no disk will be isolated, reducing the failure rate.
- Slow disks are identified and isolated only after diagnosis and repair.

Quiz

1. (Single-choice) Compared to other high-end flash storage(more than one engine), OceanStor high-end flash storage supports higher reliability as huawei can support
 - A. Dual controller failure without service interruption
 - B. One engine failure without interruption
 - C. Dual disk failure tolerance
 - D. Power failure protection

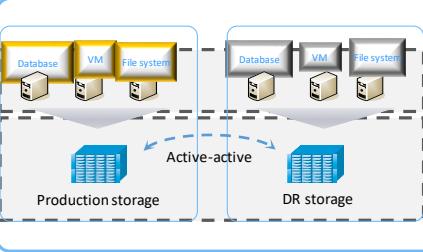
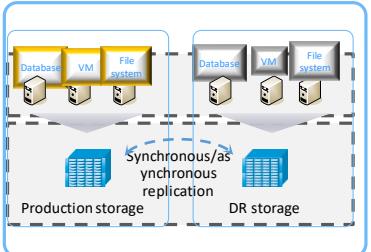
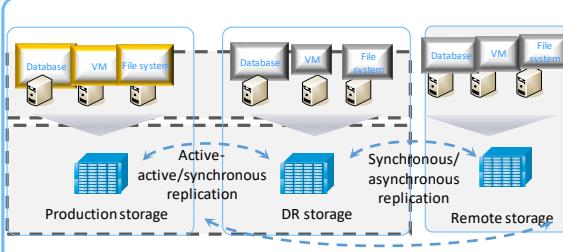
Contents

1. Storage Reliability Metrics
2. Module-Level Reliability
3. System-Level Reliability
- 4. DC-Level Reliability**
5. O&M Reliability
6. Reliability Tests and Certifications

Overview and Objectives

- This section describes the DC level reliability (disaster recovery solution) technologies.
- On completion of this section, you will be able to:
 - Describe HyperMetro solution.
 - Describe HyperReplication solution.
 - Describe 3DC solution.

Overview of Huawei DC-Level Solution

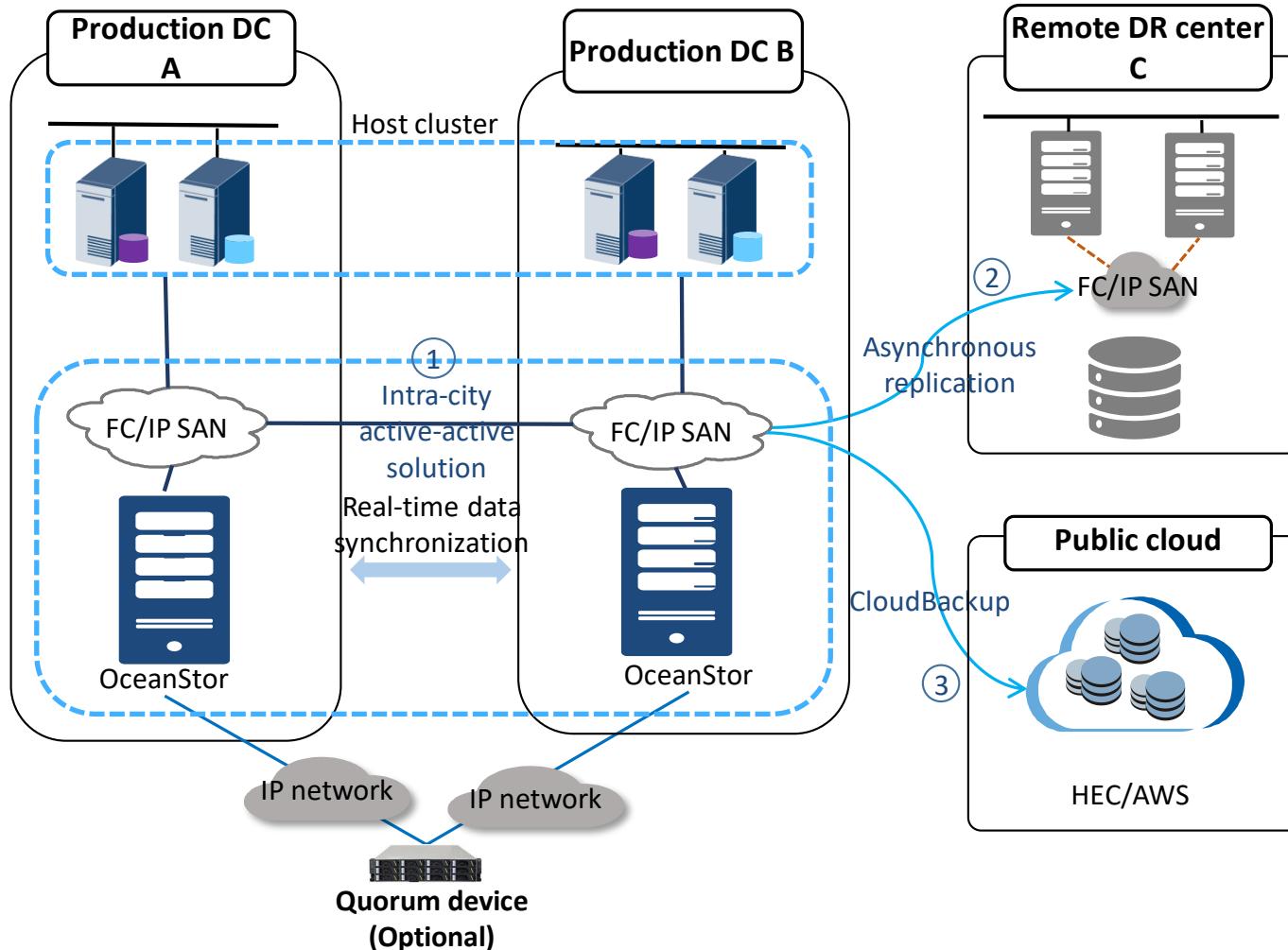
Fault Protection Scenario	Active-Active data centers solution	Active/Standby DR solution	Geo-redundant 3DC DR solution
Topology	 <p>Two data centers are shown. Each contains a Database, VM, and File system. They are connected via a dashed line labeled "Active-active". Below each data center is a Production storage unit and a DR storage unit.</p>	 <p>Two data centers are shown. Each contains a Database, VM, and File system. They are connected via a dashed line labeled "Synchronous/asynchronous replication". Below each data center is a Production storage unit and a DR storage unit.</p>	 <p>Three data centers are shown in a horizontal chain. Each contains a Database, VM, and File system. They are connected via dashed lines labeled "Active-active/synchronous replication" and "Synchronous/asynchronous replication". Below the first data center is a Production storage unit, between the first and second is a DR storage unit, and below the third is a Remote storage unit. The label "DR Star trio" is at the bottom right.</p>
DR Distance	<ul style="list-style-type: none"> Same city (< 300 km) 	<ul style="list-style-type: none"> Intra-city or remote data center (synchronous replication: < 300 km; asynchronous replication: < 3000 km) 	<ul style="list-style-type: none"> Intra-city or remote data center (synchronous replication: < 300 km; asynchronous replication: < 3000 km)
RPO	<ul style="list-style-type: none"> 0 (intra-city active-active) 	<ul style="list-style-type: none"> RPO = 0 (synchronous) or RPO = minutes (asynchronous) 	<ul style="list-style-type: none"> RPO = 0 (active-active/synchronous) or RPO = minutes (asynchronous)
RTO	<ul style="list-style-type: none"> 0 	<ul style="list-style-type: none"> Minutes 	<ul style="list-style-type: none"> Minutes or hours
Cost	<ul style="list-style-type: none"> High 	<ul style="list-style-type: none"> Medium 	<ul style="list-style-type: none"> High

DC-Level Reliability

Module Level

System Level

DC Level



Inter-system Reliability Solution

1. HyperMetro

- **Active-active architecture:** Active-active LUNs are readable and writable in both DCs and data is synchronized in real time.
- **High reliability:** Cross-site bad block repair improves system reliability.
- **Elastic scalability:** expanded to the **3DC solution** based on the remote replication solution.

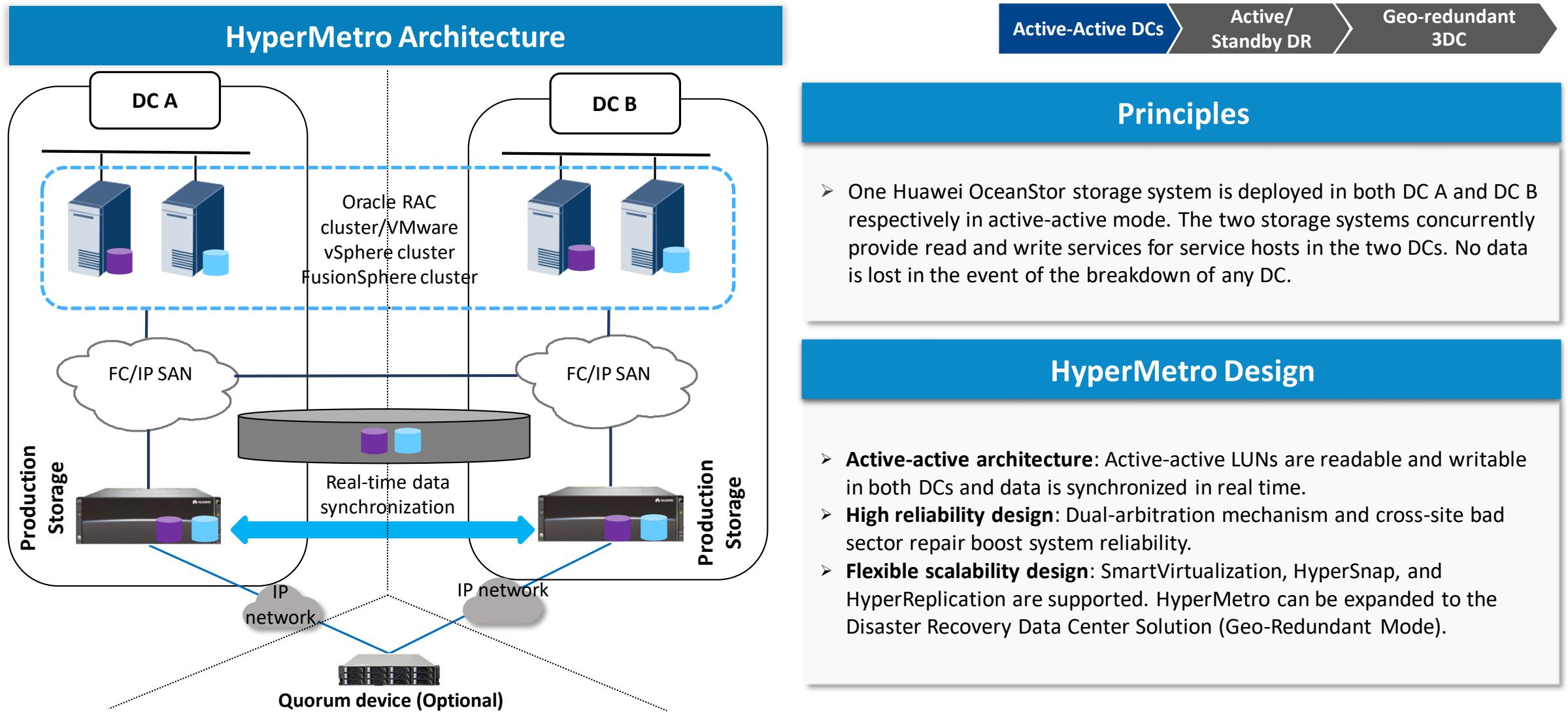
2. HyperReplication:

Synchronous and asynchronous data replication across storage systems provides intra-city and remote data protection.

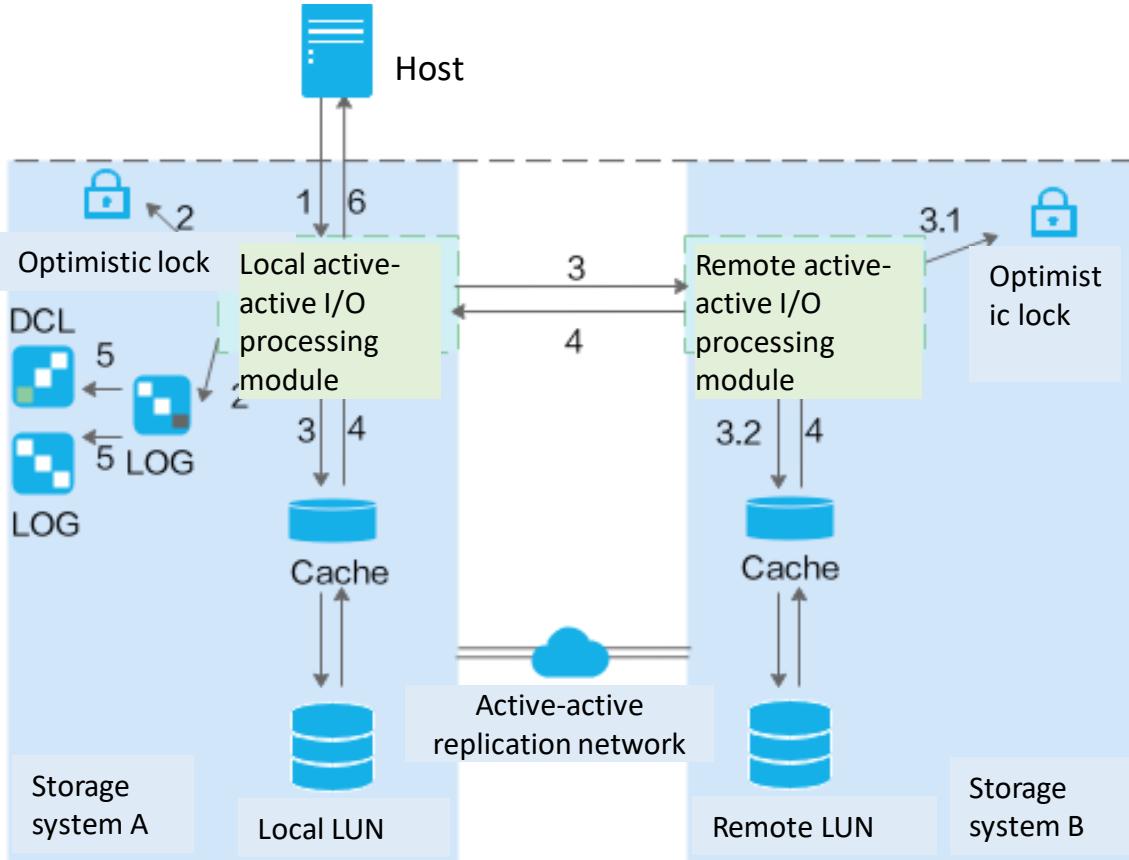
Cloud Backup Solution

- ### 3. CloudBackup:
- The public cloud object storage is used as the backup storage to prevent data loss caused by faults on storage devices in the enterprise DC.

Solution-Level Reliability: HyperMetro



HyperMetro Dual-Write Process



Active-Active DCs

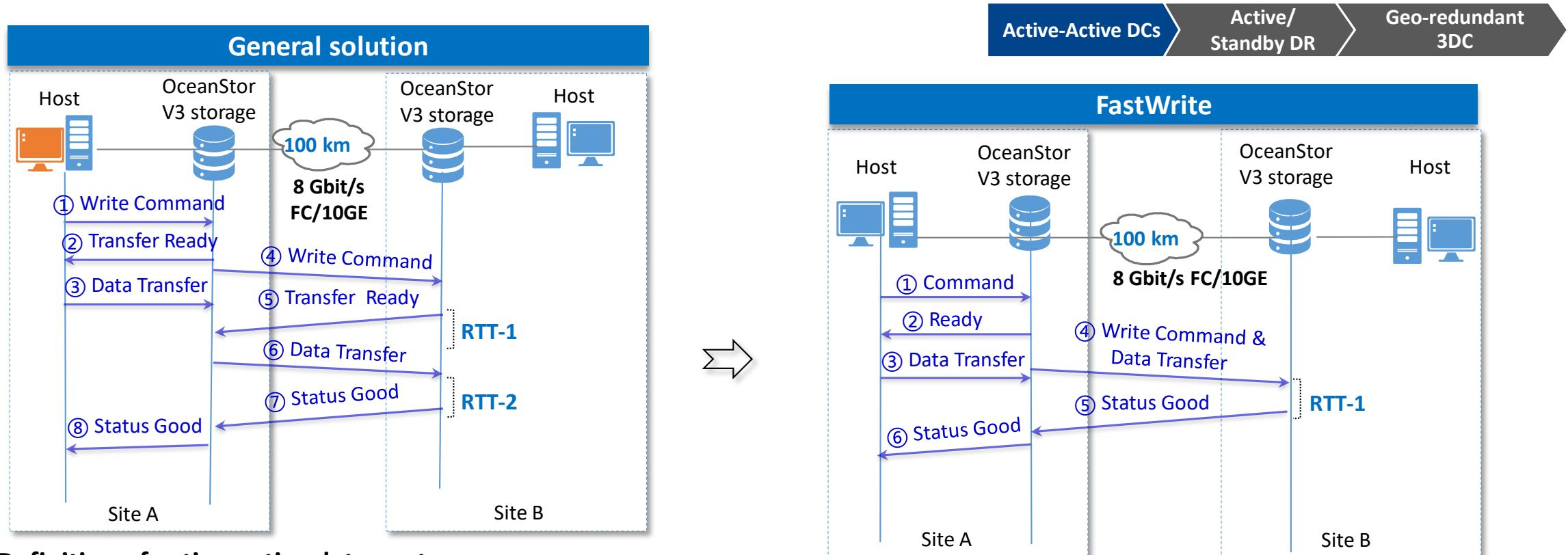
Active/
Standby DR

Geo-redundant
3DC

HyperMetro Dual-Write Process

1. A host delivers a write I/O request to the HyperMetro management module.
2. The local storage system applies for write permission to the local optimistic lock. After the write permission is obtained, a log is recorded. The log records only the address information but no data content.
3. The HyperMetro management module writes data to the caches of both the local and remote LUNs concurrently. The remote storage system upon receiving the write request also applies for write permission to the optimistic lock. After the write permission is obtained, the remote storage system writes data to the cache.
4. Both the local and remote caches return the write I/O result to the HyperMetro management module.
5. Perform the following operations based on the result of step 4:
 - If data is successfully written to both the local and remote storage systems, the log is deleted.
 - If writing data to either storage system fails, the log is converted into a DCL that records the differential data between the local and remote LUNs.The HyperMetro pair is split. The status of the HyperMetro pair becomes **To be synchronized**. I/Os are only written to the storage system on which writing data to its cache succeeded. The storage system on which writing data to its cache failed stops providing services for hosts.

FastWrite Optimizes Dual-Write Performance



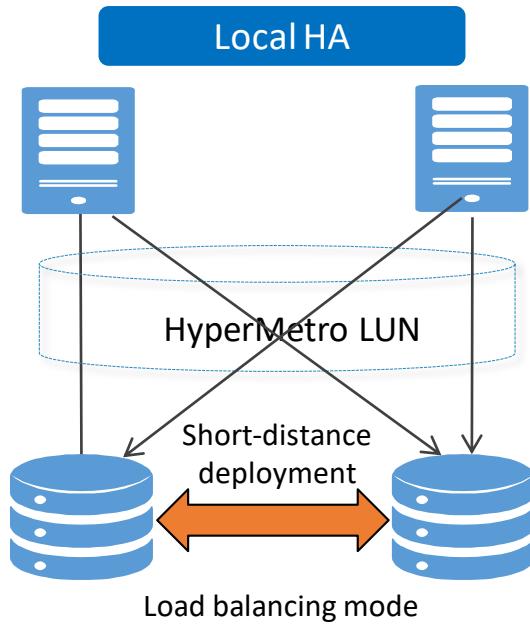
Definition of active-active data centers

- General solution: Write I/Os experience two interactions at two sites: write command and data transfer.
- Two RTTs over 100 km transmission links.

- FastWrite: optimizes the protocol to combine write command and data transfer into one interaction, **reducing cross-site write I/O interactions by half**.
- Only **one RTT** over 100 km transmission links, **improving service performance by 30%**.

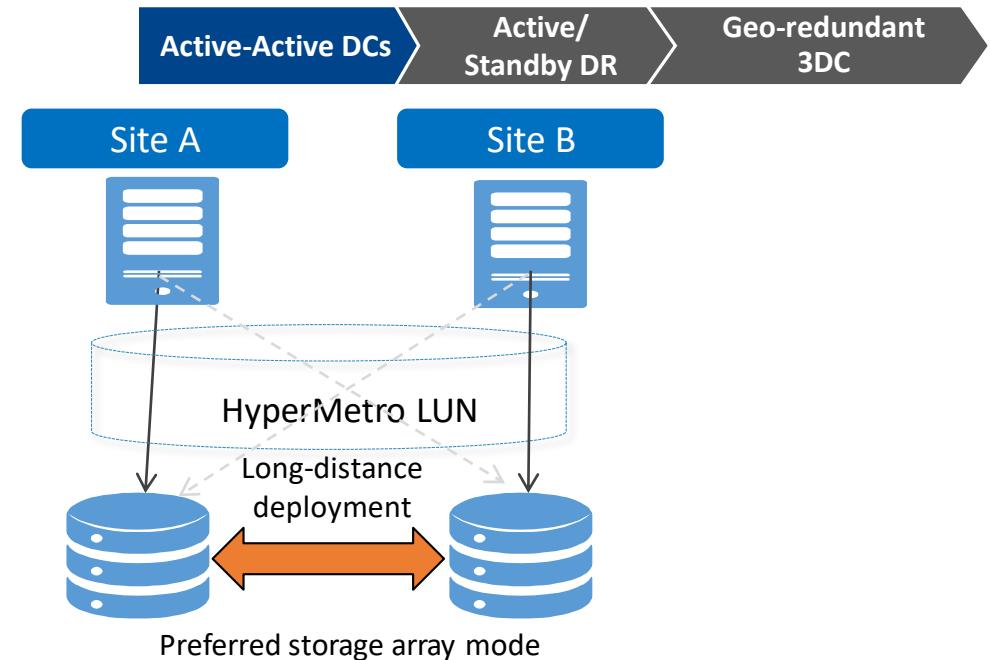
Tips: It is recommended that FastWrite be enabled when the distance between the two data centers exceeds 10 km.

Multipathing Routing Algorithm Optimization: Host Data Access Optimization



Load balancing mode (applicable to local HA scenarios)

- Cross-array I/O load balancing
- Short-distance deployment (similar to intra-equipment room deployment)
- I/Os are delivered to two storage arrays in load balancing mode, optimizing storage utilization and improving performance.

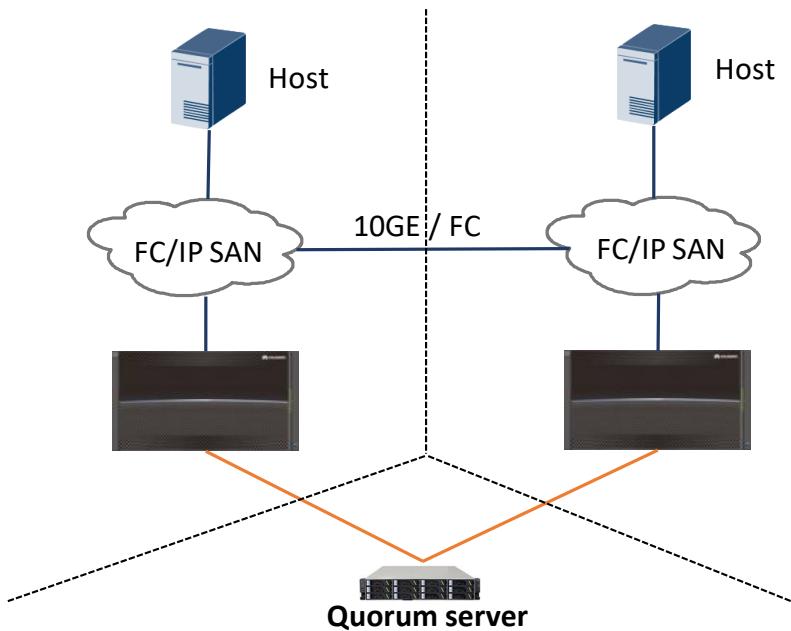


Preferred storage array mode (applicable to intra-city active-active scenarios)

- Reduced cross-site accesses and transfer latency
- Long-distance deployment
- It is set in Huawei UltraPath that the hosts at site A preferentially access the storage array at site A and the hosts at site B preferentially access the storage array at site B first. I/Os are only delivered to the preferred storage array.

Note: Huawei UltraPath is recommended in HyperMetro solution.

HyperMetro Networking Design



Tips:

1. No matter a HyperMetro replication network is an IP network or a Fibre Channel network, DCs must be interconnected over bare fibers.
2. If a HyperMetro replication network is a Fibre Channel network and the distance is less than or equal to 25 km, bare fibers can be used for direct connection. At least two pairs (four cores) of Fibre Channel networks must be configured at the storage layer. If the distance is greater than or equal to 25 km, DWDM devices must be used to interconnect data centers.
3. If a HyperMetro replication network is an IP network and the distance is less than or equal to 80 km, bare fibers can be used for direct connection. When core switches are used, at least two pairs (four cores) of Fibre Channel networks must be configured to connect core switches. If the distance is greater than or equal to 80 km, DWDM devices must be used to connect data centers.



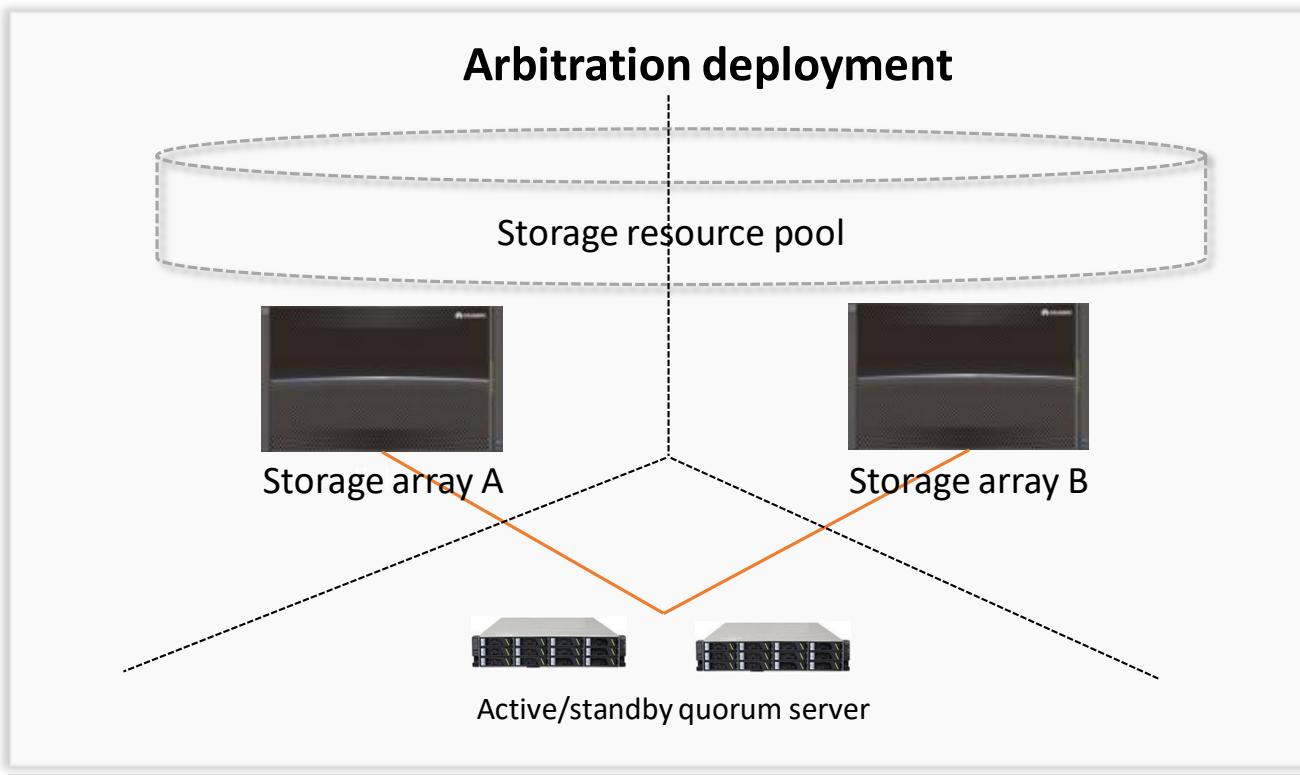
Intra-city interconnection and quorum network design

- **Network interconnection requirements**
 - Supports 10 Gbit/s, 25 Gbit/s, 40 Gbit/s, 100 Gbit/s, 8 Gbit/s Fibre Channel, 16 Gbit/s Fibre Channel, and 32 Gbit/s Fibre Channel interconnection protocols.
- **SLA design for intra-city links**
 - Link reuse (including heartbeat interconnection, HyperMetro, and replication) and simple network
 - Preferred transmission: Heartbeat > HyperMetro and synchronous replication I/O flow > asynchronous replication I/O flow
- **Quorum link design**
 - Supports 10GE and GE networks, over 2 Mbit/s bandwidth, and reachable IP addresses

Best Practices

- **Networking Rules**
 - HyperMetro intra-city interconnection network is consistent with the interconnection network of hosts and storage arrays, simplifying the network topology.
 - Storage intra-city interconnection ports are not reused as front-end host ports.

HyperMetro Arbitration



- **Quorum server:** physical server or virtual server
- **Quorum link:** IP addresses must be reachable.
- **Mechanism:** The storage array that wins the arbitration continues providing services and the storage array that loses the arbitration stops providing services.
- **Arbitration mode:** static priority mode and quorum server mode
- **Quorum granularity:** service (pair or consistency group)



Third-place quorum site (recommended)

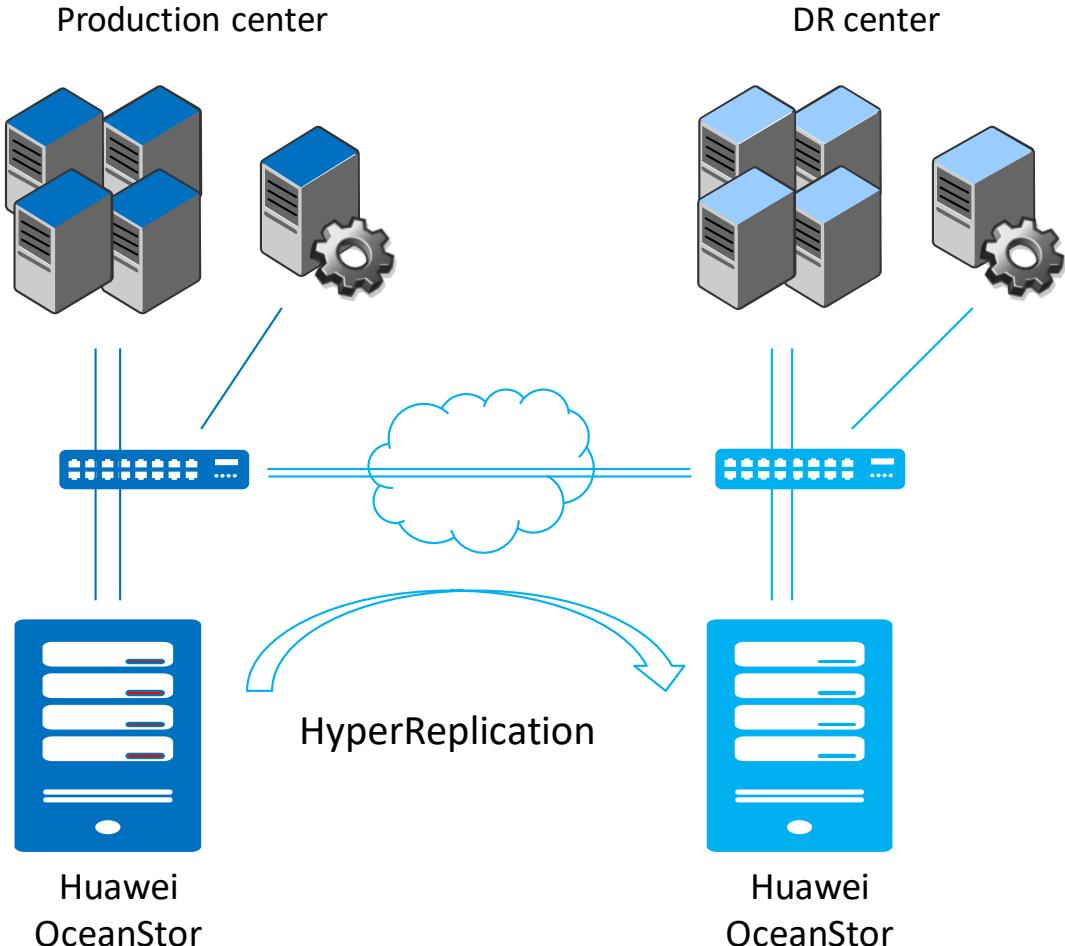
- The quorum server is deployed at the third-place quorum site and is within a domain different from the fault domains of the two active-active data centers.
- If the active-active solution is recommended, you are advised to deploy one or two quorum servers. The two quorum servers work in active/standby mode.

No third-place quorum site

- **Preferred:** The quorum server is deployed at the preferred site and an independent UPS is configured.
- **Alternative:** Do not deploy a quorum server. Instead, set the static priority between sites. (If the preferred site fails, services are stopped.) This mode must be approved by the customer in writing before implementation.

Solution-Level Reliability: HyperReplication

Active-Active DCs Active/Standby DR Geo-redundant 3DC



Overview

- HyperReplication supports both synchronous and asynchronous remote replication between storage systems. It is used in disaster recovery solutions to provide intra-city and remote data protection, preventing data loss caused by disasters and improving business continuity.

Highlights

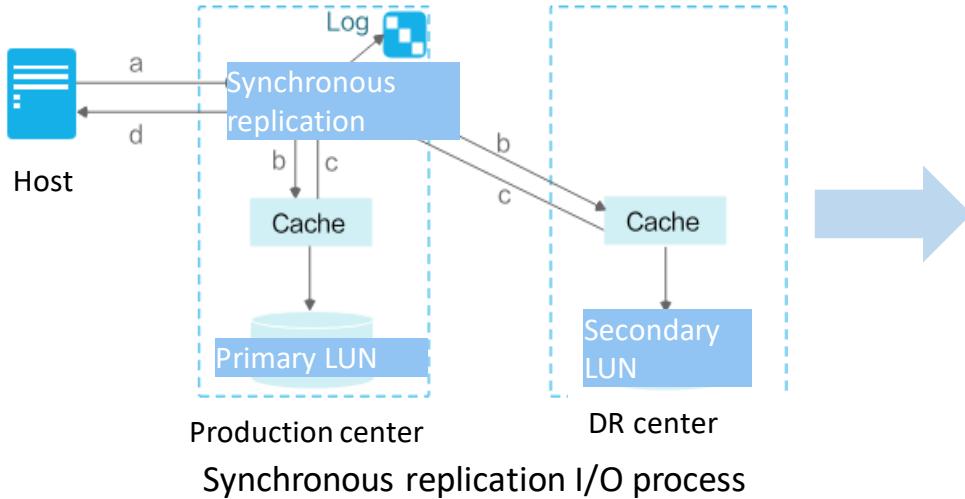
- Remote replication interworking among entry-level, mid-range, and high-end storage systems
- RPO within seconds for asynchronous replication and zero RPO for synchronous replication
- Consistency group
- Incremental synchronization
- Fibre Channel and IP links
- Network: bi-directional replication, 1:N, and N:1
- 3DC: cascading replication and parallel replication

Differences Between Synchronous Replication and Asynchronous Replication

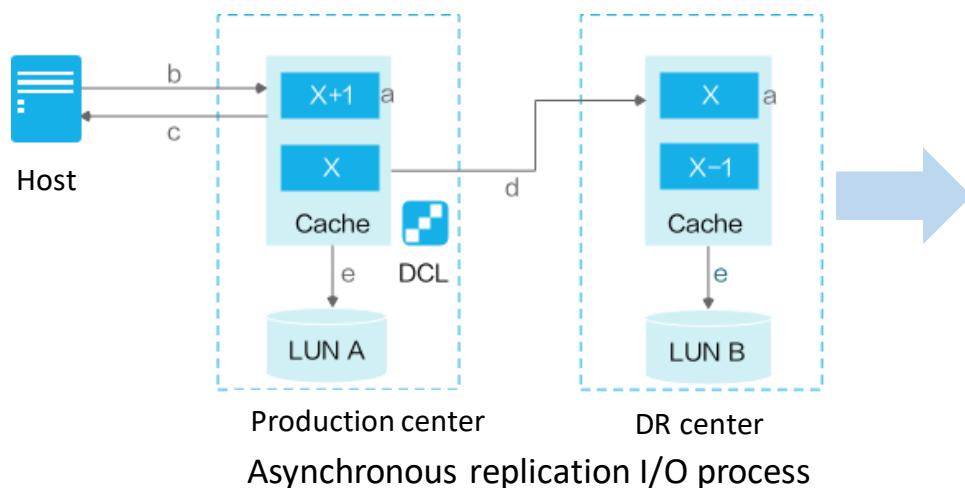


- A typical DR system consists of a production center and a DR center. The two centers can be deployed in the same city or in two cities. The remote replication mode varies depending on the distance between the two centers.
- **Remote synchronous replication (HyperReplication/S)**
- If the two centers are deployed in the same city, the transmission latency is low. HyperReplication/S is recommended. When a host writes data to the production storage array, a write success message is returned to the host only when the data is successfully written to both the production and DR storage arrays.
- **Remote asynchronous replication (HyperReplication/A)**
- If the two centers are deployed in two cities, the transmission latency is high. HyperReplication/A is recommended. When a host writes data to the production storage array, a write success message is returned to the host when the data is successfully written to the production storage array and then the data is replicated from the production storage array to the DR storage array.
- HyperReplication/A reduces service impact in long-distance deployment scenarios, but will result in certain data loss.

Synchronous and Asynchronous Replication I/O Process



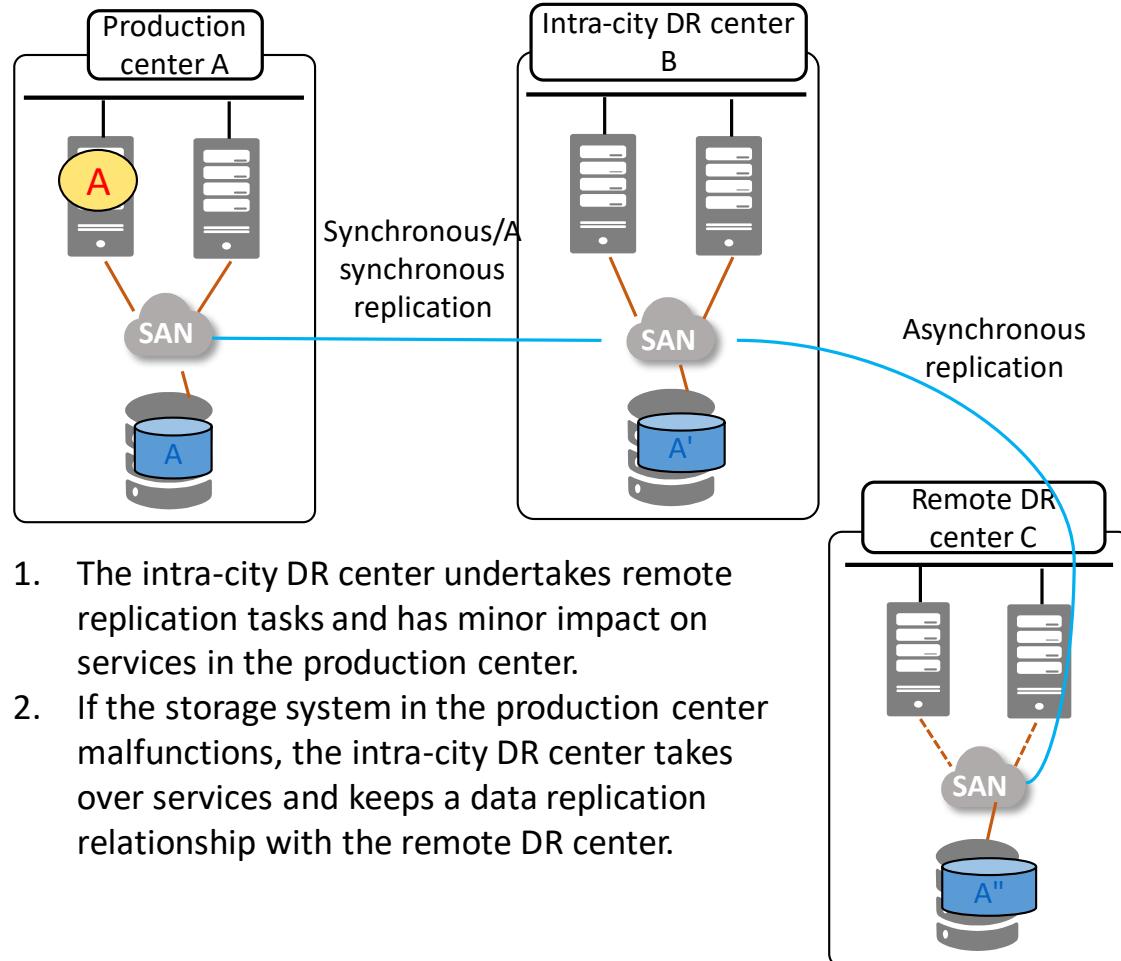
- The production storage receives a write request from the host. HyperReplication records the request in a log. The log only records the address information but no data content.
- The request is written to the primary and secondary LUNs. In write-back mode, data is written to the cache.
- HyperReplication waits for the write results from both the primary and secondary LUNs. If the write to the secondary LUN times out or fails, the remote replication relationship is disconnected. If data is successfully written to both primary and secondary LUNs, the log is cleared. Otherwise, the log is retained and written into the Data Change Log (DCL) on the disk. In this case, the synchronization is interrupted abnormally. When the synchronization is started later, the data block corresponding to the log address is replicated again.
- The write result of the primary LUN is returned to the host.



- When a remote asynchronous replication synchronization task is started in each period, snapshots of the primary and secondary LUNs are created and the points in time are updated (primary LUN snapshot at the point in time X and secondary LUN snapshot at the point in time X - 1).
- When a host writes new data, the data is cached at the point in time X + 1 in the primary LUN cache.
- The host receives a write success message.
- Data on the primary LUN at the point in time X is replicated to the secondary LUN based on the DCL.
- The primary and secondary LUNs flush the received data to disks. After the synchronization is complete, the latest data on the secondary LUN is the complete data on the primary LUN at the point in time X.

Solution-Level Reliability: 3DC DR

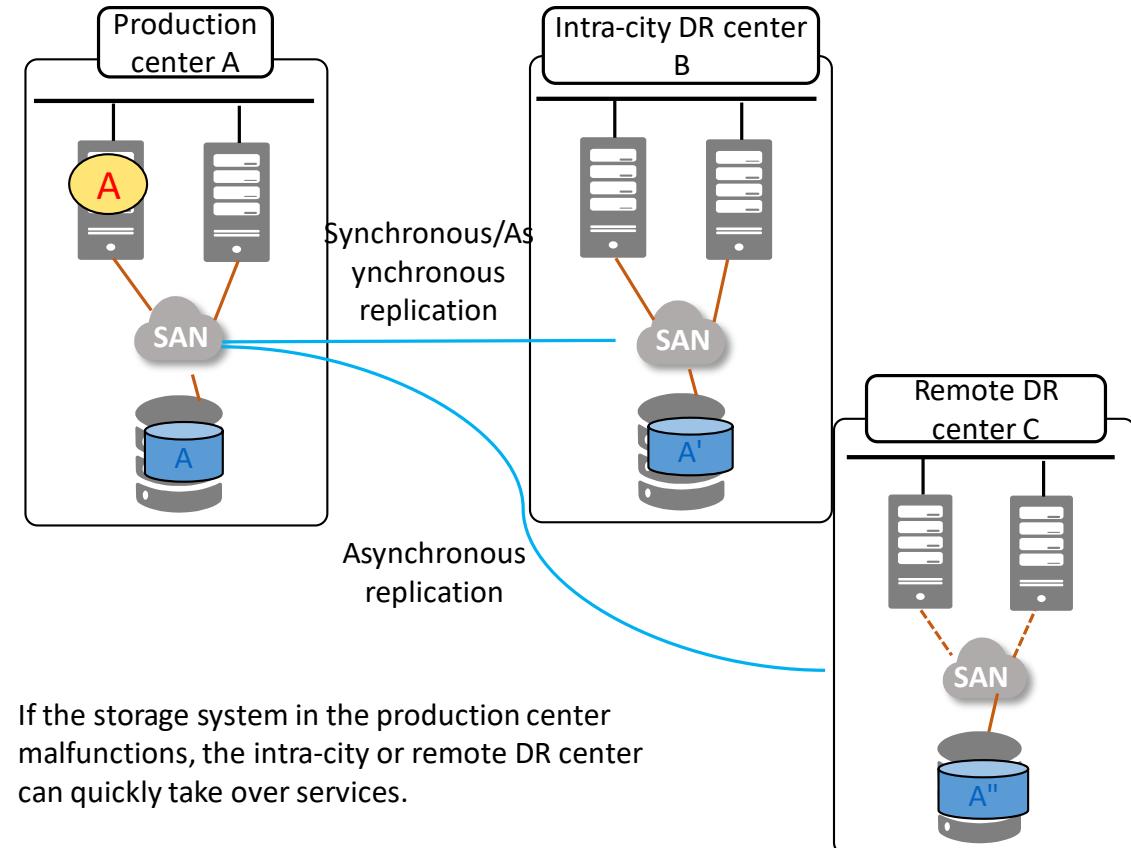
Cascaded Architecture



1. The intra-city DR center undertakes remote replication tasks and has minor impact on services in the production center.
2. If the storage system in the production center malfunctions, the intra-city DR center takes over services and keeps a data replication relationship with the remote DR center.



Parallel Architecture



Quiz

1. (Multiple-choice) What are the differences between HyperMetro and HyperReplication?
 - A. Both HyperMetro and HyperReplication can get RPO=0
 - B. Both HyperMetro and HyperReplication can get RTO almost zero.
 - C. Both HyperMetro and HyperReplication can support dual active-active host cluster.
 - D. Only HyperMetro can support active-active cluster.

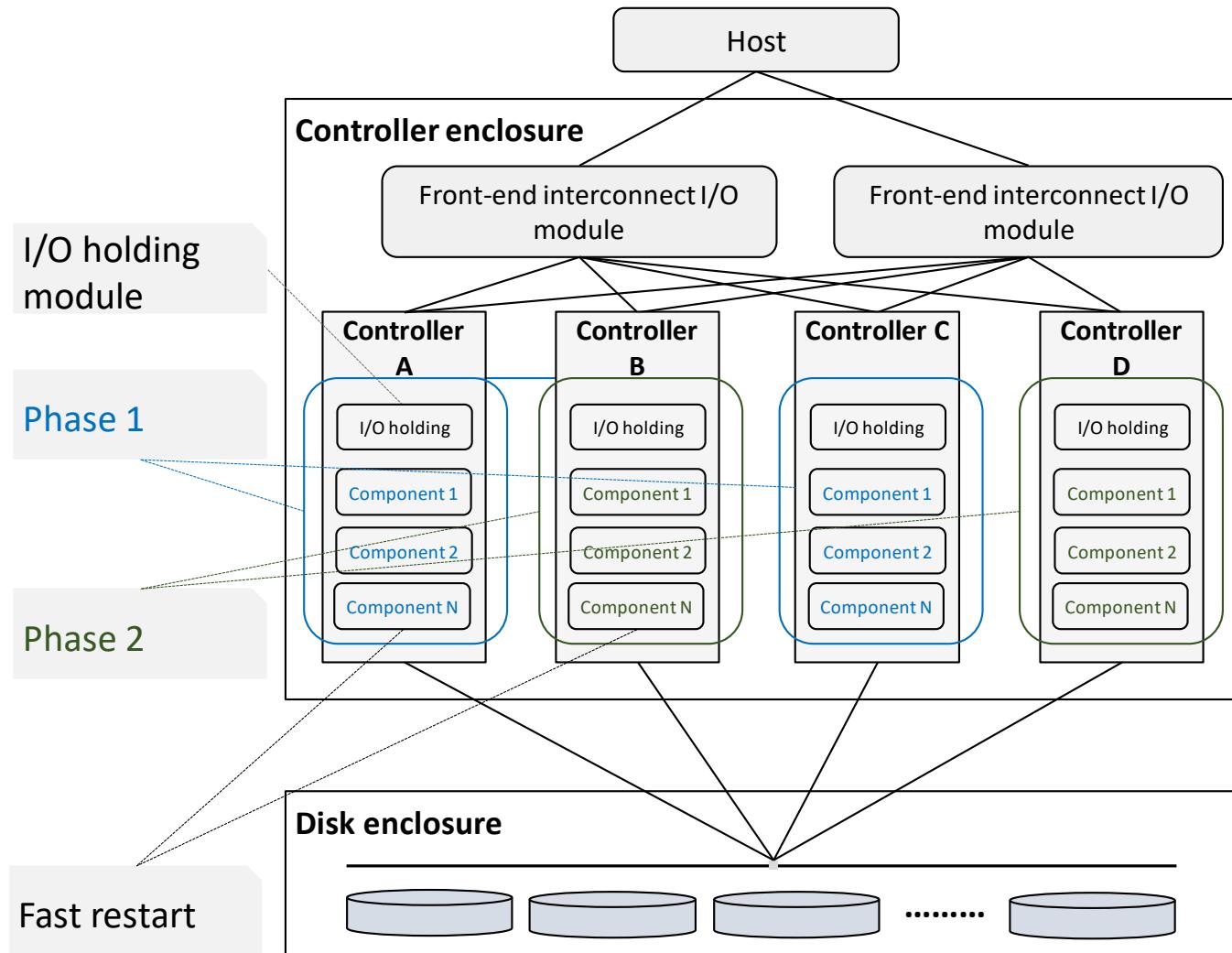
Contents

1. Storage Reliability Metrics
2. Module-Level Reliability
3. System-Level Reliability
4. DC-Level Reliability
- 5. O&M Reliability**
6. Reliability Tests and Certifications

Overview and Objectives

- This section describes O&M reliability features.
- On completion of this section, you will be able to:
 - Describe the NDU capability of OceanStor series.
 - Describe capability of intelligent O&M (DME IQ)

O&M Reliability: Fast Upgrade



Upgrade Transparent to Hosts

- Host unaware of upgrade**: Each controller has an I/O holding module, which holds current I/Os during component upgrade and restart. After being upgraded, components continue to process the I/Os so that hosts do not detect any connection interruption or I/O exception.
- Component upgrade**: The system upgrade is divided into two phases. The software components (processes) with redundant units are upgraded first. After the software packages are uploaded and the processes are restarted, the second phase is triggered.
- Zero performance loss**: Each software component restarts with 1s. The front-end interface module returns **BUSY** for failed I/Os during the upgrade. The host re-delivers the failed I/Os, and the performance is restored to 100% within 2 seconds.
- Short upgrade duration**: No host compatibility issue is involved. Host information does not need to be collected for evaluation. The entire storage system can be upgraded within 10 minutes as controllers do not need to be restarted.

O&M Reliability: Intelligent Prediction



Elimination of potential risks, improving reliability.



- Predict disk risks **14 day** ahead.
- Use the **XGBoost** algorithm to identify **80%** disk risks with only a **0.1%** mis-reporting rate.

Disk risk prediction



Massive data analysis, building mature risk disk prediction models



- Analyze **500,000+** disks and **20+ billion** feature records.
- Verify enterprise data center models for over **600 days**.

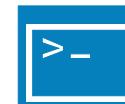
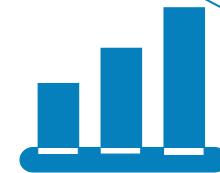


Capacity monitoring, identifying overloaded resources in advance



- Predict the capacity trend **in the next 12 months** and determine the capacity requirements.
- Predict capacity consumption and **identify overloaded resources**.

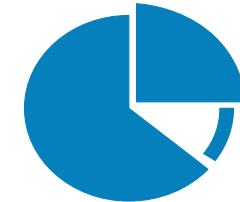
Service trend prediction



Intelligent capacity planning and on-demand procurement, reducing TCO



- Optimize idle resources to **improve resource utilization**.



- Evaluate the capacity requirements and provide detailed **expansion solutions**.

Contents

1. Storage Reliability Metrics
2. Module-Level Reliability
3. System-Level Reliability
4. Solution-Level Reliability
5. O&M Reliability
- 6. Reliability Tests and Certifications**

Reliability Tests and Certifications

Hardware Tests and Certifications



Type	Purpose	Authentication Result
EMC test	Ensure that the product meets the requirements of the corresponding EMC standards, real-world electromagnetic environment, and device or system compatibility.	Met the mandatory admission certification requirements of each region/country and organization: <ul style="list-style-type: none">China: CCCEuropean Commission: CEJapan: VCCI-ARussia: CU...
Safety test	Ensure the personal safety when using the products, reduce the injury caused by electric shock, fire, heat, mechanical damage, radiation, chemical damage, and energy, and meet the admission requirements of each country.	
Environment (climatic) test	Check that the products meet the requirements. Expose defects in design, process, and materials.	
Environmental (mechanical) test	Improve the environment adaptability of the products to mechanical stress during storage and transportation to ensure qualified product appearance, structure, and performance and ensure that the product can withstand the adverse impact caused by external mechanical stress on the equipment.	Passed some optional certifications, such as China's: <ul style="list-style-type: none">Earthquake resistance certificationEnvironmental Labeling certification
HALT test	Find the weak points of the products and improve product reliability.	

Ecosystem Compatibility Certifications

- After 10+ years of technical accumulation and continuous investment, Huawei has established the largest storage interoperability lab in Asia. The lab provides 10,000+ pages of interoperability lists and 1 million+ verified service scenarios, covering 4000+ software and hardware versions of mainstream applications, operating systems, virtualization products, servers, and switches. Huawei has cooperated with mainstream vendors and earned 1500+ certificates. Huawei products are compatible with mainstream vendors' new products as soon as the new products are launched.
- Huawei is the storage vendor that has the most upstream and downstream partner resources. Huawei is the top strategic partner of Seagate, Western Digital, Intel, SAP, and Microsoft.

Summary

Key Reliability Technologies of OceanStor Flash Storage

- **High Service Availability (Tolerate 7 out of 8 controllers failures)**
Controller Failover within Seconds, Continuous Mirroring, HyperMetro-Inner
- **Solid Data Reliability**
Multiple Cache Copies, RAID 2.0+, E2E Data Protection
Fast Reconstruction, Reconstruction Offloading, Dynamic Reconstruction
- **High Disk Fault Tolerance**
Wear Leveling/Anti-wear Leveling, Bad Sector/Block Scanning and Repair, Quick Response to Slow I/Os, Intra-disk RAID
- **Disaster Recovery Solution**
HyperSnap, HyperClone, HyperReplication, 3DC(Cascaded and Parallel Architecture)
- **O&M Reliability**
Fast Upgrade

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 OceanStor Pacific Distributed Storage Architecture and Key Technologies



Foreword

- Huawei OceanStor Pacific is an intelligent distributed storage series with scale-out capability designed to support the business needs of today and tomorrow. It features a wide range of storage systems that provide the high performance of traditional parallel storage and meet the needs of mission-critical and emerging workloads.
- This chapter describes the hardware and software architecture, functions and features, and Typical Scenarios of Huawei OceanStor Pacific .

Objectives

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

- Understand the hardware and software architecture
- Understand the technical features of superior performance
- Understand the technical features of high reliability
- Understand the technical features of high efficiency
- Understand the technical features of solid security and stability
- List the typical scenarios

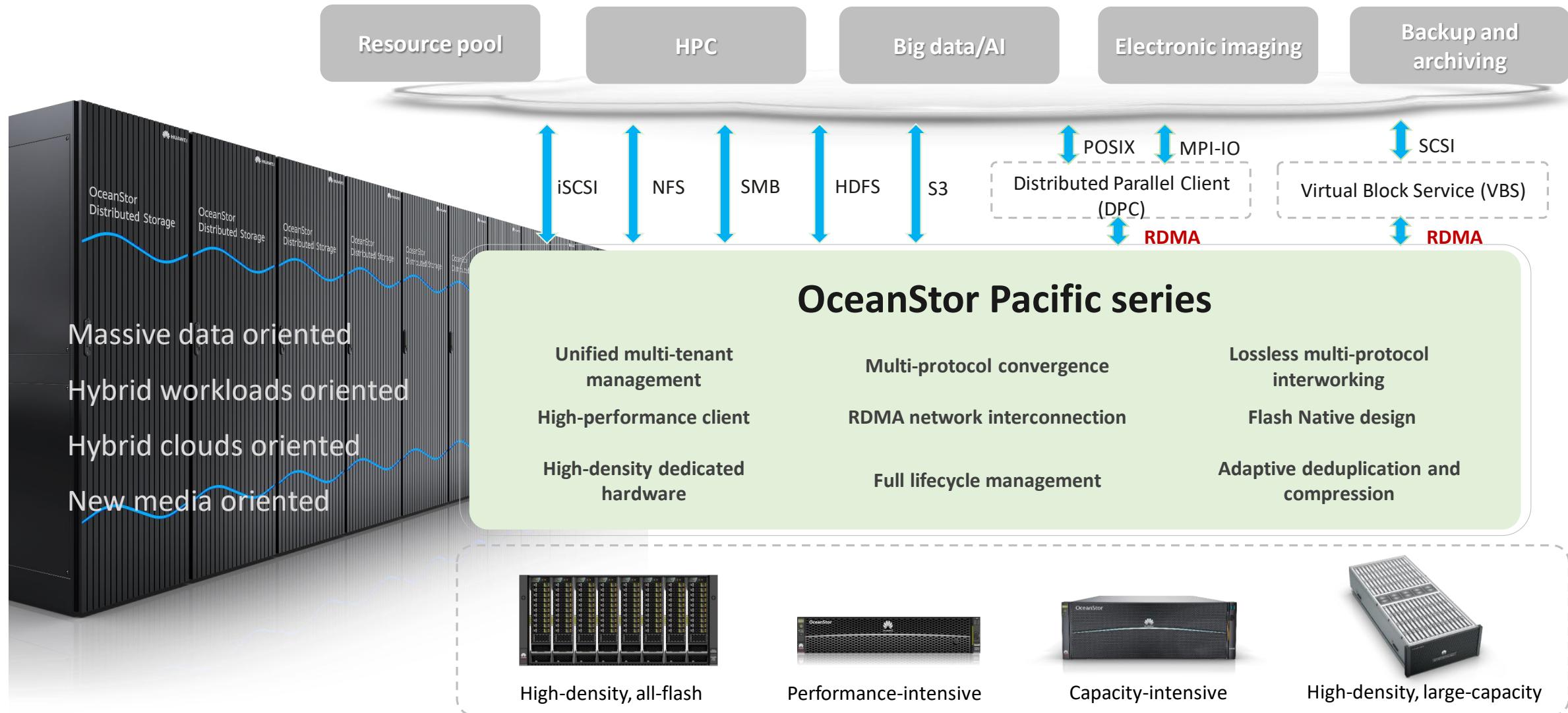
Contents

- 1. Product Overview**
2. Hardware Architecture
3. Software Architecture
4. Superior Performance
5. High Reliability
6. High Efficiency
7. Solid Security and Stability
8. Typical Scenarios

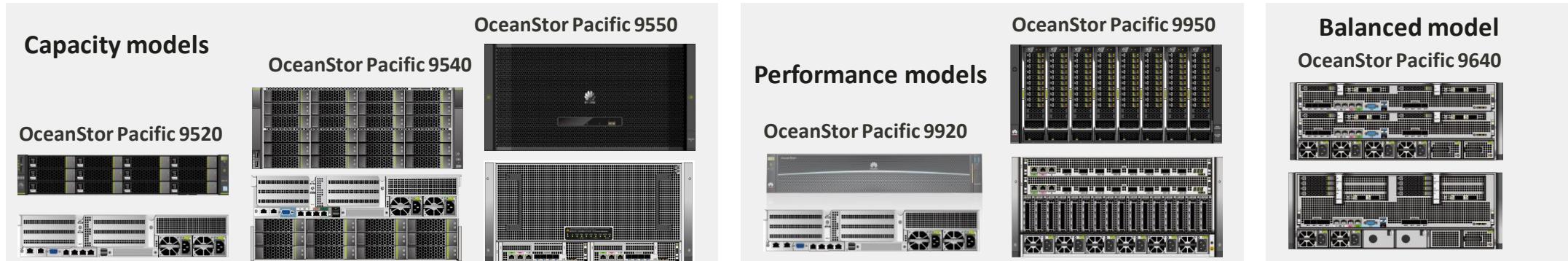
Overview and Objectives

- This section describes the overall introduction to Huawei OceanStor Pacific and the specifications of each model.
- On completion of this section, you will be able to:
 - List the unstructured protocols supported by OceanStor Pacific
 - Understand the models and classifications of OceanStor Pacific

OceanStor Pacific Series Overview



Specifications of OceanStor Pacific (Unstructured Service)



Category	Capacity models					Performance models		Balance model		
	OceanStor Pacific 9520		OceanStor Pacific 9540		OceanStor Pacific 9550	OceanStor Pacific 9920	OceanStor Pacific 9950	OceanStor Pacific 9640 (1 node)	OceanStor Pacific 9640 (2 nodes)	
Product model	x86	Kunpeng920	x86	Kunpeng920	Kunpeng920	Kunpeng920	Kunpeng920	Kunpeng920	Kunpeng920	
Chassis form factor	2 U/1 node	2 U/1 node	4 U/1 node	4 U/1 node	5 U/2 nodes	2 U/1 node	5 U/8 nodes	4U/1 nodes	4U/2 nodes	
CPUs/node	Two x86 processors	Two Kunpeng 920 processors	Two x86 processors	Two Kunpeng 920 processors (48-core)	Kunpeng 920 processor (64-core)	Two Kunpeng 920 processors (48-core)	Kunpeng 920 processor (64-core)	Two Kunpeng 920 processors (48-core)	Two Kunpeng 920 processors (48-core)	
Max. number of primary storage disks/node	12	12	36	36	60	12	10	60	30	
Max. memory/node	256 GB	512 GB	256 GB	512 GB	256 GB	512 GB	256 GB	256G	256G	
Max. cache/node	1 to 4 NVMe SSDs	1 to 4 NVMe SSDs	2 to 4 NVMe SSDs	2 to 4 NVMe SSDs	4 half-palm NVMe SSDs	NA	NA	4 palm SSDs	2 palm SSDs	
System disks/node	2 x 600 GB SAS HDDs	2 x 600 GB SAS HDDs	2 x 600 GB SAS HDDs	2 x 600 GB SAS HDDs	2 x 480 GB SSDs	2 x 600 GB SAS HDDs	2 x 480 GB SSDs	2 x 480 GB SSDs	2 x 480 GB SSDs	
BBU	NA	Yes	NA	Yes	Yes	Yes	Yes	Yes	Yes	
Data disk type	SATA HDD					SAS SSD	half-palm NVMe SSD	SATA HDD	SATA HDD	
Front-End Service Networks	<ul style="list-style-type: none"> • 10GE, 25GE, or 100GE TCP/IP • 25GE or 100GE RoCE • 100 Gb/s EDR InfiniBand 		<ul style="list-style-type: none"> • 10GE, 25GE, or 100GE TCP/IP • 25GE or 100GE RoCE • 100 Gb/s EDR InfiniBand 		<ul style="list-style-type: none"> • 10GE, 25GE, or 100GE TCP/IP • 25GE or 100GE RoCE • 100 Gb/s EDR/HDR InfiniBand 		<ul style="list-style-type: none"> • 10GE, 25GE or 100GE TCP/IP • 100GE RoCE • 100 Gb/s EDR/HDR InfiniBand 		<ul style="list-style-type: none"> • 10GE, 25GE, or 100GE TCP/IP • 25GE or 100GE RoCE • 100 Gb/s EDR/HDR InfiniBand 	
Storage Interconnection Networks	<ul style="list-style-type: none"> • 10GE, 25GE or 100GE RoCE • 100 Gb/s EDR InfiniBand 		<ul style="list-style-type: none"> • 10GE, 25GE or 100GE RoCE • 100 Gb/s EDR InfiniBand 		<ul style="list-style-type: none"> • 25GE RoCE 		<ul style="list-style-type: none"> • 25GE or 100GE RoCE • 100 Gb/s EDR InfiniBand 		<ul style="list-style-type: none"> • 10GE, 25GE or 100GE RoCE • 100 Gb/s EDR InfiniBand 	
Applicable storage services	Object and HDFS	Object and HDFS	Object and HDFS	File, object, and HDFS, with multi-protocol interworking						

Quiz

1. (True or False) OceanStor Pacific can provides unstructured data service through iSCSI/NFS/SMB/HDFS/S3 protocols.

2. (Single-choice) Which statement is NOT true about OceanStor Pacific?
 - A. Each 5U chassis of OceanStor Pacific 9950 houses 80disks
 - B. Each 5U chassis of OceanStor Pacific 9550 houses 120 disks
 - C. Each 4U chassis of OceanStor Pacific 9640 houses 60 disks
 - D. Each 2U chassis of OceanStor Pacific 9920 houses 24 disks

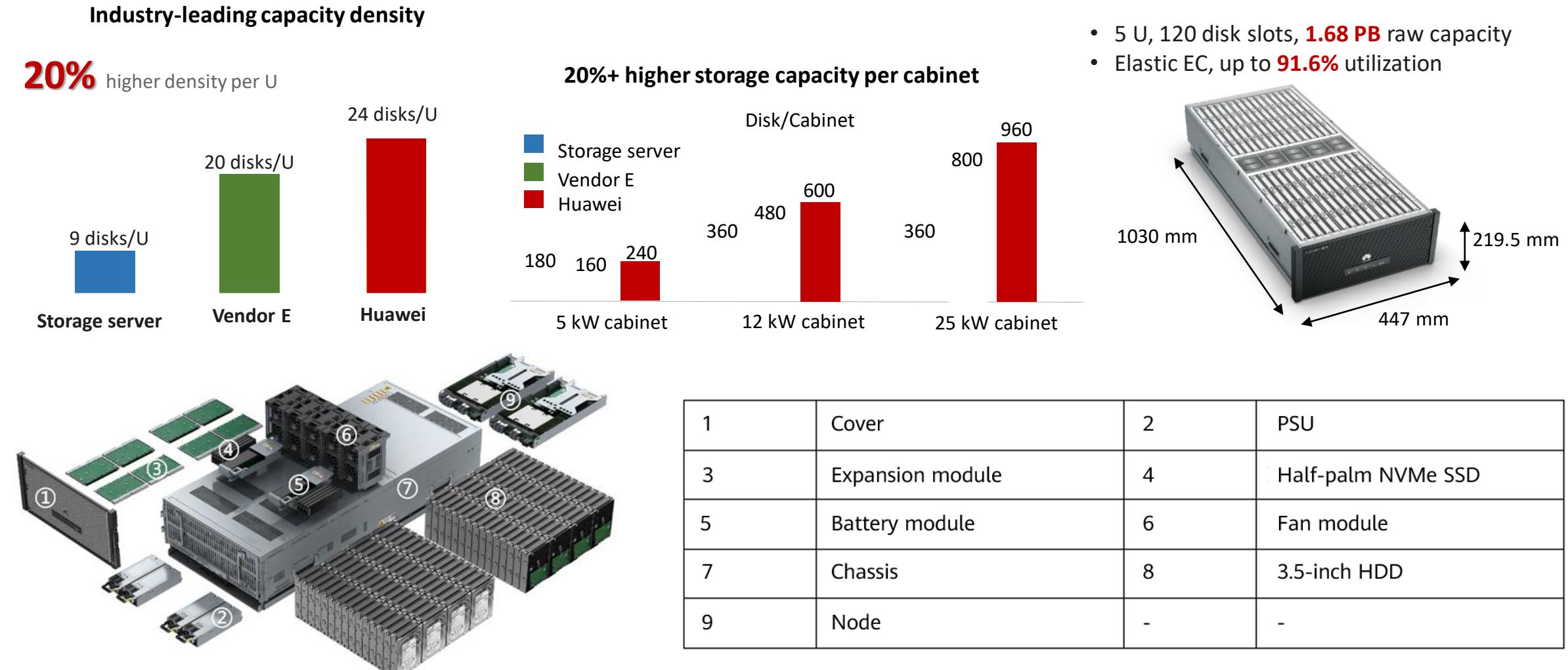
Contents

1. Product Overview
- 2. Hardware Architecture**
3. Software Architecture
4. Superior Performance
5. High Reliability
6. High Efficiency
7. Solid Security and Stability
8. Typical Scenarios

Overview and Objectives

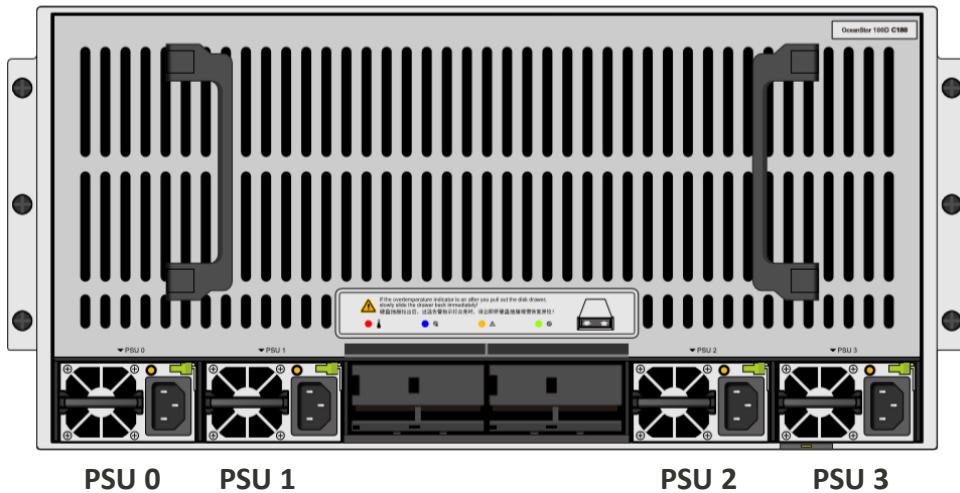
- This section describes the hardware architecture of Huawei OceanStor Pacific 9550 and 9950.
- On completion of this section, you will be able to:
 - Understand the hardware structure, components, and key hardware design of the OceanStor Pacific 9550
 - Understand the hardware structure, components, and key hardware design of the OceanStor Pacific 9950

OceanStor Pacific 9550 High-Density Large-Capacity Model

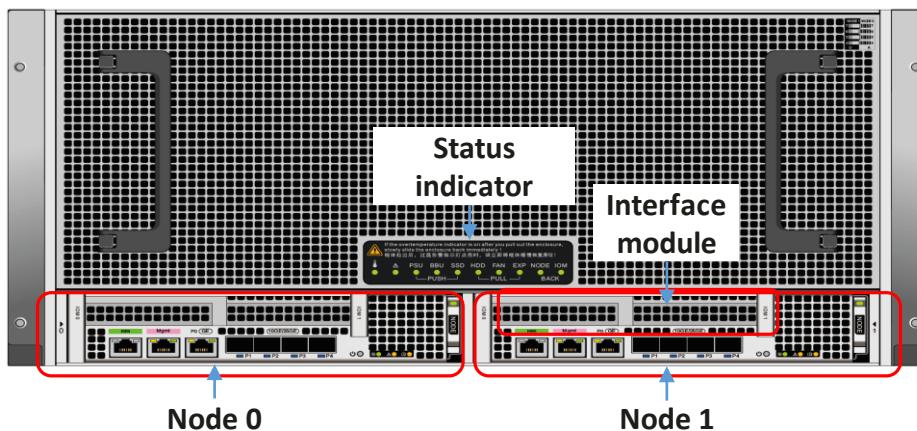


Physical Form of OceanStor Pacific 9550

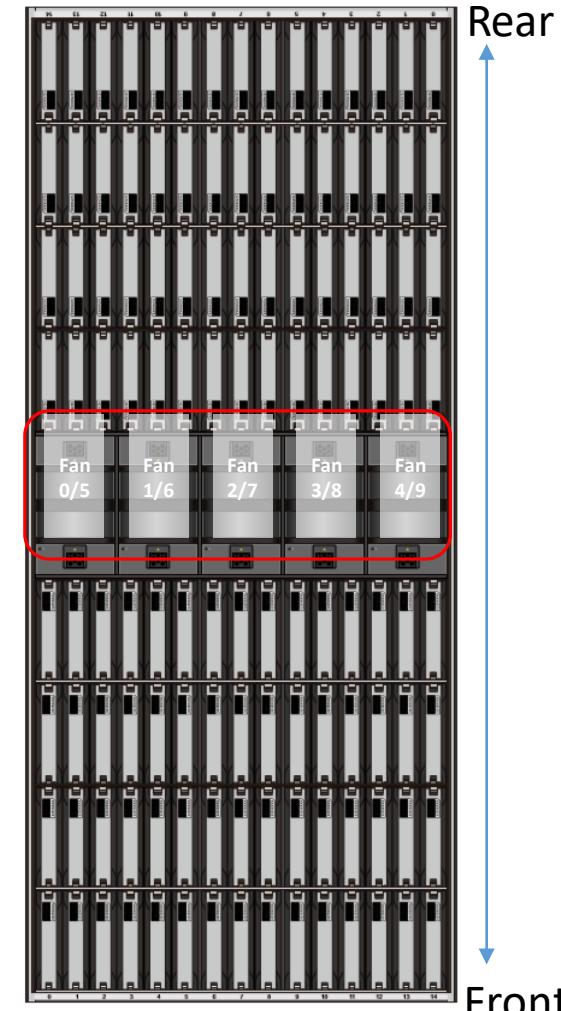
Front view (cover removed)



Rear view



Top view



OceanStor Pacific 9550 Key Hardware Designs

Bi-directional pulling + Holding rail-free

Less shift of center of gravity and more secure operations
Larger space on the left and right, accommodating more media



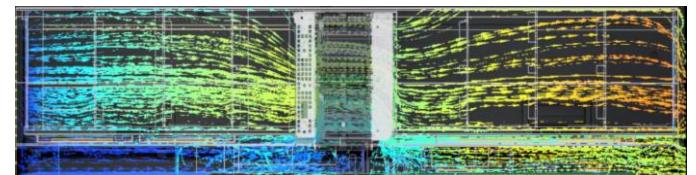
Retraction design

Saving space for power supplies and cables
Industry's first ultra-high-density device that meets the requirements of 1.1 m cabinets

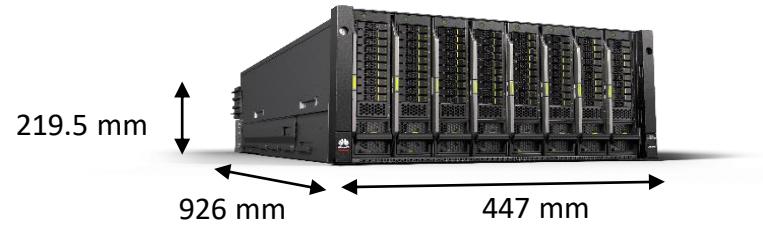
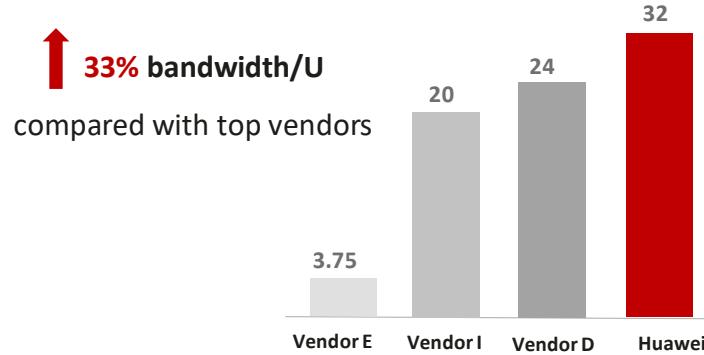


Dual air channels + Counter-rotating pressurized fans

Upper disks and lower nodes and dual air channels, enhancing heat dissipation capabilities
Counter-rotating pressurized fans in the middle, implementing good ventilation in air channels



OceanStor Pacific 9950 High-Density High-Performance Model

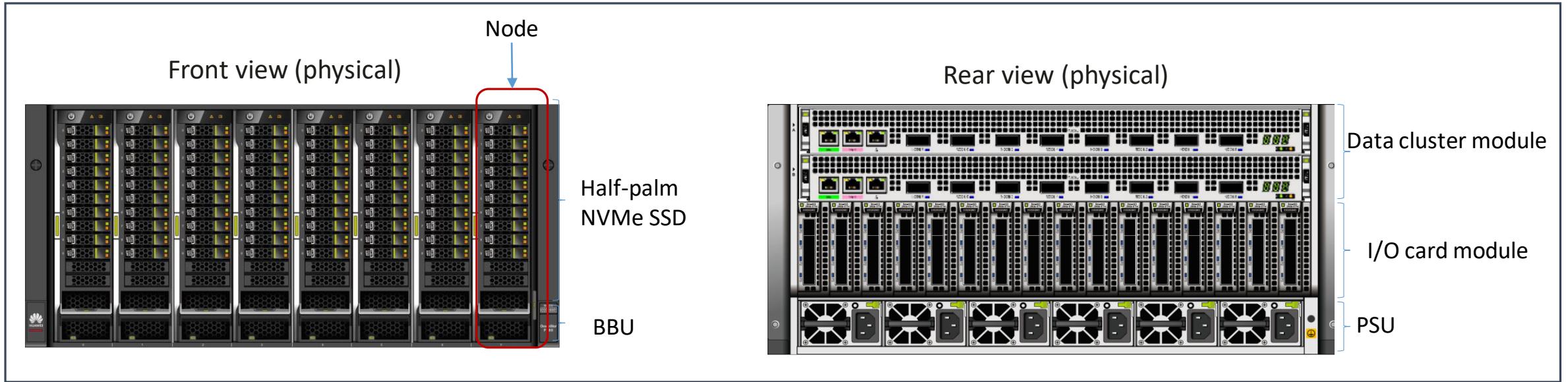


5 U, 8 nodes, 80 disk slots, **160 GB/s read bandwidth or 6.4 million read IOPS** of the entire device



1	Front panel cover	2	System subrack
3	BBU	4	Node
5	Half-palm NVMe SSD (main storage)	6	Fan module
7	PSU	8	I/O card module
9	Data cluster module	-	-

Physical Form of OceanStor Pacific 9950

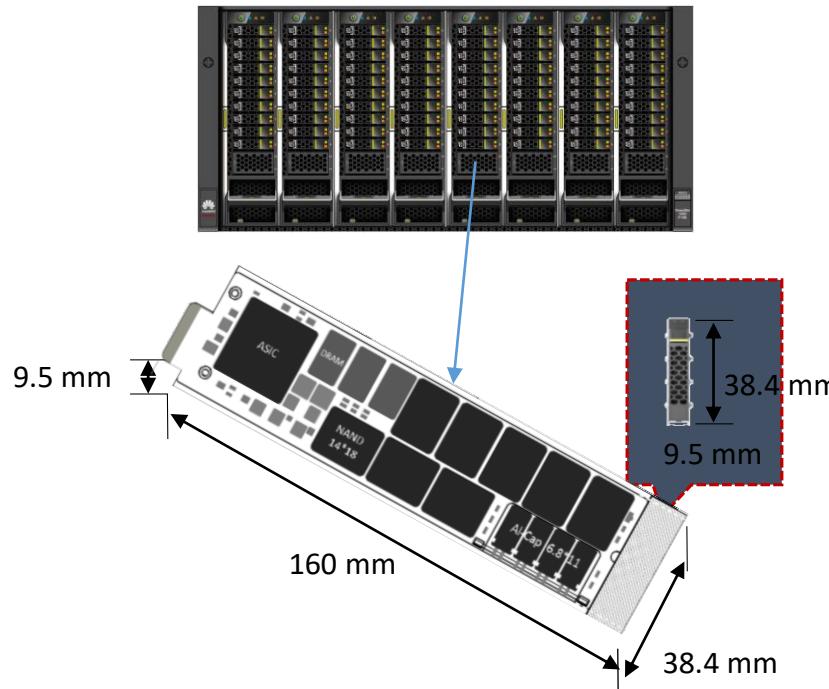


OceanStor Pacific 9950 Key Hardware Designs

High-density multi-node + NVMe SSD

A single device can function as a cluster, with **160 Gbit/s** hardware channel capability.

Each U space can accommodate **33%** more all-flash media.



Dedicated backup power

32 GB battery-protected memory per node, providing

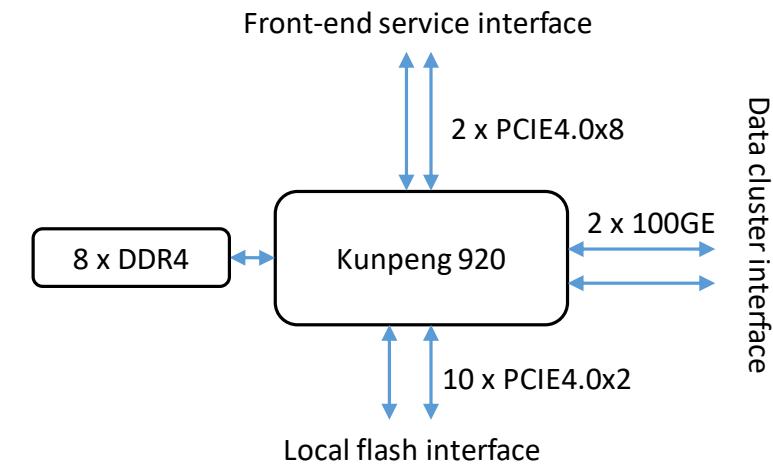
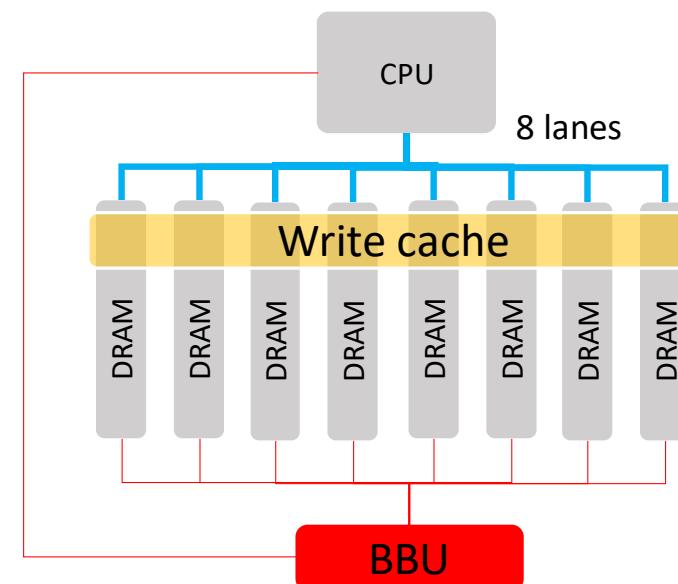
memory-level latency for key data

Concurrent memory channel bandwidth, up to 8 times that
of **NVDIMM**

Optimal bandwidth ratio

Front-end : Interworking : Back-end = 1:1:1.25

Optimal bandwidth ratio, 20 Gbit/s bandwidth
performance per node



Quiz

1. (Multiple-choice) Which of the following key hardware designs are used for high density in OceanStor Pacific 9550?
 - A. Bi-directional pulling&Holding rail-free
 - B. Retraction design
 - C. Dual air channels
 - D. Counter-rotating pressurized fans

2. (Multiple-choice) Which of the following key hardware designs are used in OceanStor Pacific 9950?
 - A. High-density multi-node
 - B. Dedicated backup power
 - C. Half-palm NVMe SSD
 - D. Optimal bandwidth ratio

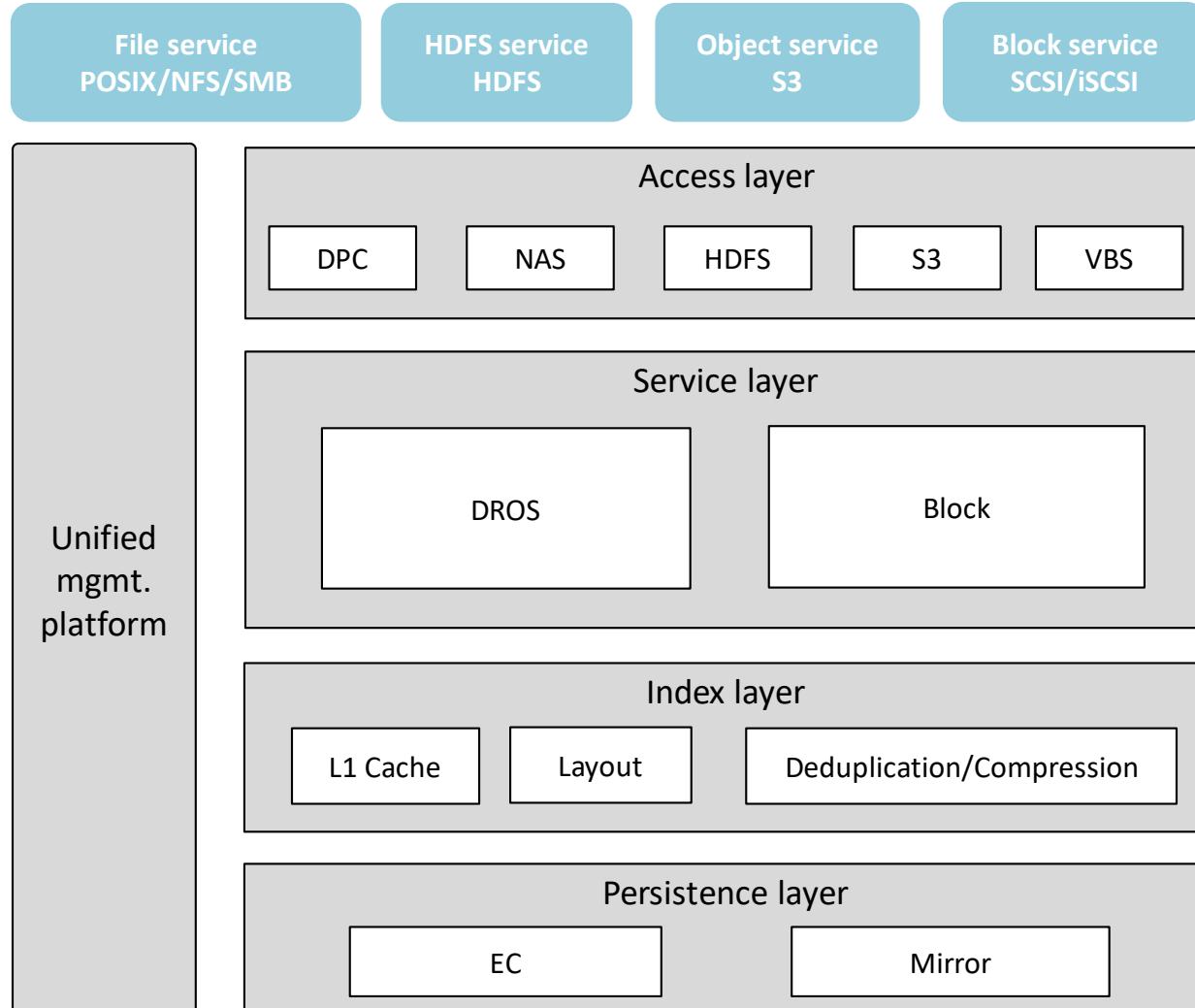
Contents

1. Product Overview
2. Hardware Architecture
- 3. Software Architecture**
4. Superior Performance
5. High Reliability
6. High Efficiency
7. Solid Security and Stability
8. Typical Scenarios

Overview and Objectives

- This section describes the software architecture of Huawei OceanStor Pacific.
- On completion of this section, you will be able to:
 - Understand the overall software architecture
 - Understand the design of key technologies for convergence and interworking

Overall Software Architecture



DPC for HPC

- DPC is compatible with standard POSIX and MPI-IO semantics and provides high-performance and parallel storage capabilities for HPC scenarios.

Unified Data Service Base

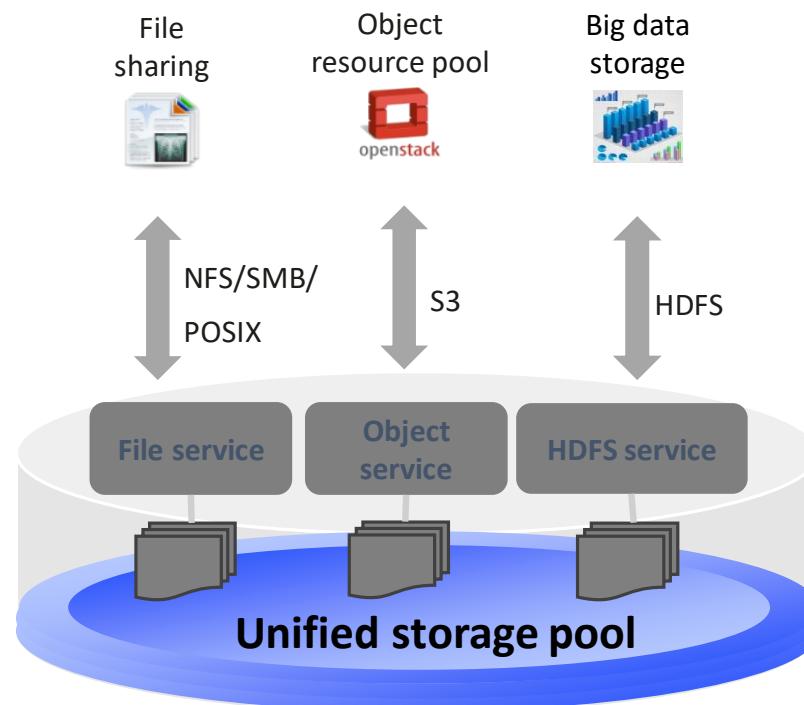
- Unstructured services (object, file, and HDFS) are processed on the DROS platform in a unified manner to implement multi-protocol interworking and enterprise features such as HyperSnap, SmartQuota, SmartTier, and HyperReplication.
- The structured service (block) is processed by an independent block module.

Unified Distributed Persistence Layer

- The object, file, HDFS services share storage pools, reducing costs.

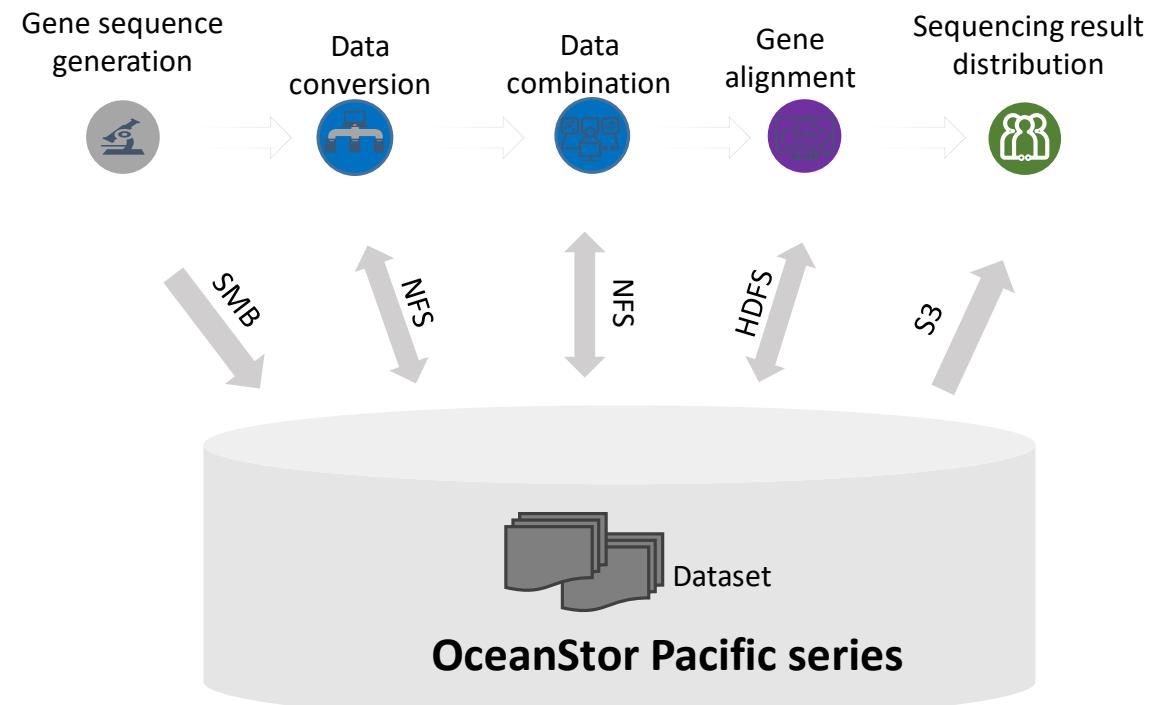
Convergence and Interworking: One Storage System Meets Diversified Storage Access Requirements

Multiple services in Converged resource pool
enables resource sharing



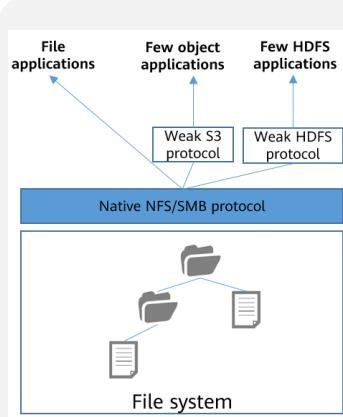
Multi-protocol interworking ensures
zero data copy

Gene sequencing analysis



Convergence and Interworking: Unified Metadata Management for Multiple Protocols Avoid Semantics Loss

Traditional Architecture

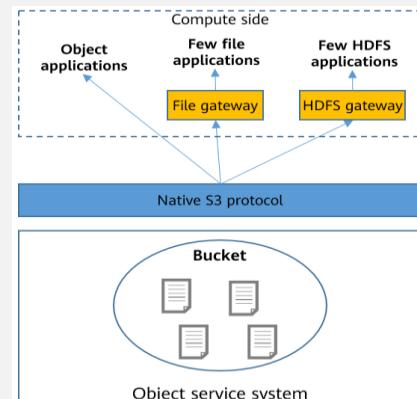


NAS + gateway

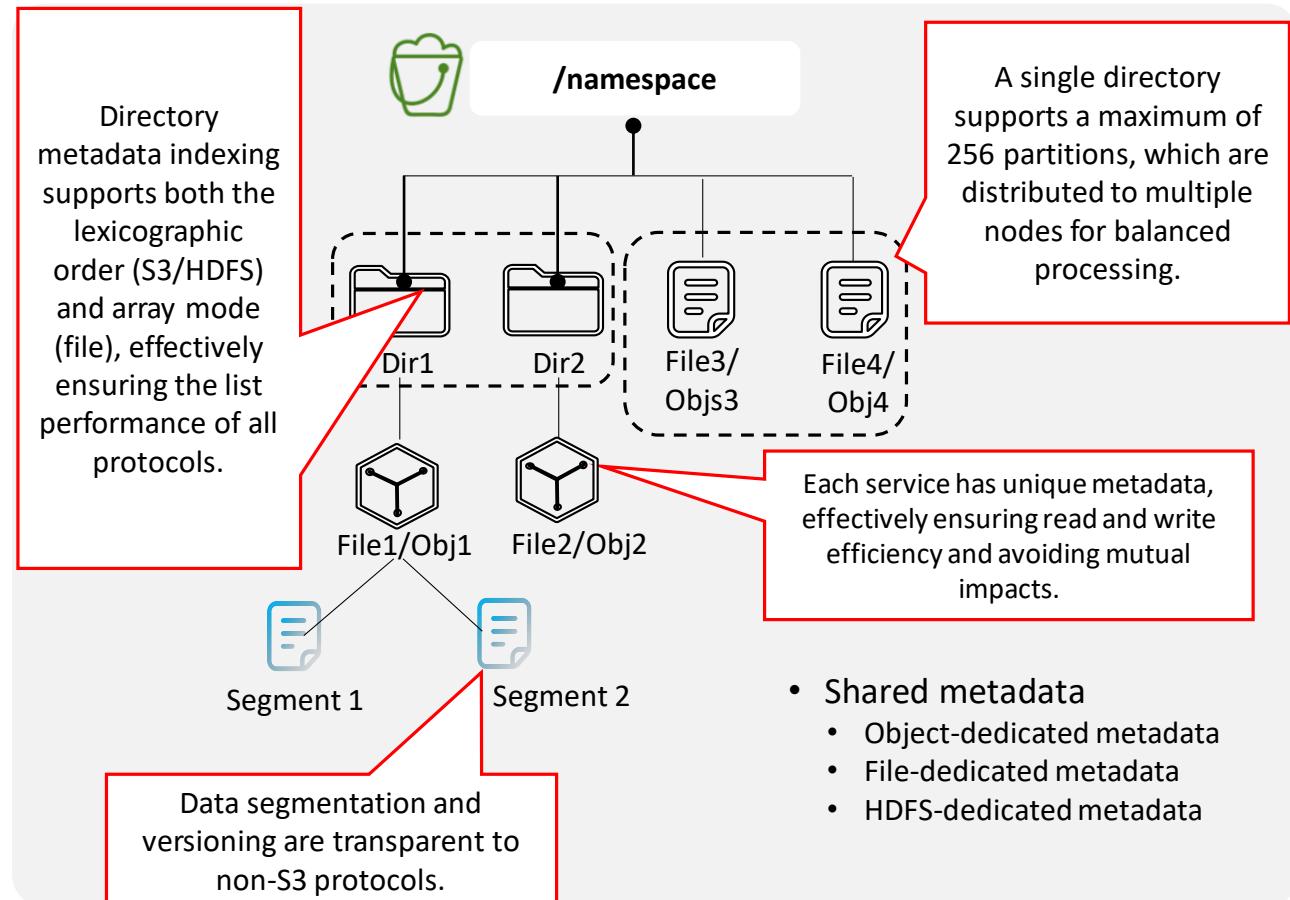
- The object protocol does not support functions such as versioning, multi-part upload, quota, and user-defined tags.
- The number of objects supported by a bucket is limited by the file system capability, and the object list operation performance is poor.
- The HDFS directory statistics performance is poor.

Object + gateway

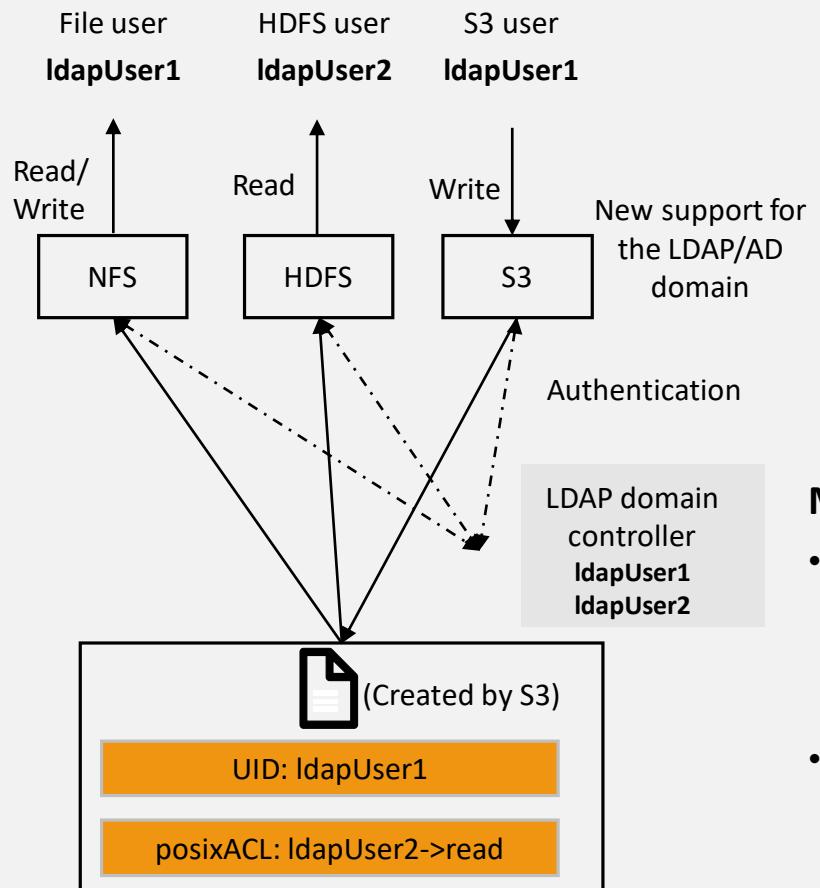
- Directory permissions cannot be modified.
- File locks are not supported, so data consistency cannot be ensured.
- The directory renaming (copy + deletion) performance is poor.
- The operation performance of large files and small I/Os is poor.



OceanStor Pacific Architecture



Convergence and Interworking: Unified Authentication of Multi-Protocol Users, Minimum Authorization, and Extended Locks Ensure Data Security



Unified multi-protocol control: unified authentication

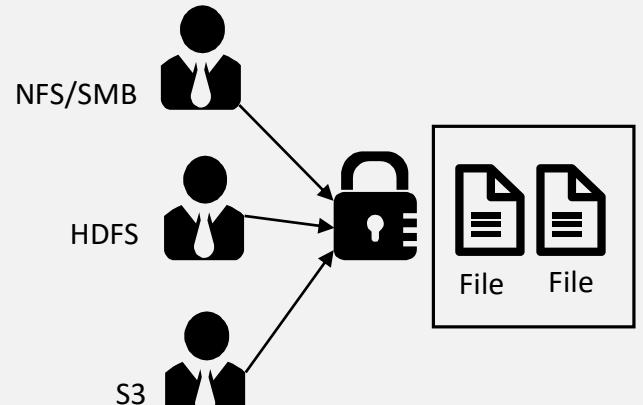
- S3: newly supports LDAP and AD domain authentication
- NFS, SMB, S3, and HDFS support unified domain controller users or user mapping modes to ensure permission interworking between protocols.

Minimum authorization

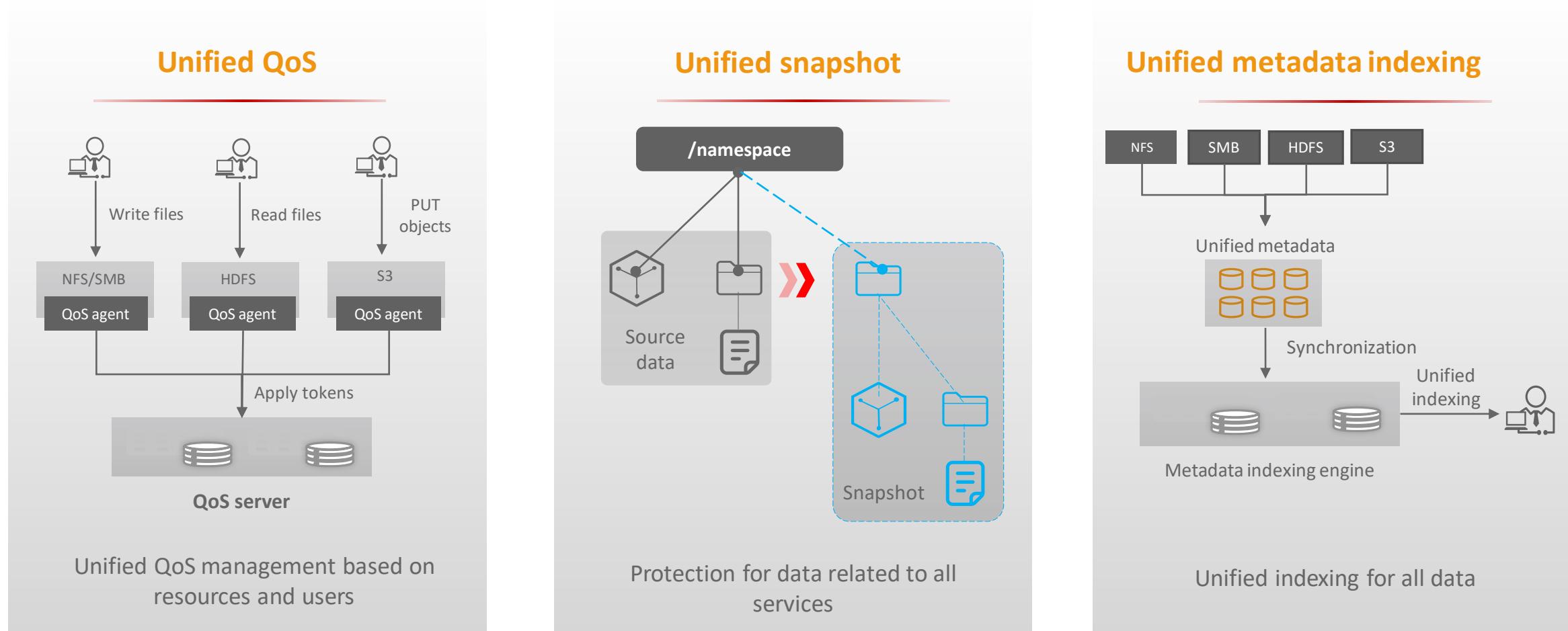
- By default, tenants of the object service cannot access each other's resources. They can do so only after the access is authorized by the tenant owner.
- For the file service, access authorization is available only for the users and users groups in a tenant.

Unified cross-protocol I/O lock management

- Object service: newly supports the Lease lock. When NFS, CIFS, S3, and HDFS are used to concurrently access the same file, forcible lock protection is supported.



Convergence and Interworking: Multiple Protocols Share Value-added Features of Storage Services



Quiz

1. (Multiple-choice) Which of the following services can be deployed in the same storage pool of OceanStor Pacific?
 - A. File
 - B. HDFS
 - C. S3
 - D. Block

2. (Multiple-choice) Which of the following enterprise features are shared by multiple protocols in OceanStor Pacific?
 - A. QoS
 - B. Snapshot
 - C. Authentication
 - D. Metadata indexing

Contents

1. Product Overview
2. Hardware Architecture
3. Software Architecture
- 4. Superior Performance**
5. High Reliability
6. High Efficiency
7. Solid Security and Stability
8. Typical Scenarios

Overview and Objectives

- This section describes the high-performance design of Huawei OceanStor Pacific.
- On completion of this section, you will be able to:
 - Understand the key technologies of high bandwidth optimization
 - Understand the key technologies of high IOPS optimization
 - Understand the key technologies of hybrid load conflict optimization

High-Performance Design: One Storage System Delivers Both High Bandwidth and High IOPS

Oil exploration

20 GB/s bandwidth per PB or higher



Autonomous driving

More than 50 GB/s bandwidth for the simulation
Millions of IOPS for AI training



Animation rendering

500 K IOPS or higher per PB



High bandwidth optimization

Large I/O
passthrough

SmartEqualizer

RDMA

DPC

Hybrid load conflict optimization

QoS

I/O auto-
adaptation

Dynamic space
management

Intelligent multi-
core CPU

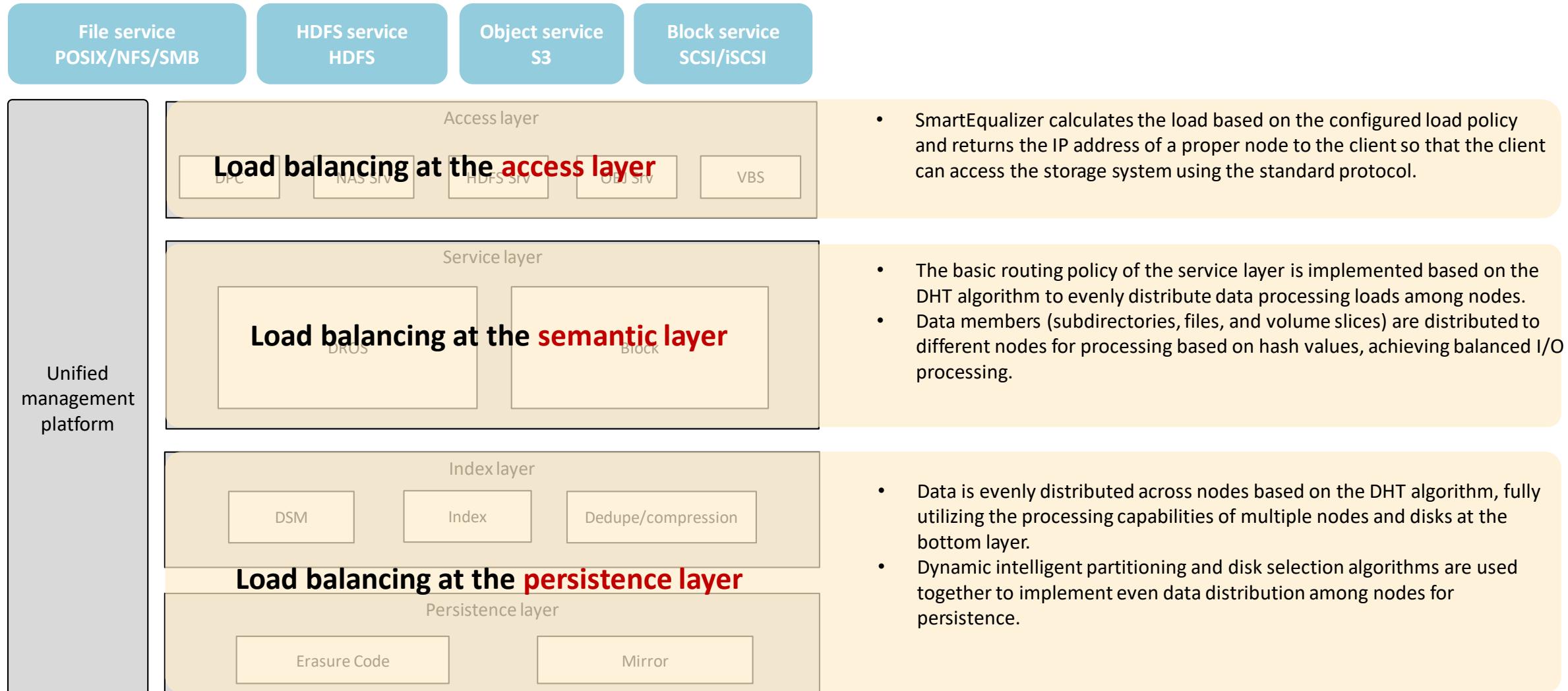
High IOPS optimization

Small I/O
aggregation

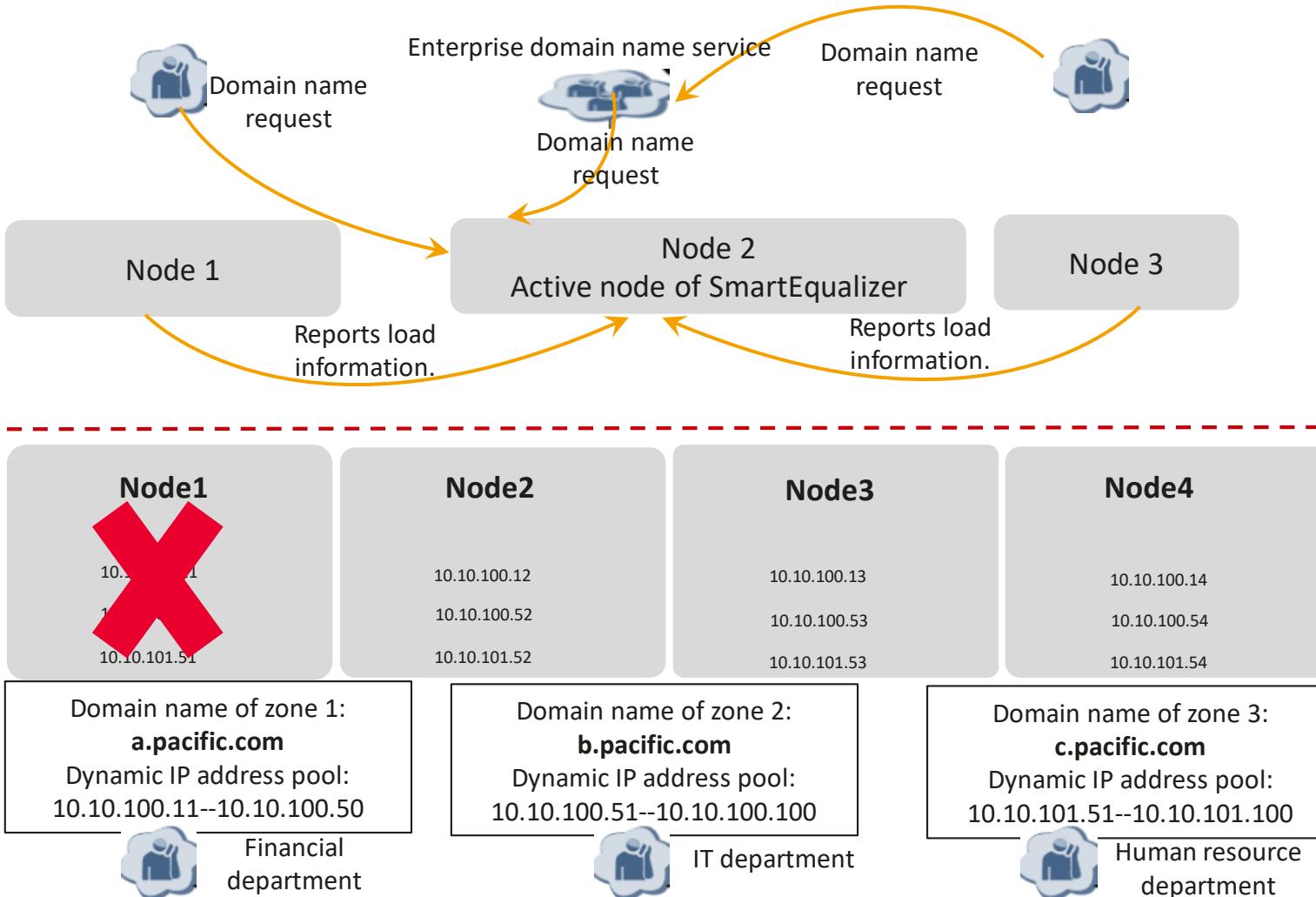
Balanced data
processing

No distributed
lock

High-Performance Design: End-to-End Load Balancing



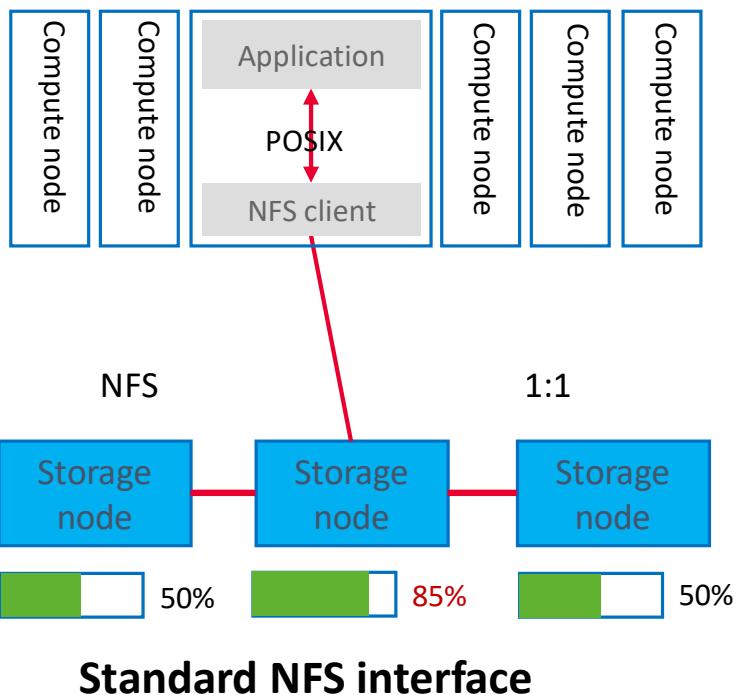
SmartEqualizer Implements Load Balancing and Failover at the Access Layer (for Unstructured Services)



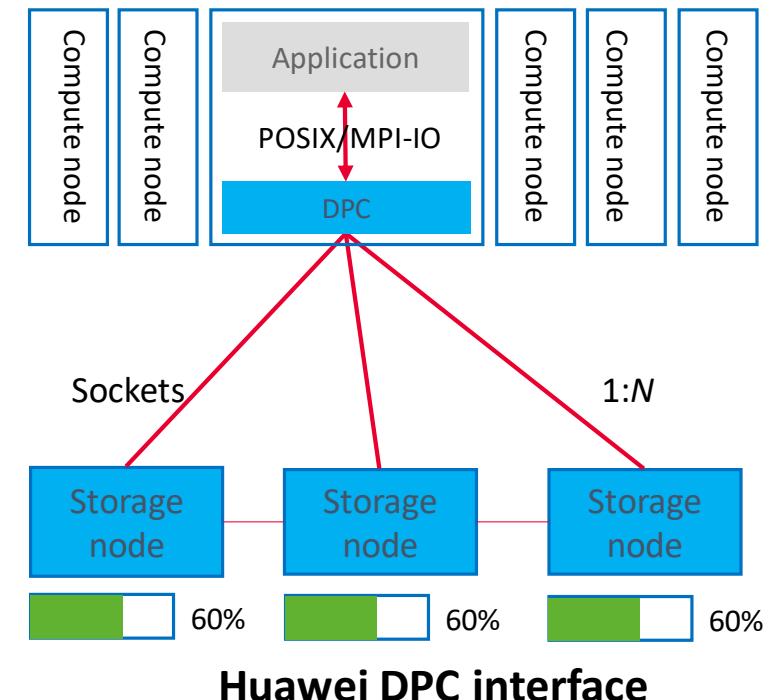
- **Domain name-based load balancing**
SmartEqualizer calculates the load based on the configured load policy and returns the IP address of a proper node to the client so that the client can access the storage system.
- **IP resource pooling**
A single network port supports 16 network zones. Each zone can belong to different tenants and subnets to implement security isolation and intra-zone load balancing.
- **Load balancing and automatic failover**
The IP address pool in a zone is dynamically allocated to each node. If a node is faulty, IP addresses can be automatically taken over by multiple nodes to ensure reliability and load balancing. You can also configure the static IP address resource pool mode in which IP addresses do not automatically float.

DPC Implements End-to-End I/O Load Balancing (for the File Service)

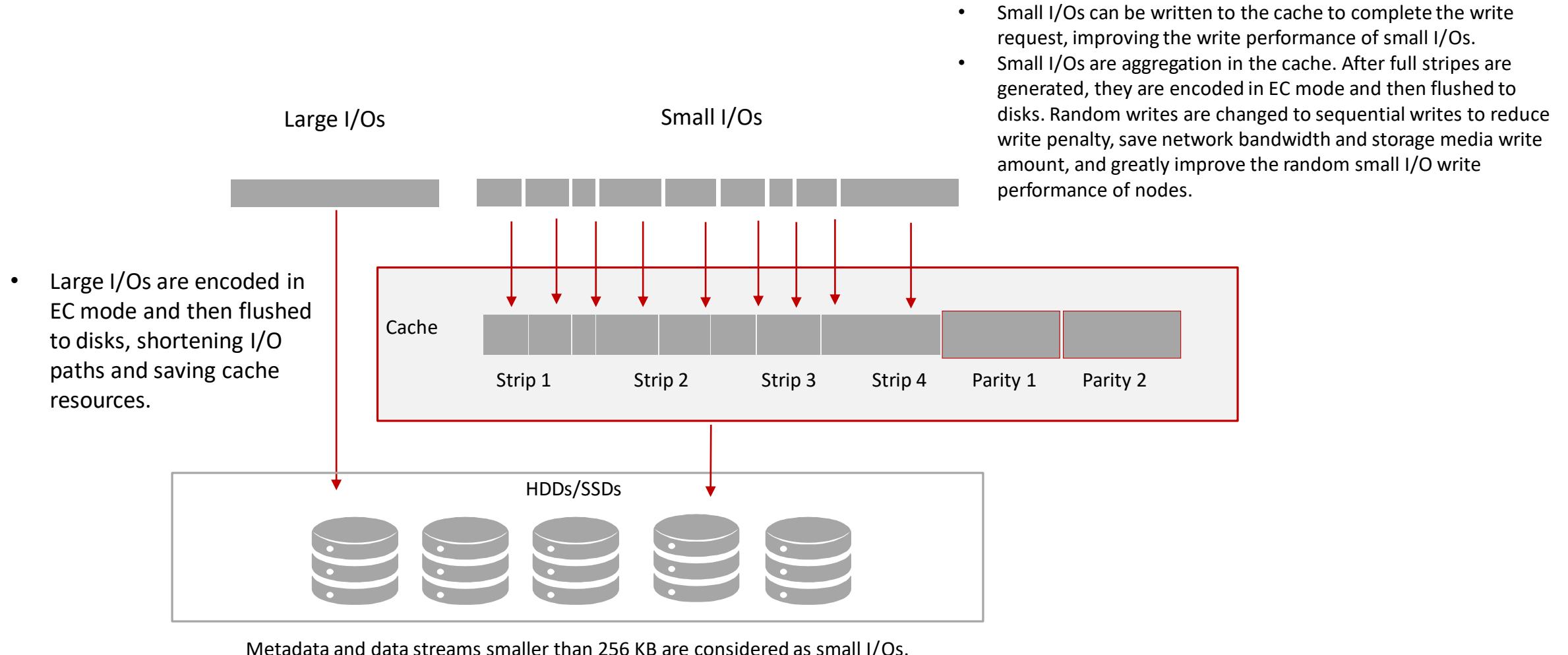
- DPC communicates with storage nodes using RDMA, effectively reducing the CPU usage of compute and storage nodes.
- DPC can be responsible for **I/O addressing, load balancing, and EC operations**. A single client can access multiple nodes at the same time, preventing service load forwarding between storage nodes.
- The client works with MPI-IO to support concurrent processing of a single file among multiple storage nodes and implement cross-node load balancing.



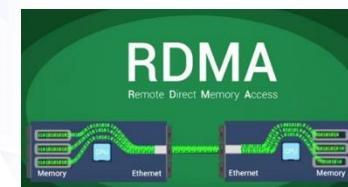
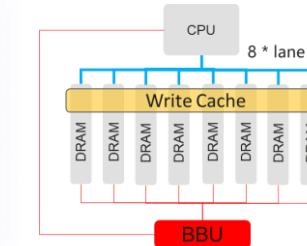
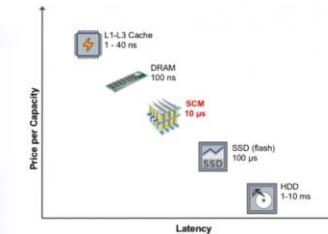
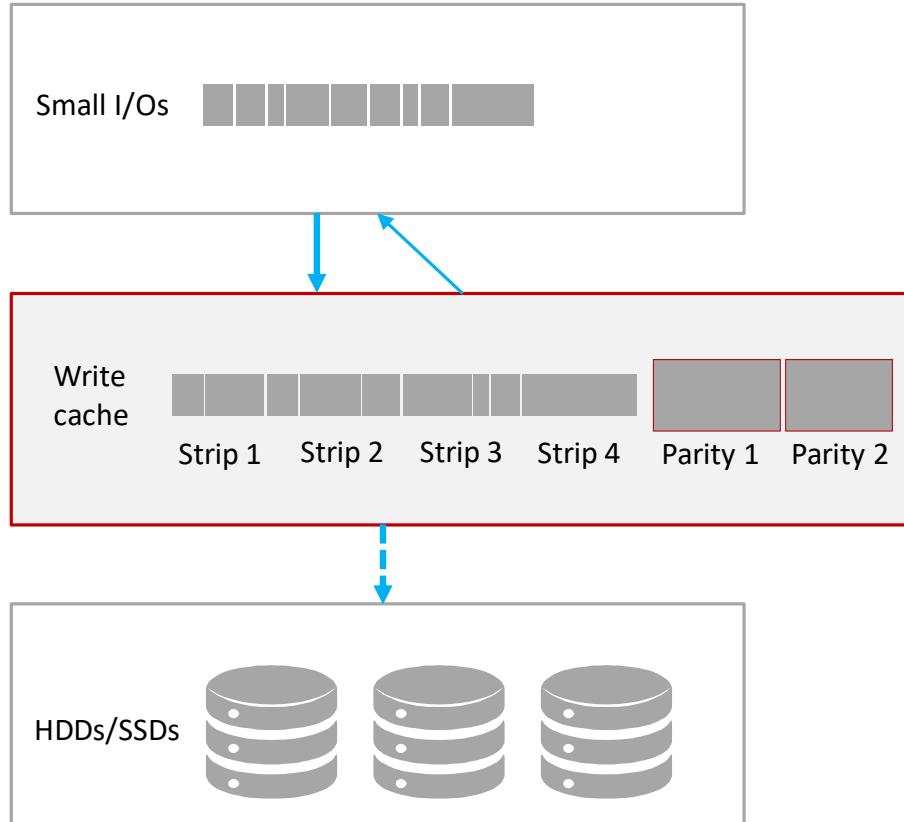
VS



Intelligent I/O Auto-Adaptation, Meeting Different Performance Requirements of Large and Small I/Os



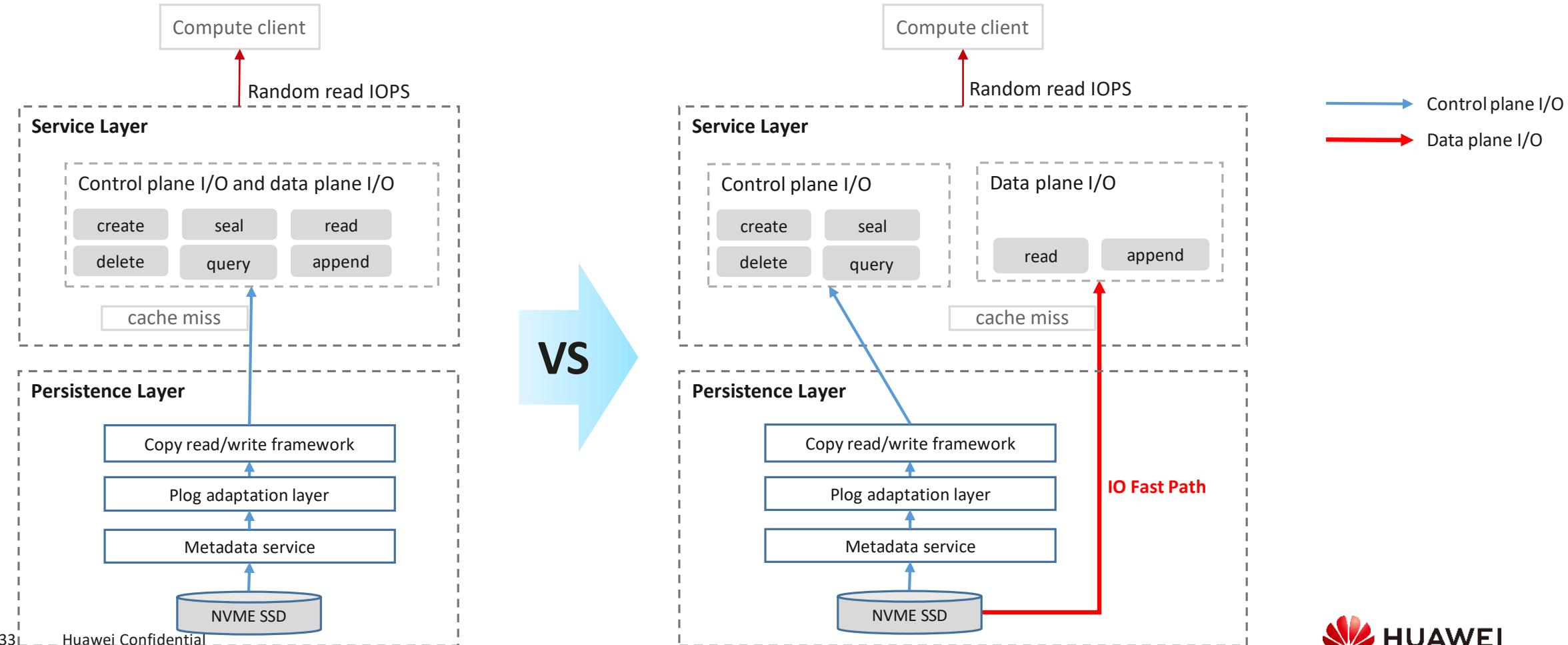
Efficient Write Cache Design and Three Key Technologies for Improving Small I/O Write Performance



- When DRAM is used as the cache, the I/O operation latency is 1% of that of SCM and 1‰ of that of SSDs.
- Unstructured services use the innovative BBU + DRAM architecture. The I/O bandwidth is eight times that of the traditional NVDIMM architecture.
- The RDMA technology is supported for cache mirroring between nodes. Data synchronization does not require CPU intervention, reducing the latency by more than 50%.

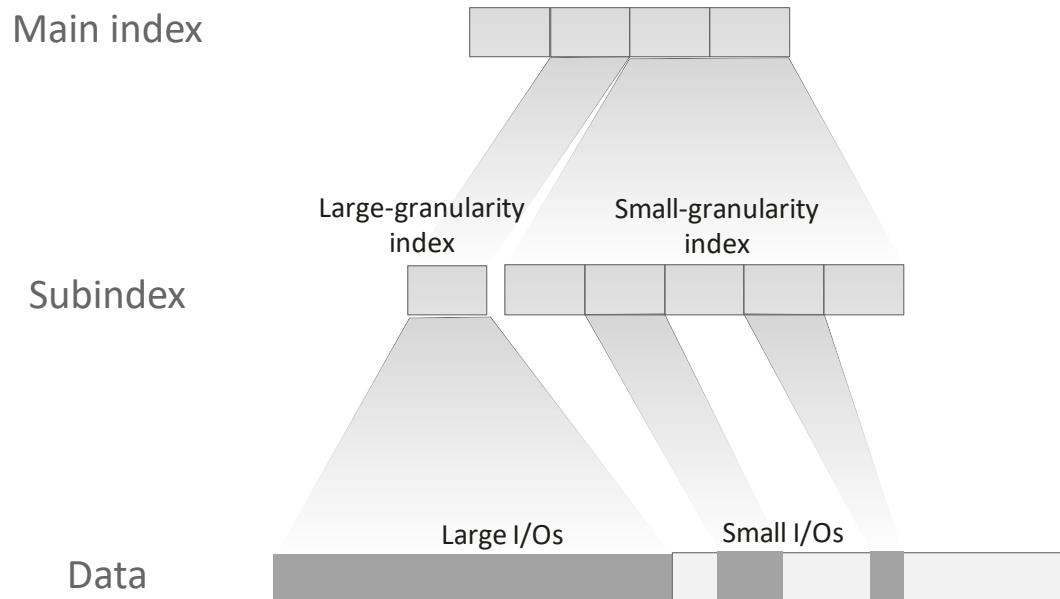
In Read Cache Miss Scenarios, the Data and Control Separation Technology Reduces the I/O Latency of the All-Flash System By 50%.

In the random read IOPS scenario, the cache hit ratio is low and depends on the disk performance. The data and control separation technology separates the data plane I/O path from the control plane I/O path. The I/O function call stack and latency at the storage layer are reduced by 50%, and the random read IOPS of each OceanStor Pacific 9950 node is improved to over 800,000.



Dynamic Granularity Adaptation of Index Space, Ensuring I/O Performance of Both Large and Small Sizes

Flexibly coping with bandwidth-intensive and IOPS-intensive services



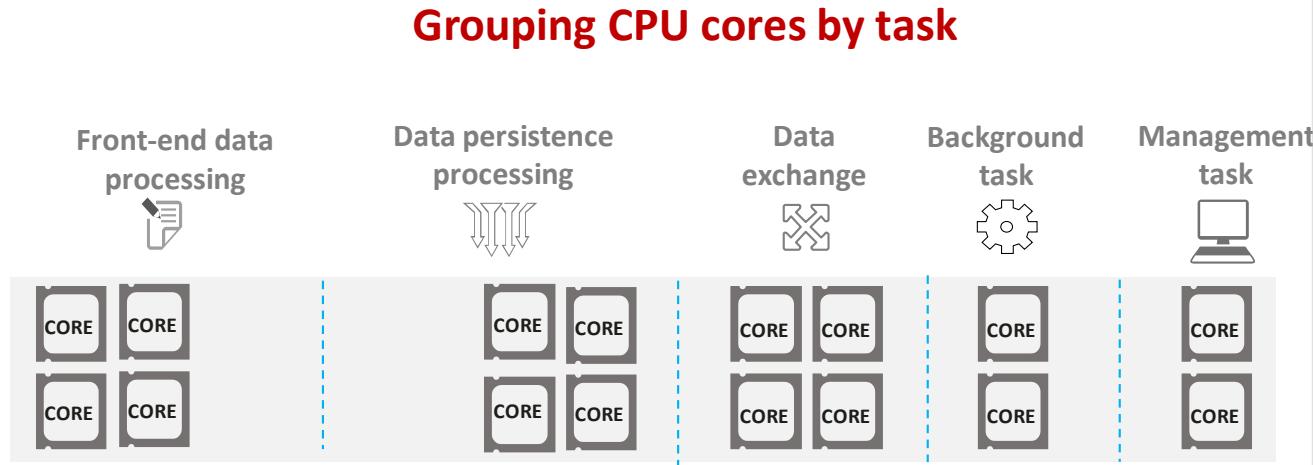
20% ↑ large I/O write bandwidth improvement

Use large-granularity indexes to reduce index space and resource consumption.

60% ↑ 4 KB random write performance improvement

Small-granularity variable-length indexes are used for small I/O writes to prevent read/write amplification.

Intelligent Multi-Core Technology, Efficiently Utilizing CPU Resources



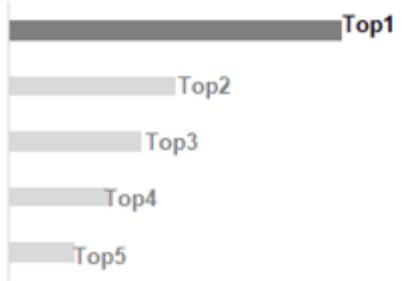
Increasing IOPS-intensive services

CPU cores are grouped by task to avoid interference between tasks.

- Mission-critical services have dedicated cores to ensure sufficient resources.
- The number of cores in a group is dynamically adjusted based on IOPS and bandwidth service loads to ensure the optimal CPU resource ratio in hybrid load scenarios.

I/O priority-based intelligent scheduling

Data read and write
Advanced features
Cache batch write
Disk rebuilding
Garbage collection



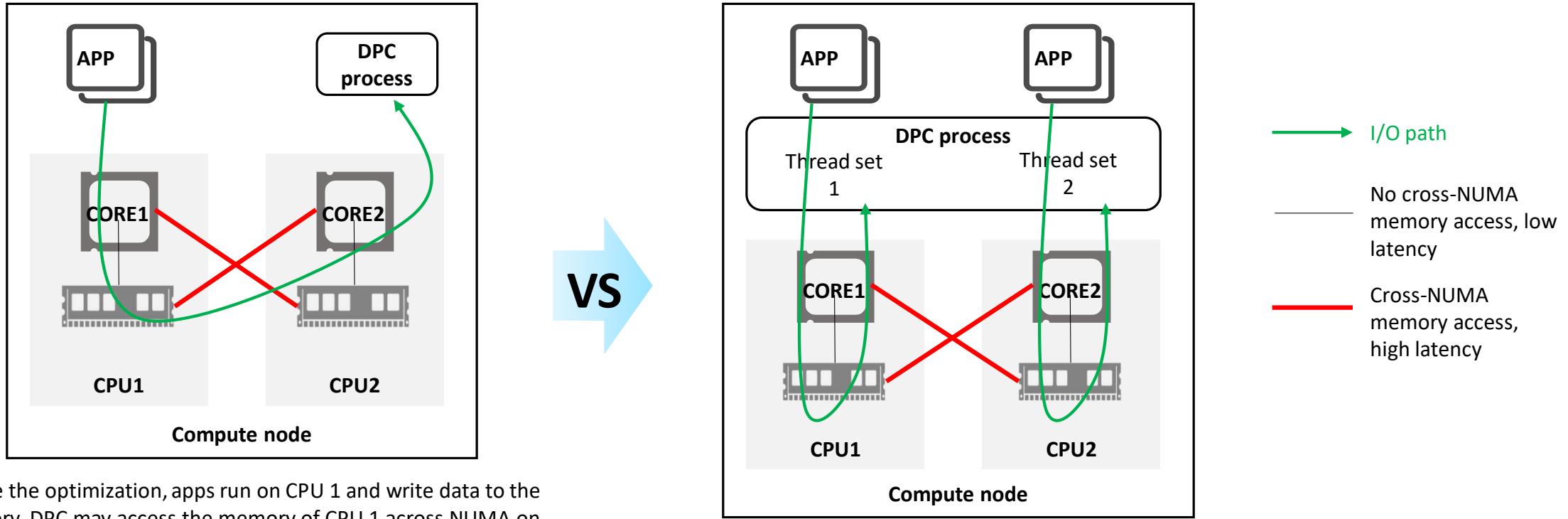
High I/O priority reduces the impact on services.

Data read and write I/Os always have the highest priority to ensure the data read and write I/O performance. Other I/Os are processed later.

DPC Supports NUMA Affinity with Computing Applications, Maximizing Multi-NUMA Performance

The latency of cross-NUMA memory access by the CPU on the compute side is high, affecting the I/O bandwidth and latency.

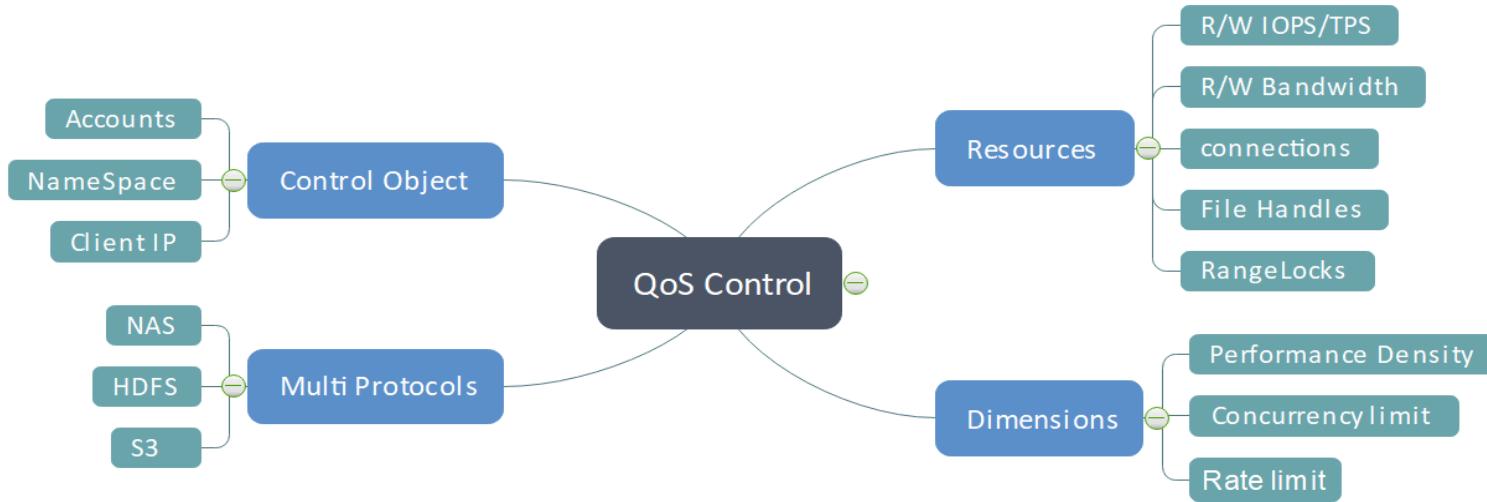
DPC supports NUMA affinity with compute applications, eliminating the latency of cross-NUMA memory access and improving the bandwidth performance of a single DPC client by over 30% in typical scenarios.



Before the optimization, apps run on CPU 1 and write data to the memory. DPC may access the memory of CPU 1 across NUMA on CPU 2, resulting in high latency.

After the optimization, apps run on CPU 1 and write data to the memory. Thread set 1 of DPC accesses the memory of CPU 1, and the latency is low.

SmartQoS, Ensuring Performance of Multiple Storage Services



Functions of QoS

- The QoS control of namespaces, tenants, clients, and volumes prevents mutual interference.
- Supports unified access control of NAS, HDFS, and S3 protocols.
- Flexibly controls the bandwidth and OPS upper limits to adapt to multi-service scenarios.
- Supports object read/write OPS and BPS control.
- Supports concurrent control of file handles and range locks (NAS).
- Supports concurrent connection control (NAS and HDFS).
- Supports the performance density control mode (BW).

Quiz

1. (True or False) The DPC client works with MPI-IO to support concurrent processing of a single file among multiple storage nodes and implement cross-node load balancing.
2. (Single-choice) Which of the following statements about improving performance of OceanStor pacific is not correct?
 - A. Small I/Os can be written to the cache to complete the write request
 - B. Small I/Os are aggregation in the cache
 - C. Small I/Os are encoded in EC mode and then flushed to disks
 - D. Random writes are changed to sequential writes

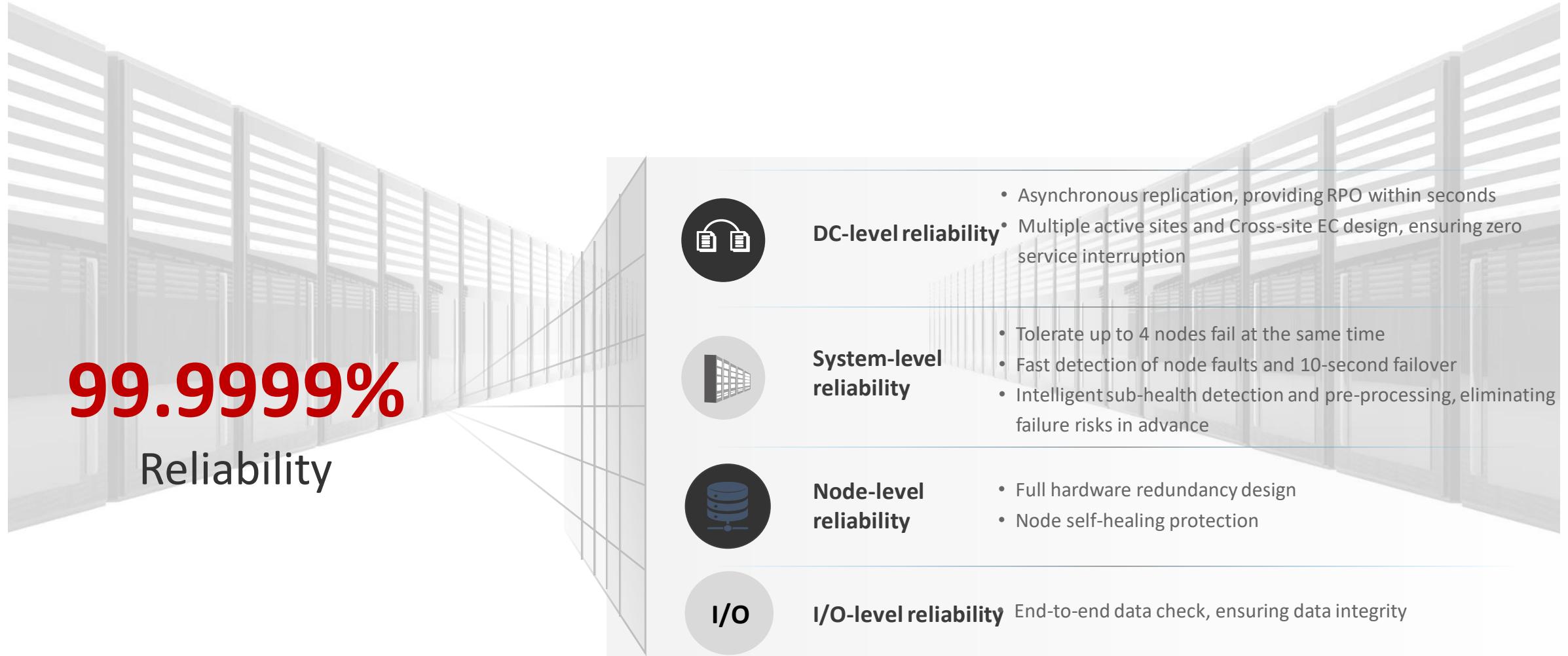
Contents

1. Product Overview
2. Hardware Architecture
3. Software Architecture
4. Superior Performance
- 5. High Reliability**
6. High Efficiency
7. Solid Security and Stability
8. Typical Scenarios

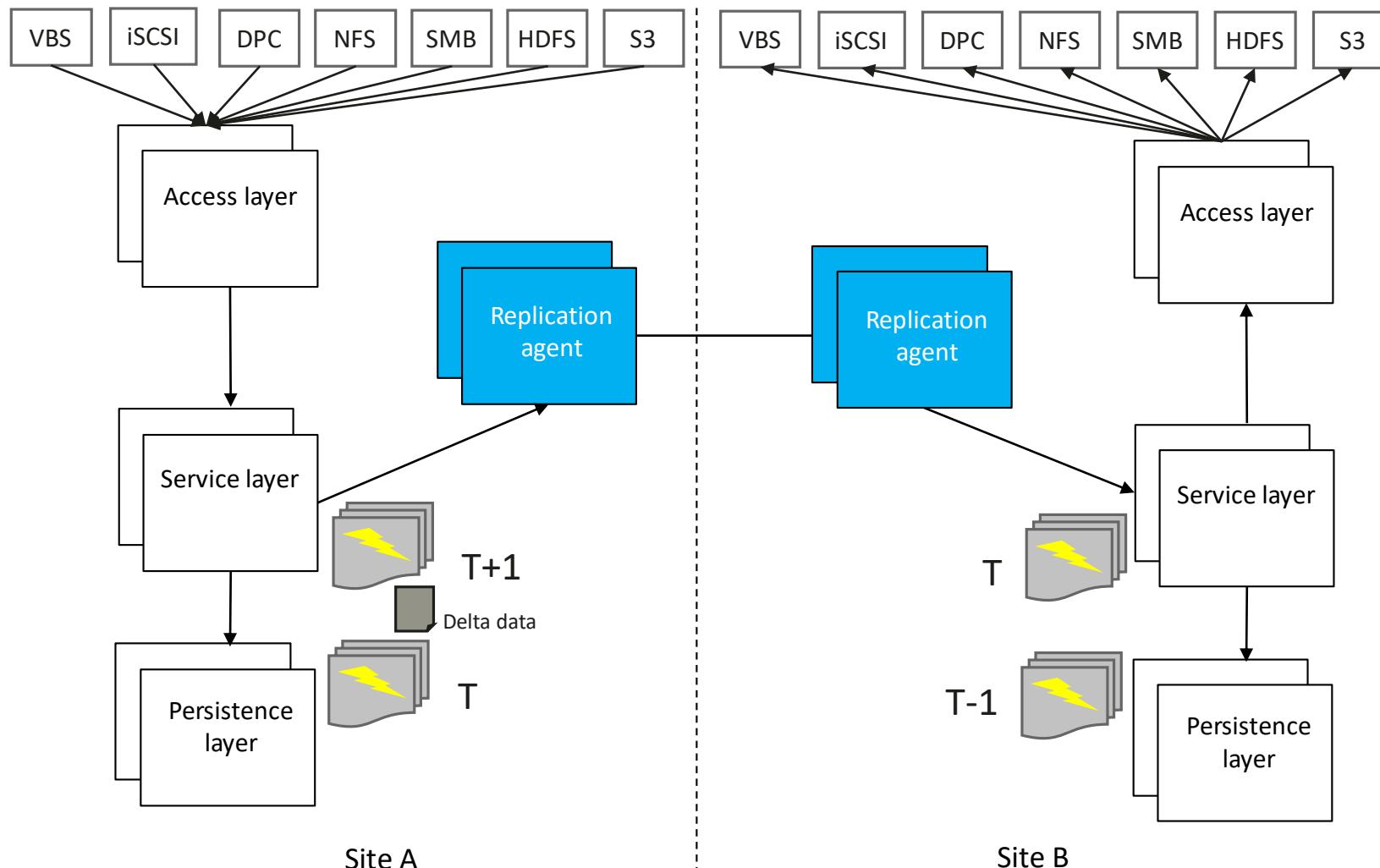
Overview and Objectives

- This section describes the high-reliability design of Huawei OceanStor Pacific.
- On completion of this section, you will be able to:
 - Understand the key technologies of DC-level reliability
 - Understand the key technologies of system-level reliability
 - Understand the key technologies of node-level reliability
 - Understand the key technologies of I/O-level reliability

Multi-Level Reliability Assurance for 24/7 Service Continuity

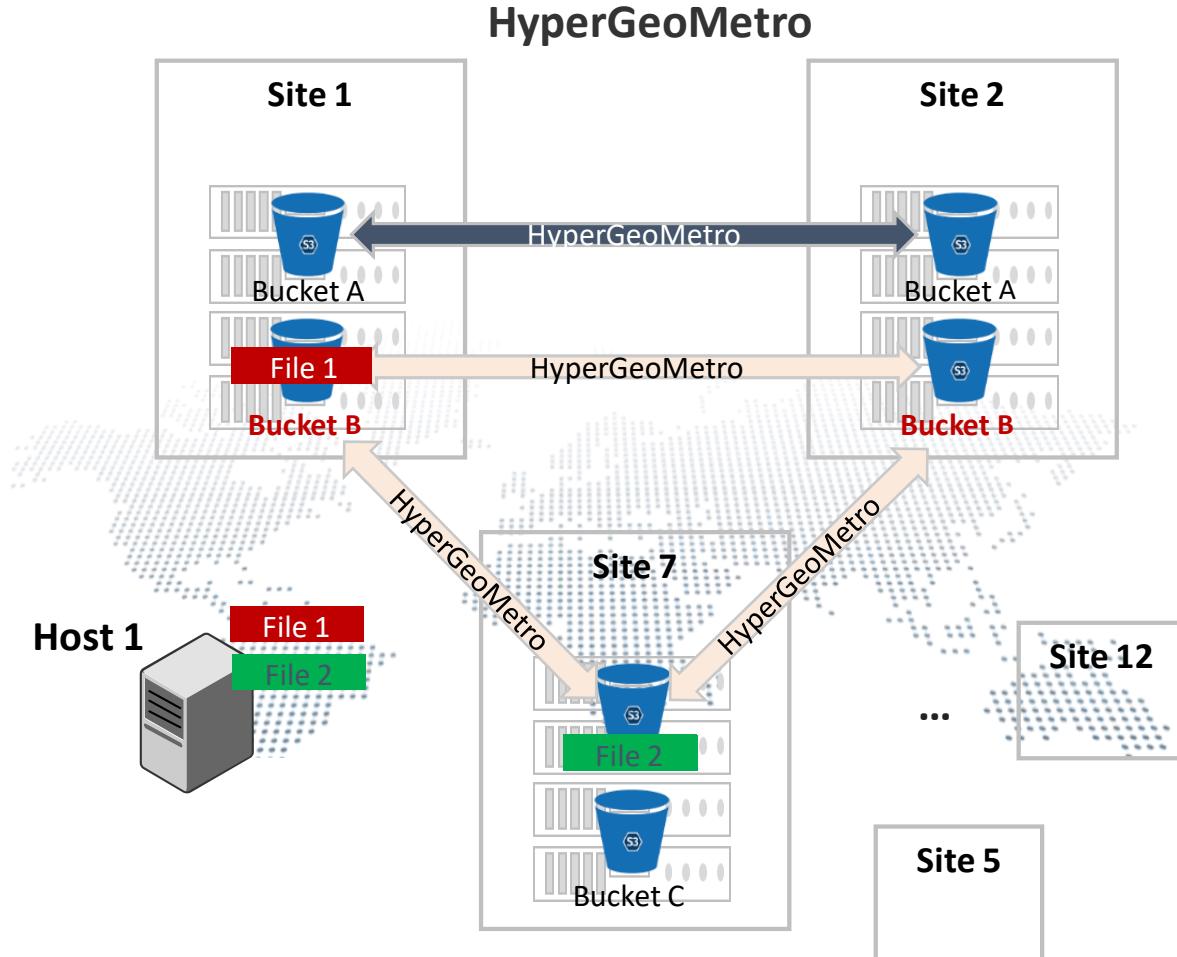


Efficient Remote Replication for Cross-Region Data Protection



- High concurrency: Multiple nodes participate in replication concurrently (a maximum of 64 nodes per cluster).
- Multi-protocol sharing, simplifying management
- Differential synchronization based on the granularity of volume, file, and object data blocks, saving bandwidth and reducing RPO
- The metadata version tracing technology is used to quickly compare differences between snapshots, reducing the RPO.
- Unstructured services support network topologies (3DC) and two-level topologies (one-to-many and many-to-one)
- Replication links support QoS control.

HyperGeoMetro: Multiple Active Sites (Multiple Copies) Across Regions Safeguard Critical Services



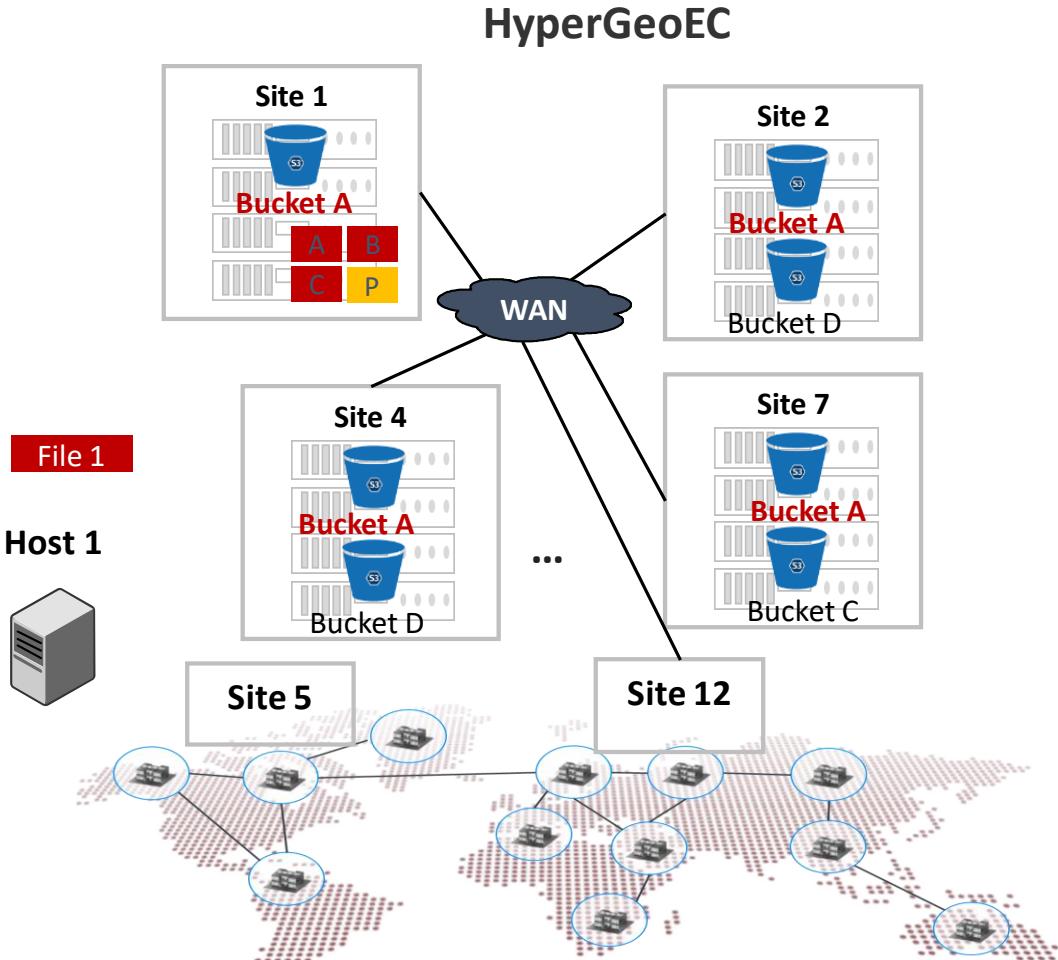
Scenario characteristics

- Each site has a data copy, enabling nearby access for optimal service experience.
- ... If one site fails, hosts can read/write to another site.
- The feature is applicable to production environments in sectors like finance, government, and healthcare.

Key Messages:

- **Up to 12 active sites**
 - Each site is in active state.
 - At least 2 sites are required. Up to 12 sites can be configured.
- **Seconds level RPO**
 - Each site maintains a copy of data, minimizing the RPO to seconds in the event of switchover upon a site failure.
- **Bucket-level flexible configuration**
 - A HyperGeoMetro policy is flexibly configured at the bucket level. And you could mix HyperGeoEC and HyperGeoMetro in same site.

HyperGeoEC: EC Across Active Sites for Site-Level DR and Cost Balancing



Scenario characteristics

- One data copy is maintained between multiple sites. The performance in cross-site read is low.
- The feature is applicable to backup/archiving environments in sectors like finance, government, and healthcare.
- A buffer storage is configured for local object data to accelerate the read performance.

Key Messages:

- **EC across up to 12 sites**
 - Each site is in active state.
 - At least 3 sites are required. Up to 12 sites can be configured.
- **Tolerates the failure of up to 2 sites**
 - Redundant data is evenly distributed across all sites to tolerate the failure of 1 or 2 sites.
 - If a node or disk fails, reconstruction is performed in the owning cluster to avoid cross-site reconstruction.
- **Bucket-level flexible configuration**
 - A HyperGeoEC policy is flexibly configured at the bucket level. And you could mix HyperGeoEC and HyperGeoMetro in same site.

System-Level Reliability: Data Redundancy Protection Tolerates Simultaneous Failures of Four Nodes Without Service Interruption

Distributed storage pool

Take N+M=8+4 for Example

file A | B | C | D || E | F | G | H | +1 | +2 | +3 | +4

write
read

Node 1		
01	02	03
E	05	06
Node 6		
01	02	03
04	05	06
Node 11		
01	02	03
04	05	06

Node 2		
01	02	03
04	05	06
Node 7		
01	02	06
04	+1	

Node 3		
01	02	03
04	05	06
Node 8		
01	C	06
04	05	06

Node 4		
01	02	03
04	05	06
Node 9		
01	02	03
04	05	06

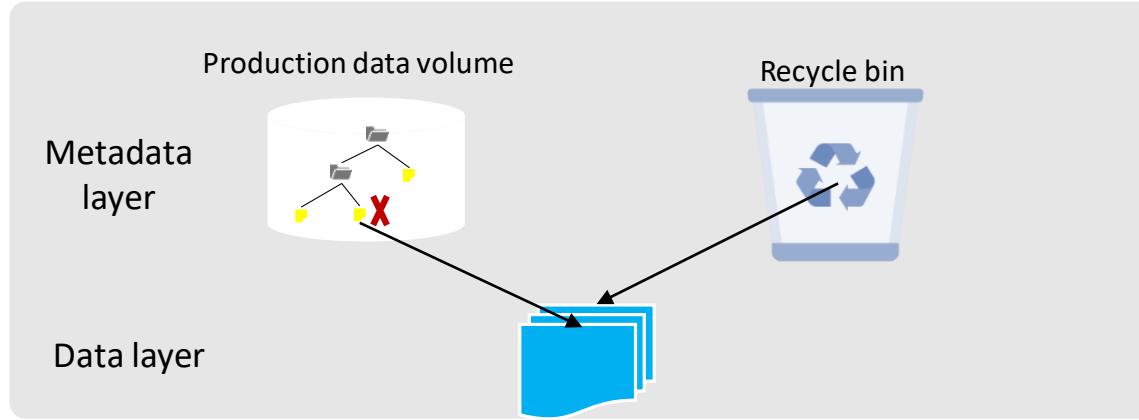
Node 5		
01	02	03
04	05	06
Node 10		
01	02	03
04	05	06
Node 14		
X	02	03
04	05	F

Cross-node EC redundancy mechanism:

- Node-level redundancy: The system tolerates simultaneous failures of a maximum of **four nodes**.
- Disk-level redundancy: The system tolerates simultaneous failures of a maximum of **four disks**.

Intelligent Anti-misdeletion, Providing Enhanced Data Reliability (for Unstructured Services)

**Deleted data is migrated to the recycle bin,
without physical data transfer.**



View data in the recycle bin in Linux.

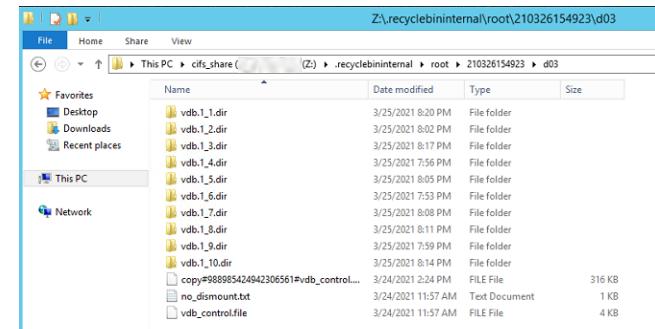
```
ctup000101759:/mnt/fs2/A/B # ls  
1.txt  
ctup000101759:/mnt/fs2/A/B # pwd  
/mnt/fs2/A/B  
ctup000101759:/mnt/fs2/A/B # rm 1.txt ----- 删除文件  
ctup000101759:/mnt/fs2/A/B # ls ..../recycleb  
ctup000101759:/mnt/fs2/.recyclebininternal # cd ..  
ctup000101759:/mnt/fs2/.recyclebininternal/root  
ctup000101759:/mnt/fs2/.recyclebininternal/root/current/A/B # ll  
total 0  
-rw-r--r-- 1 root root 0 Mar 26 17:58 1.txt ----- 进入回收站  
ctup000101759:/mnt/fs2/.recyclebininternal/root/current/A/B # cd ..  
/mnt/fs2/.recyclebininternal/root/current/A/B  
ctup000101759:/mnt/fs2/.recyclebininternal/root/current/A/B # ls  
ctup000101759:/mnt/fs2/.recyclebininternal/root/current/A/B # cd /mnt/fs2/  
1.txt A  
ctup000101759:/mnt/fs2 #
```

Annotations in Chinese:

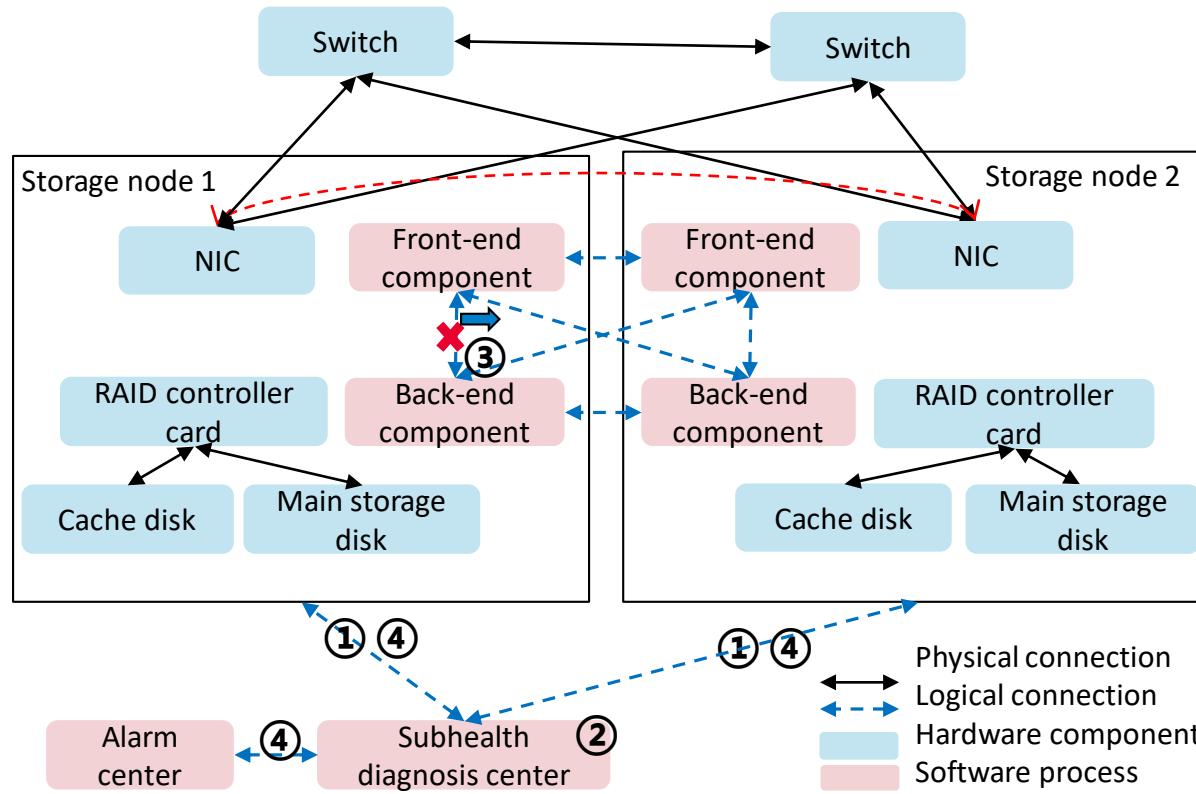
- 删除文件 (Delete file)
- 进入回收站 (Enter the recycle bin)
- 移出回收站 (Restore the file from the recycle bin)

- The recycle bin function is disabled by default. You can enable or disable it at any time.
- When receiving an explicit command for deleting a file, the system only hides the file but does not move the data. The metadata is modified in the background and the file is moved to the recycle bin.
- The retention period of files in the recycle bin is configurable. A file that has been retained in the recycle bin for a period longer than the configured period will be automatically deleted and data is permanently destroyed.
- Deletion operations using multiple protocols (S3, HDFS, NFS, and SMB) are supported.

View data in the recycle bin in Windows.



Intelligent Subhealth Detection Addresses Risks Before a Fault Occurs

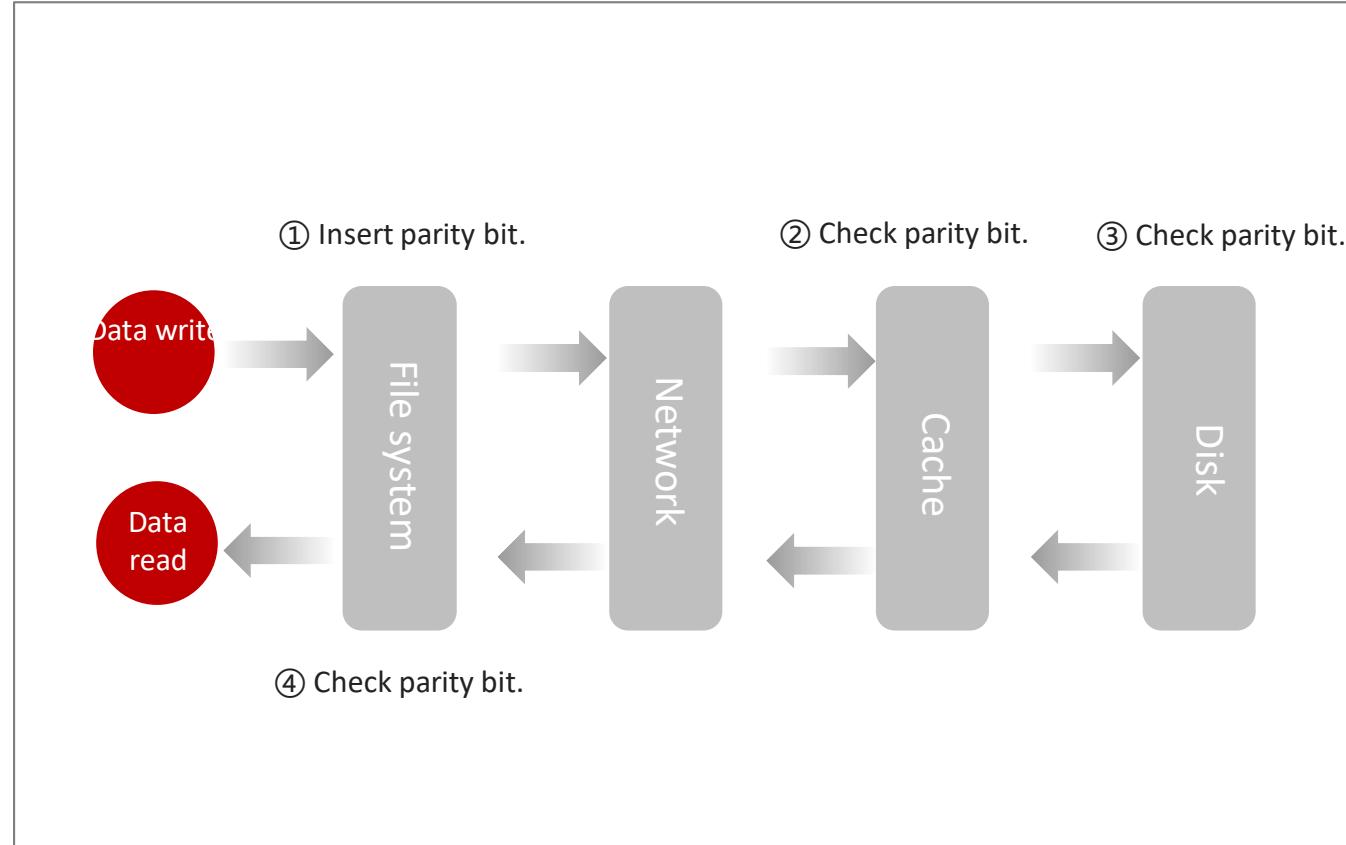


● Intelligent detection ①:

- **Disk detection:** collects disk information such as Self Monitoring Analysis and Reporting Technology (SMART), statistical I/O latency, real-time I/O latency, and I/O errors.
- **Network detection:** quickly detects exceptions such as intermittent disconnections, packet errors, negotiated rates, latency, and packet loss.

- **I/O flow detection:** checks the access I/O latency between components. If the latency exceeds the threshold, an exception is reported.
- **Intelligent diagnosis ②:**
 - **Disk diagnosis:** uses the clustering or slow-disk detection algorithm to diagnose abnormal disks or RAID controller cards.
 - **Network diagnosis:** intelligently diagnoses network port, NIC, and link exceptions based on the networking model and network exception information.
 - **Service diagnosis:** diagnoses components with abnormal latency using the majority voting or clustering algorithm based on the abnormal latency reported by each component.
- **Path-switching retry ③:**
 - **Triggering mechanism:** During the detection and diagnosis of key I/O flows, path-switching retry is performed first to ensure service continuity.
 - **Implementation mechanism:** For read I/Os, data is read from another copy or recalculated again using EC. For write I/Os, data is written to new space.
- **Isolation and warning ④:**
 - **Disk isolation:** isolates a disk according to the diagnosis result and reports an alarm.
 - **Network isolation:** isolates a network port, link, or node according to the diagnosis result and reports an alarm.
 - **Service isolation:** isolates a component according to the diagnosis result and reports an alarm.

End-to-End DIF Data Consistency Check, Ensuring Data Integrity



- **Online check**

The parity bit is written to the host. Data is checked when being written to the disk and read from the host. This addresses data damage issues brought by disk read/write offsets and NIC RAM changes.

- **Background check**

When the system load is light, data is automatically and periodically checked in the background. This helps detect and recover disk silent data corruption in advance.

- **Zero-impact self-healing**

Corrupted data is repaired using local redundant data (such as copies and EC fragments).

Quiz

1. (Multiple-choice) Which of the following reliability dimensions are supported by OceanStor Pacific series?
 - A. DC-level reliability
 - B. System-level reliability
 - C. Node-level reliability
 - D. I/O-level reliability
2. (Single-choice) Which of the following statements about HyperGeoEC is correct by OceanStor Pacific series?
 - A. One data copy is maintained between multiple sites
 - B. Only one site is in active state
 - C. At least 2 sites are required
 - D. Tolerates the failure of up to 3 sites

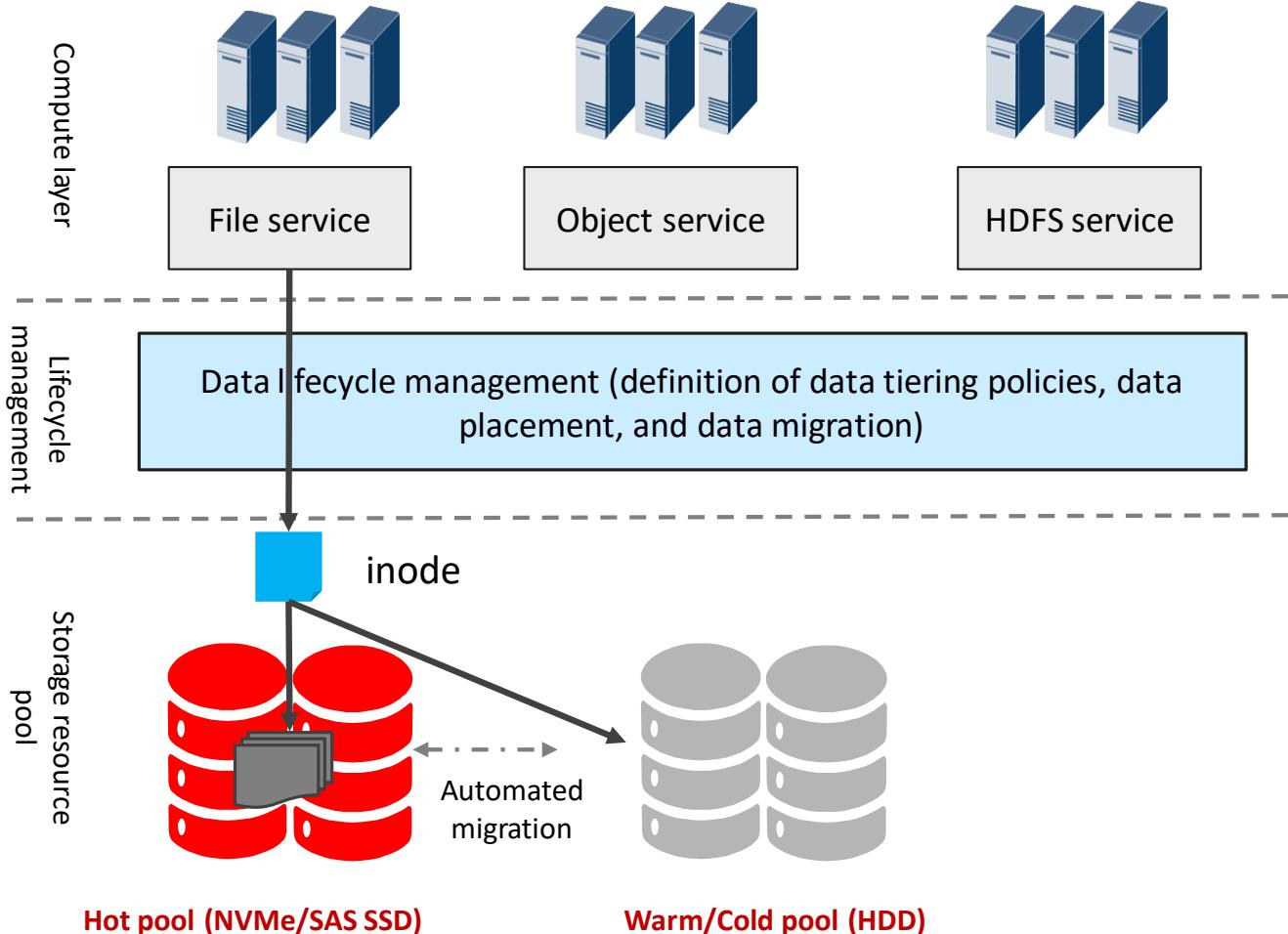
Contents

1. Product Overview
2. Hardware Architecture
3. Software Architecture
4. Superior Performance
5. High Reliability
- 6. High Efficiency**
7. Solid Security and Stability
8. Typical Scenarios

Overview and Objectives

- This section describes the high-efficiency design of Huawei OceanStor Pacific.
- On completion of this section, you will be able to:
 - Understand the key technologies of data tiring
 - Understand the key technologies of unified global metadata Indexing
 - Understand the key technologies of multi-tenant
 - Understand the key technologies of management

SmartTier – Intelligent Data Tiering, Building the Highest Cost-Effectiveness (for Unstructured Services)

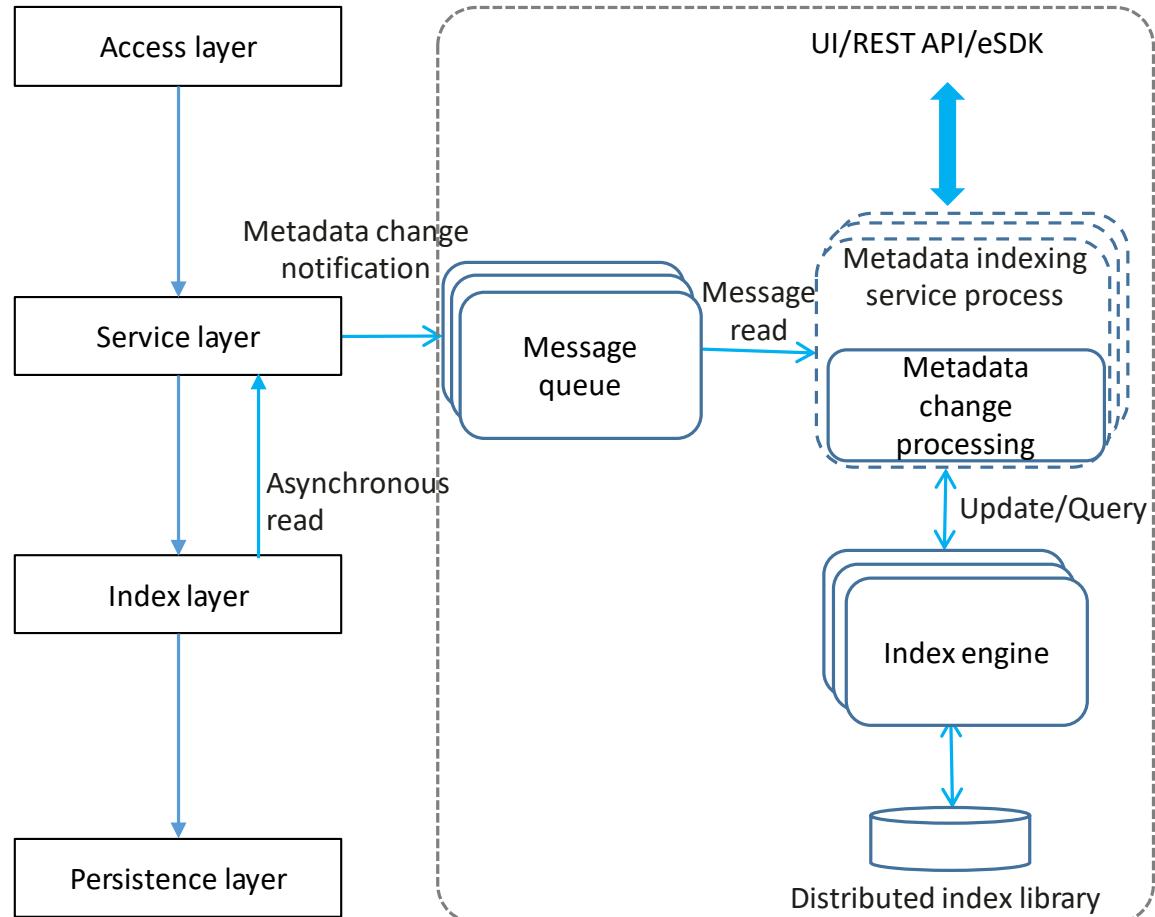


- **Supports file, object, and HDFS services.**
- **Supports two tiers of storage resource pools.** Users can configure proper storage media as required.
- Determines whether a file needs to be migrated based on the disk pool capacity watermark, file name, timestamp, size, or UID/GID.
- Supports a combination policy of multiple rules in AND or OR relationships.
- **Cross-pool data migration has no impact on upper-layer applications.**

The migration speed can be dynamically adjusted based on service loads. Data can be directly accessed after the migration without being migrated back.

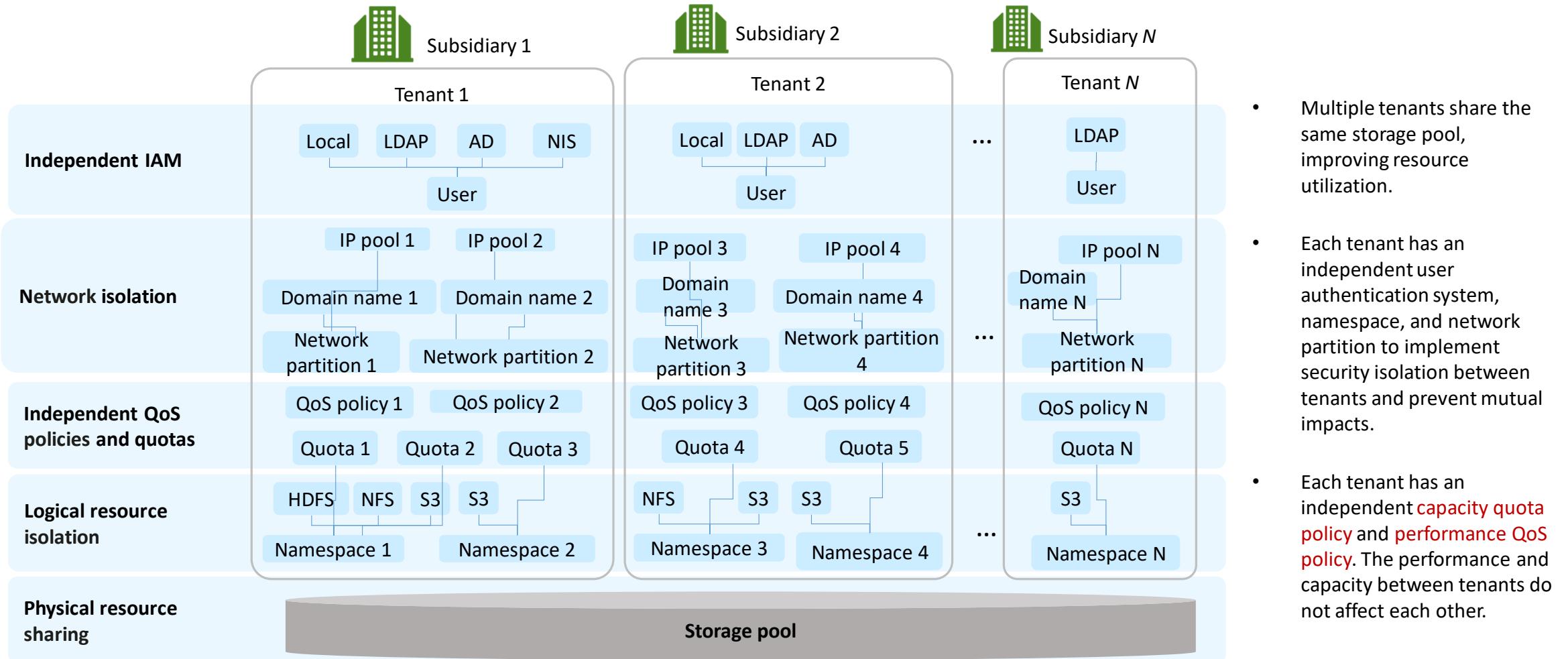
- **Supports full scan and differential scan.** Differential scan is implemented based on snapshots, and the scan speed is irrelevant to the total data volume.
- **Full lifecycle management** controls new data writes, hot and cold data migration, and deletion of expired data.

Unified Global Metadata Indexing, Implementing Efficient Data Management (for Unstructured Services)

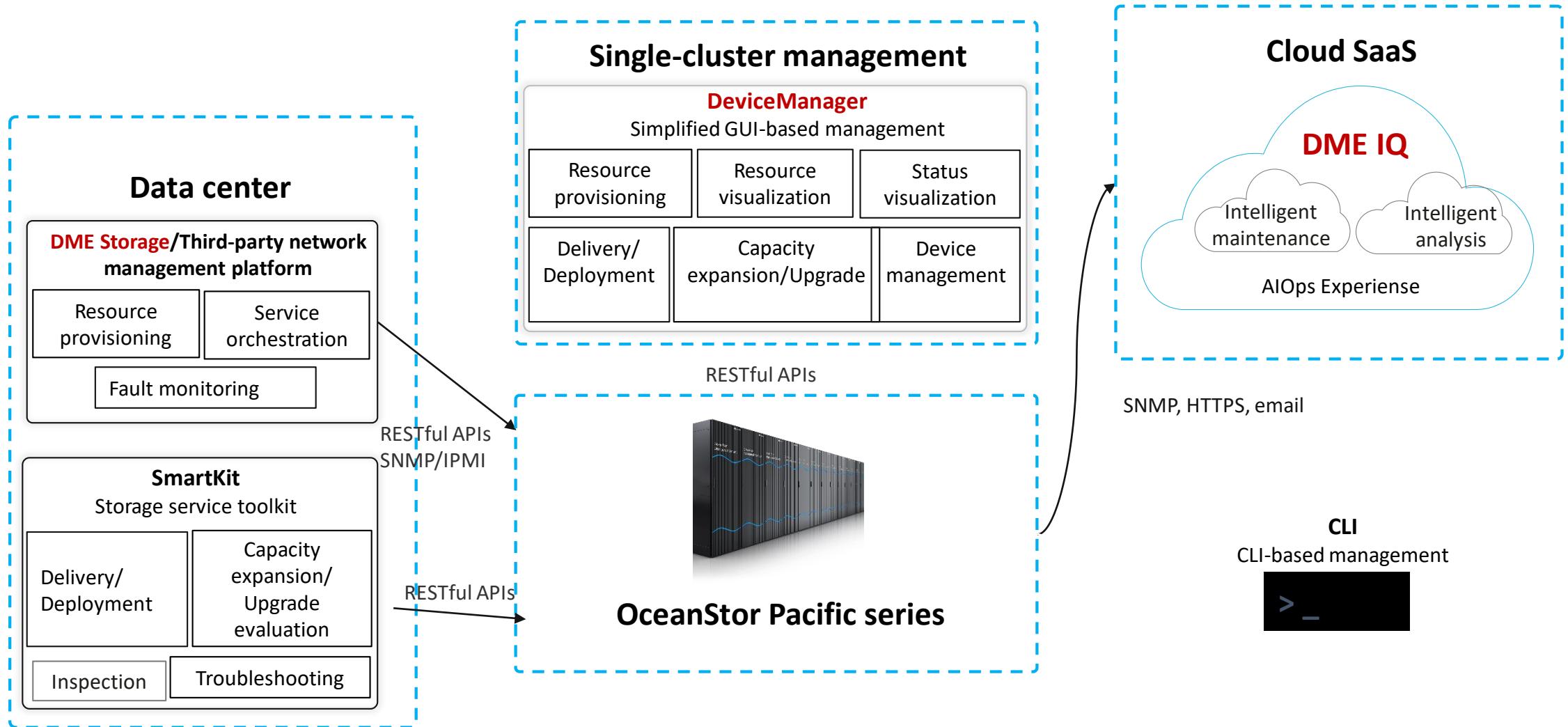


- **Unified indexing of unstructured data**
Supports three types of metadata: object, file, and HDFS.
- **Metadata change, proactive asynchronous notification, localized processing**
Local affinity avoids cross-node network interactions, helps efficient data processing, and reduces the data query latency to 30 seconds.
- **Distributed concurrent processing of user query requests**
The index engine is deployed in distributed mode to avoid query bottlenecks.
- **Rich search attributes**
Include system metadata, custom metadata fields (up to 30), and object tags.
- **Rich search conditions**
By default, attributes of the character type support prefix matching, and attributes of the numeric type support range and size condition matching.

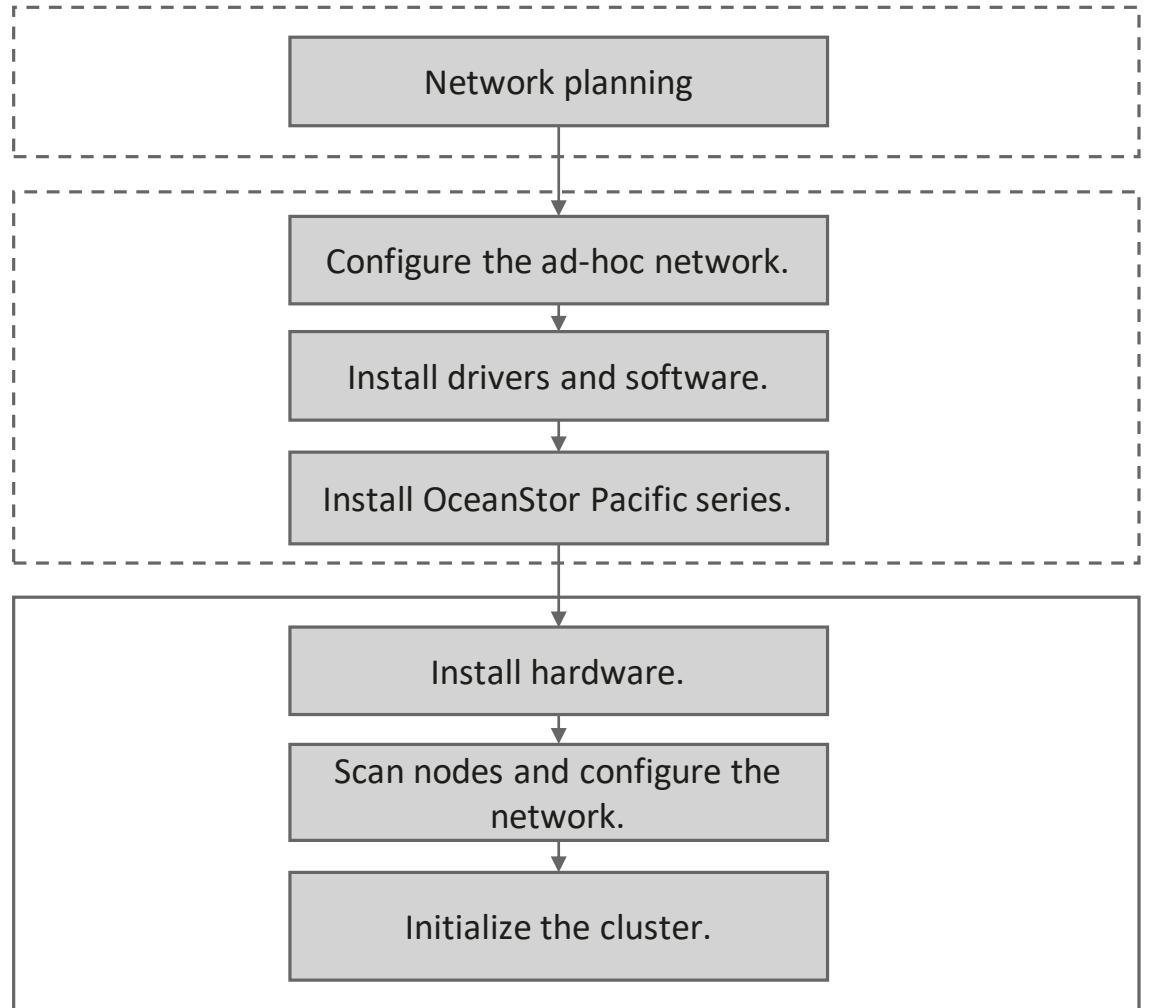
Multi-Tenant – Resource Sharing, Security Isolation, and Performance and Capacity Isolation (for Unstructured Services)



Efficient and Easy-to-Use Cluster Management

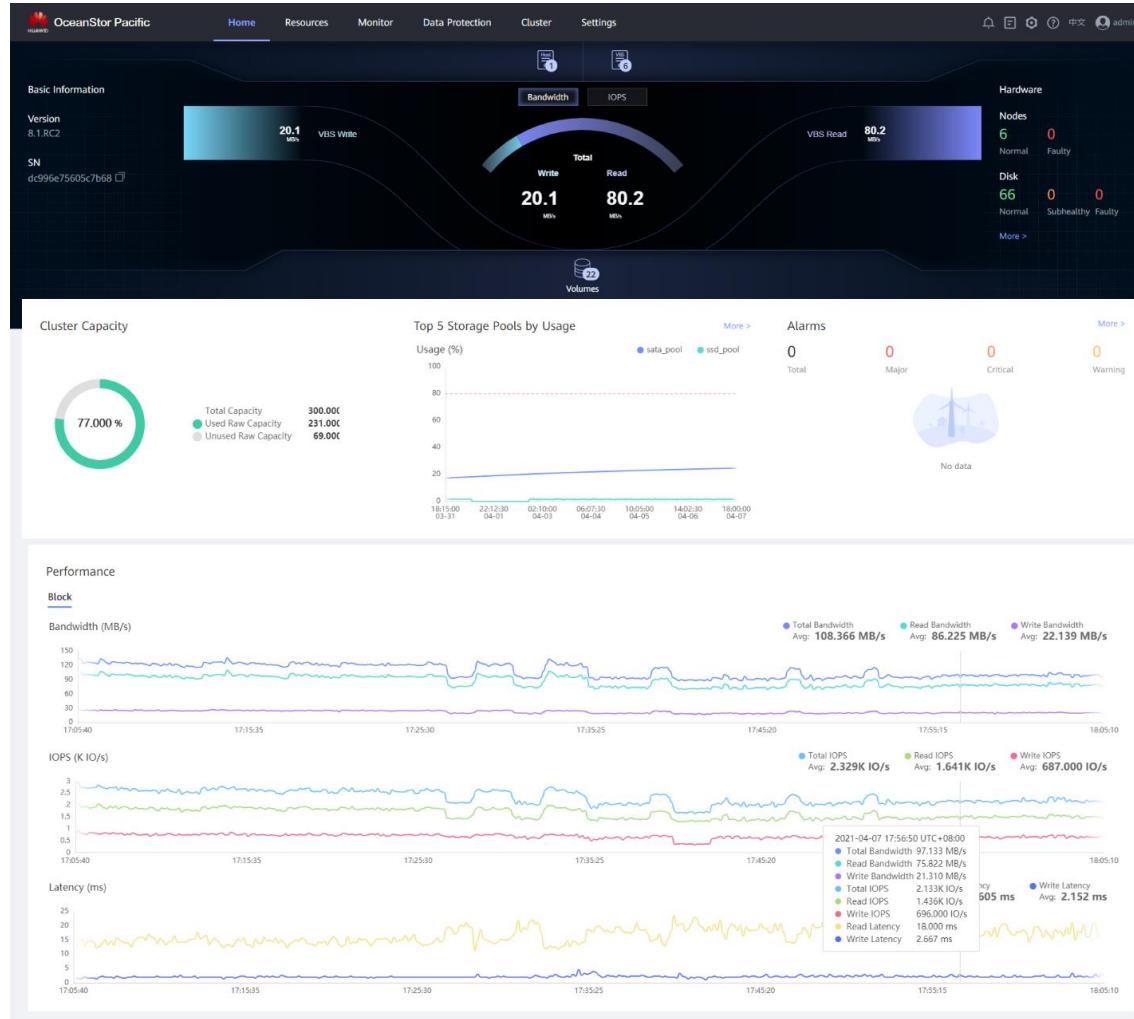


Automatic Deployment: Out-of-the-Box Implements Deployment Within 1 Hour



- The LLD online tool provides wizard-based and template-based network planning, which takes only **30 minutes**.
- **In-depth pre-installation** is completed before delivery.
- The deployment assistant tool SmartKit supports batch node discovery, basic network information configuration, and automatic firmware and driver updates. The cluster deployment can be completed **within 1 hour**.

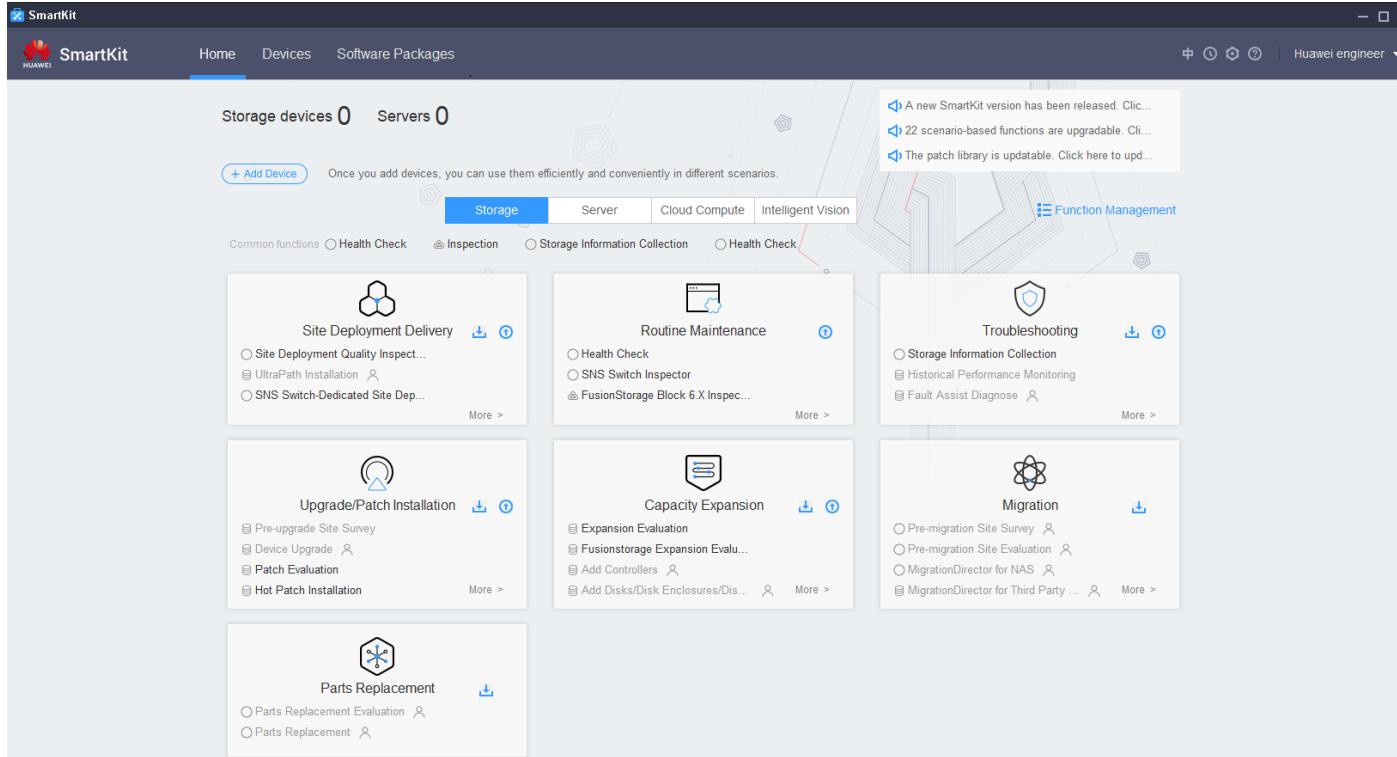
DeviceManager: Built-in Efficient Management Portal



Built-in management software DeviceManager

- Simplifies configuration and facilitates service provisioning, service protection, and system configuration.
- Enables converged management of unstructured data.
- Implements online fast upgrade and smooth capacity expansion and reduction.
- Manages network topologies in a visualized manner.

SmartKit Simplifies Operations in Complex Scenarios



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 in one-stop manner.

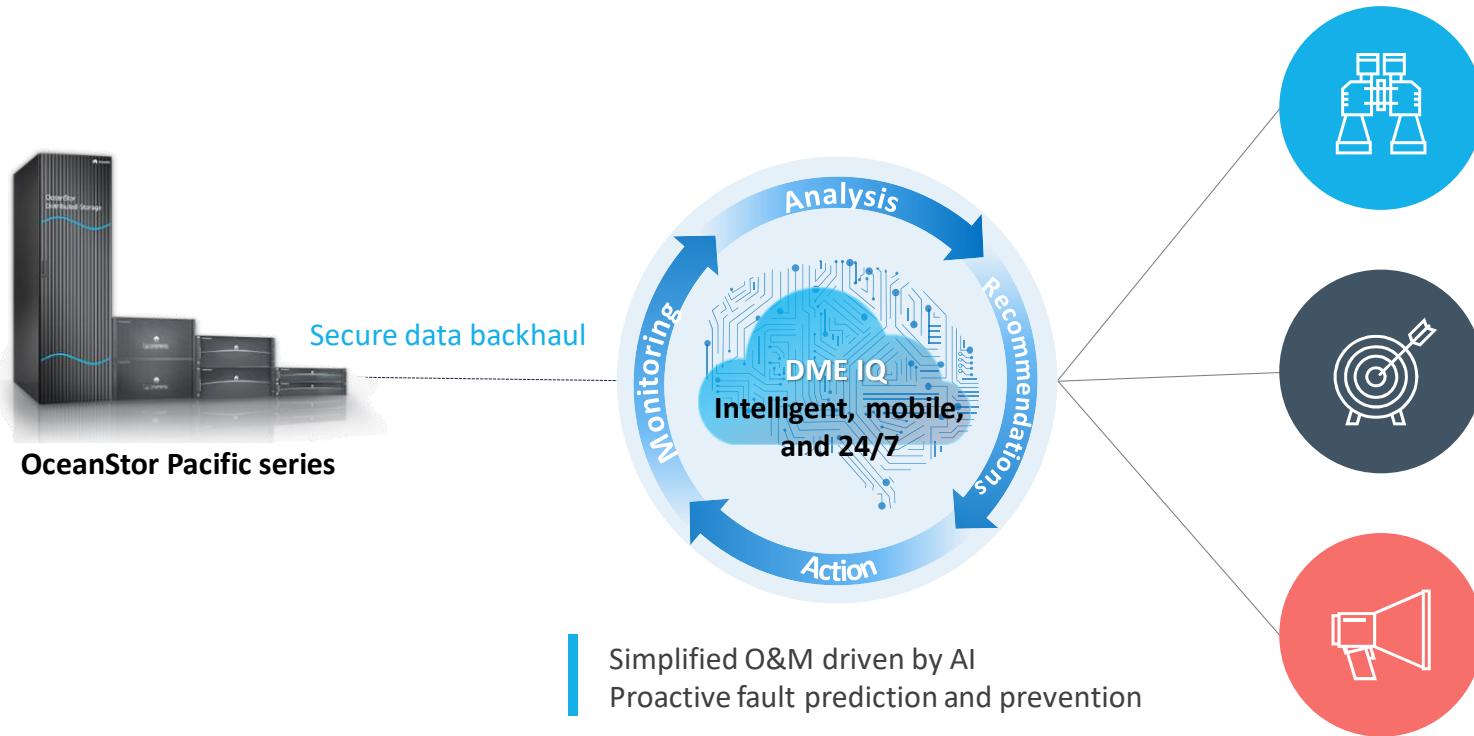


3. Standardized operations

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

Huawei SmartKit provides a unified service tool platform for products in the storage, server, and cloud computing fields. The platform contains various tools required for IT device deployment, maintenance, and upgrade to help users, service engineers, and maintenance engineers perform precise operations on devices during the preceding processes, simplifying operations and improving work efficiency.

DME IQ Implements Proactive, Predictive, and Intelligent O&M



Intelligence

- ✓ Capacity prediction: predicts the capacity trend in 12 months and reports alarms.
- ✓ Disk fault prediction: predicts HDD faults 14 days in advance.

Mobile app

- ✓ Message notification: notifies device alarms and risks in real time.
- ✓ Fault analysis: displays fault details to understand the impact scope.

All-day

- ✓ Intelligent monitoring: 24/7 monitoring and automatic trouble ticket creation
- ✓ Troubleshooting: 1000+ fault modes, quickly locating root causes

Huawei DME IQ is a cloud-based intelligent O&M platform. It leverages big data analysis and AI technologies to provide services such as automatic fault reporting, capacity prediction, performance prediction, disk risk prediction, and fault handling progress tracking for data infrastructures such as Huawei storage devices and servers. Such proactive and predictive O&M anytime and anywhere for data infrastructure simplifies O&M and improves O&M efficiency.

Quiz

1. (True or False) In order to enhancing Data Management efficiently, OceanStor Pacific uses Unified Global Metadata Indexing feature.
2. (Multiple-choice) What policy can OceanStor Pacific Multi-Tenant feature set to each tenant ?
 - A. Capacity quota
 - B. Performance QoS
 - C. Domain name
 - D. Directory

Contents

1. Overview
2. Overall System Design
3. Superior Performance
4. High Reliability
5. High Efficiency
- 6. Solid Security and Stability**
7. Typical Scenarios

Overview and Objectives

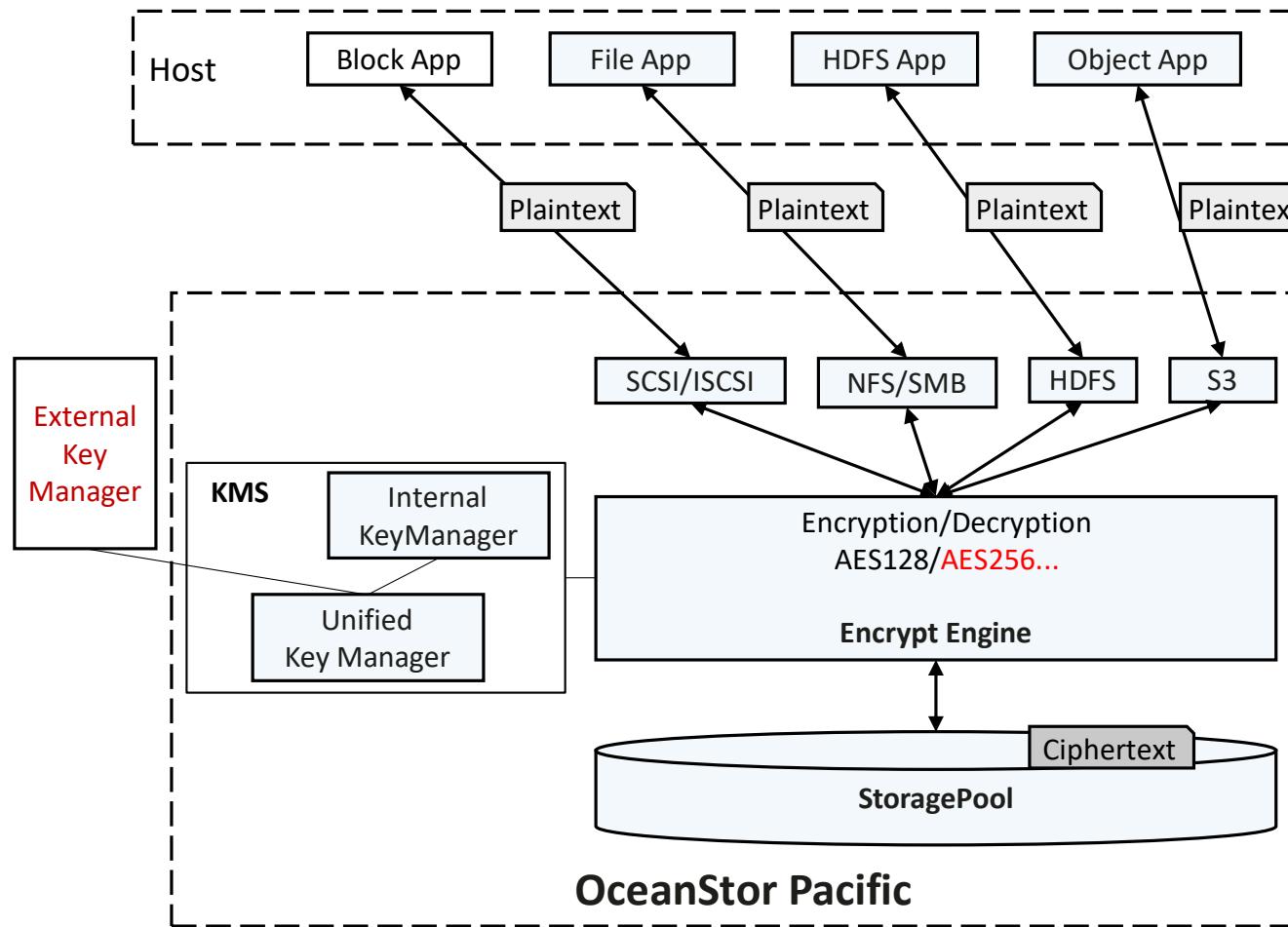
- This section describes the security and stability design of Huawei OceanStor Pacific.
- On completion of this section, you will be able to:
 - Understand the full-stack security design
 - Understand the key features and technologies of security design

Full-Stack Security Design Implements In-depth Defense and Provides All-round Data Protection

Risks

Key capabilities							
		Service access authentication	Management access authentication	Access authentication	Certificate management	Security audit	
Host access risk Unauthorized access by management personnel	Unauthorized access prevention	CHAP authentication AK/SK authentication Kerberos authentication AD/LDAP/NIS domain authentication Pre-shared key authentication	Two-factor authentication AD/LDAP domain authentication Single sign-on (SSO)	Default role RBAC Password complexity Forced password change upon first login Weak password dictionary	Unified certificate management (import, revocation, and expiration alarms) Automatic certificate application and issuance	Reporting alarms and security logs in syslog format	
Data security risk Tampering, spoofing, and leakage	Data leakage prevention	Static data encryption Self-encrypting disk (SED) Software-based encryption of unstructured data	Key management External/Internal Key Manager	Data destruction Key destruction Tool destruction	Service isolation NAS multi-tenancy	Secure communication HTTPS TLS 1.2	Security audit Log audit
Software risk Tampering and virus implantation during release, installation, and running	Software anti-tampering	Software package integrity Integrity verification for the software package (manual) Digital signature verification for inner software (automatic)		Secure boot Digital signature-based integrity verification during system startup			Security Privacy Resilience Compliance review
Security management risk Unauthorized operations by management personnel, such as stopping services and destroying data	Resilient system	Traffic overload QoS policy Overload	Security protocol TLS 1.2 SSH V2 FTPS	Vulnerability protection NSFOCUS, GSM, and AWVS scanning NMAP tool scanning	System security EulerOS hardening Binary security compilation options Minimum permissions (without the root permission for processes)		

Software-based Data Encryption (Data-at-Rest) Prevents Service Data Leakage



Unified data software encryption platform, supporting structured and unstructured services

Remarks: Block service encryption and deduplication and compression are mutually exclusive.

Functions

- Implements software-based static data encryption (Data-at-Rest).
- Data encryption can be flexibly enabled by **tenant, LUN, and namespace**.
- Supports all structured and unstructured standard protocols (SCSI/ISCSI/NFS/SMB/S3/HDFS).
- Supports international standard algorithms AES128-XTS/**AES256**-XTS and AES-NI instruction acceleration.
- The TLS certificate-based secure channel ensures key transmission security. The certificate can be updated.

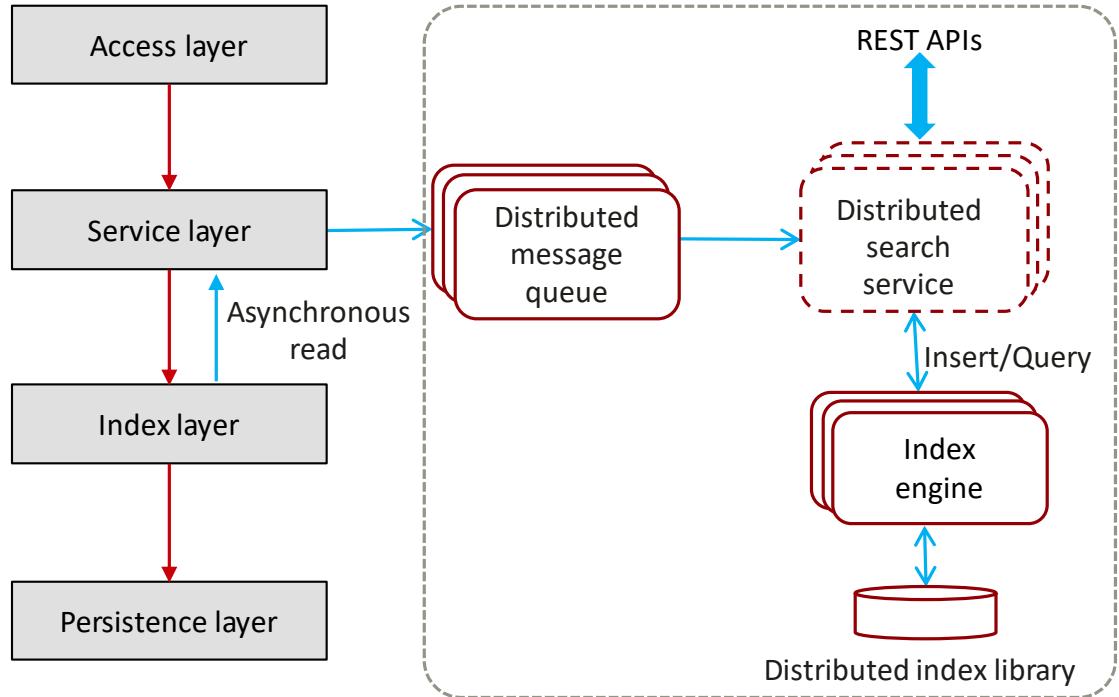
Key Manager Service

- **Internal Key Manager** and **External Key Manager** are supported.
- External Key Manager: complies with the KMIP protocol and supports KMIP1.0, KMIP1.1, and KMIP1.2.
- Internal Key Manager: The secure random number generates access keys. It uses a three-layer architecture to protect key security and supports key backup, restoration, update, and destruction.

Encryption and decryption processes

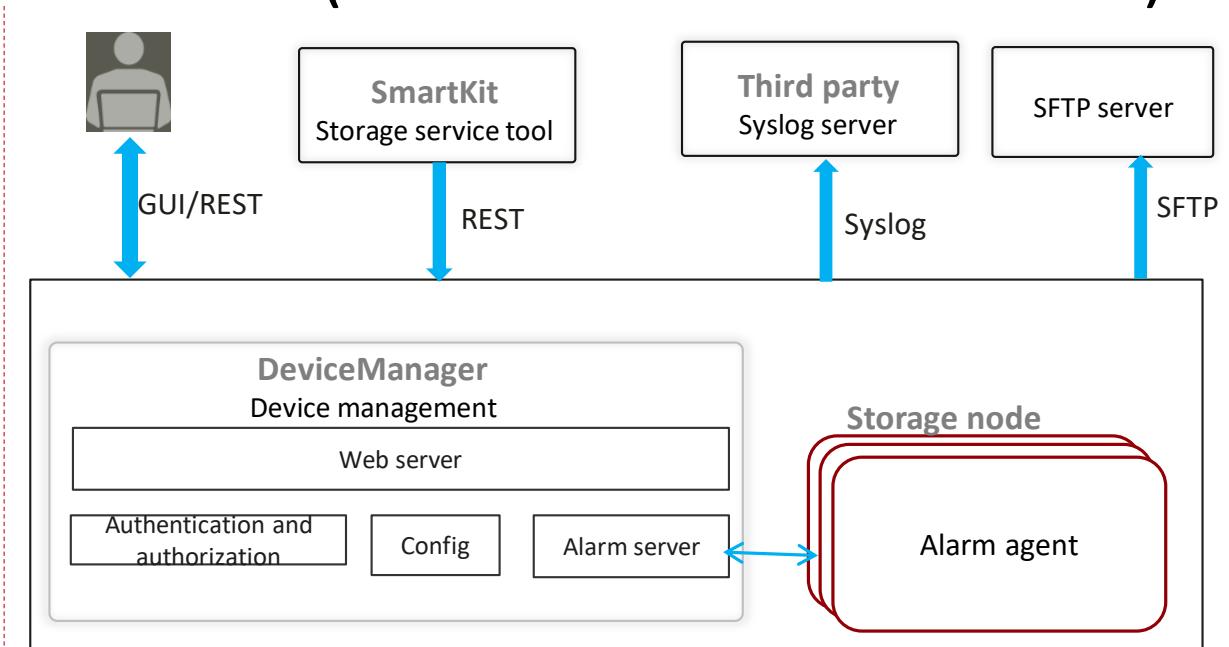
- Obtain the plaintext AK from Internal Key Manager and parse the ciphertext DEK into the plaintext DEK.
- During a write operation, the encryption engine adopts the plaintext DEK to encrypt the written plaintext data into ciphertext data and writes the ciphertext data into the storage pool.
- During a read operation, the encryption engine adopts the plaintext DEK to decrypt the ciphertext data read from the storage pool into plaintext data and returns the plaintext data to the application.

Log Audit Makes All Data Operations Traceable (for Unstructured Services)



Protocol access audit log

- Audit logs can be recorded for multi-protocol operations on unstructured data (NFS/SMB/S3/HDFS).
- The audit function can be enabled or disabled based on namespaces.
- The operation words of audit logs can be set based on namespaces. Creation, deletion, and renaming are supported.
- The audit logs of multi-protocol data access operations are recorded in a unified format.
- REST API search and multi-field search (event ID, name, protocol type, time segment, and main IP address) are supported.
- High reliability and data consistency: Audit logs and metadata/data operations are processed in the same transaction to ensure the consistency of audit records and operations.



System management audit log

- All operations involving system modification are recorded in audit logs.
- The log content supports post-event audit, including the user ID, time, event type, name of the accessed resource, address or ID of the access initiator, and access result.
- Management audit logs can be searched, filtered, and exported on the GUI.
- Management audit logs can be interconnected with third-party log analysis tools through the syslog protocol.
- Management audit logs can be dumped to the log archive server using SFTP. The log retention period meets requirements.
- Management audit logs are transmitted using secure transmission protocols, such as HTTPS, SFTP, and TLS1.2.

Quiz

1. (Multiple-choice) Which elements can be set by Data encryption of OceanStor Pacific?
 - A. Tenant
 - B. LUN
 - C. Directory
 - D. Namespace

Contents

1. Overview
2. Overall System Design
3. Superior Performance
4. High Reliability
5. High Efficiency
6. Solid Security and Stability
- 7. Typical Scenarios**

Overview and Objectives

- This section describes the typical scenarios of Huawei OceanStor Pacific.
- On completion of this section, you will be able to:
 - Understand the scenarios/requirements/advantages of OceanStor Pacific in the HPDA field
 - Understand the scenarios/requirements/advantages of OceanStor Pacific in object resource pools

Storage Solution Oriented to the Next-Generation HPDA

HPDA Scenarios

- Oil & Gas exploration
- Autonomous driving
- Life science
- Supercomputing centers

High-density design

2.7X higher capacity density

1.3X higher performance density

	Vendor E	Vendor D	HUAWEI
1.5 PB 12 U, 108 disks	24GBps/U	3.75GBps/U	32GBps/U
General-purpose high-density servers	5 U, 120 disks	1.68 PB	HUAWEI

Less footprint:
multiple cabinets -> 1 cabinet

Seamless multi-protocol interworking

Category	Vendor E	Vendor D	OceanStor Pacific
Native semantics	√	✗	√
Gateway-free	√	✗	√
Semantic integrity	Intermediate	Low	High
Impact on performance	Intermediate	Large	Small

High efficiency:
multiple data copies -> 1 data copy

Hybrid-workloads Oriented

Single-stream write IOPS, 20% higher than Vendor D

1 device for N phases

Seismic data processing Seismic data interpretation

High bandwidth High OPS

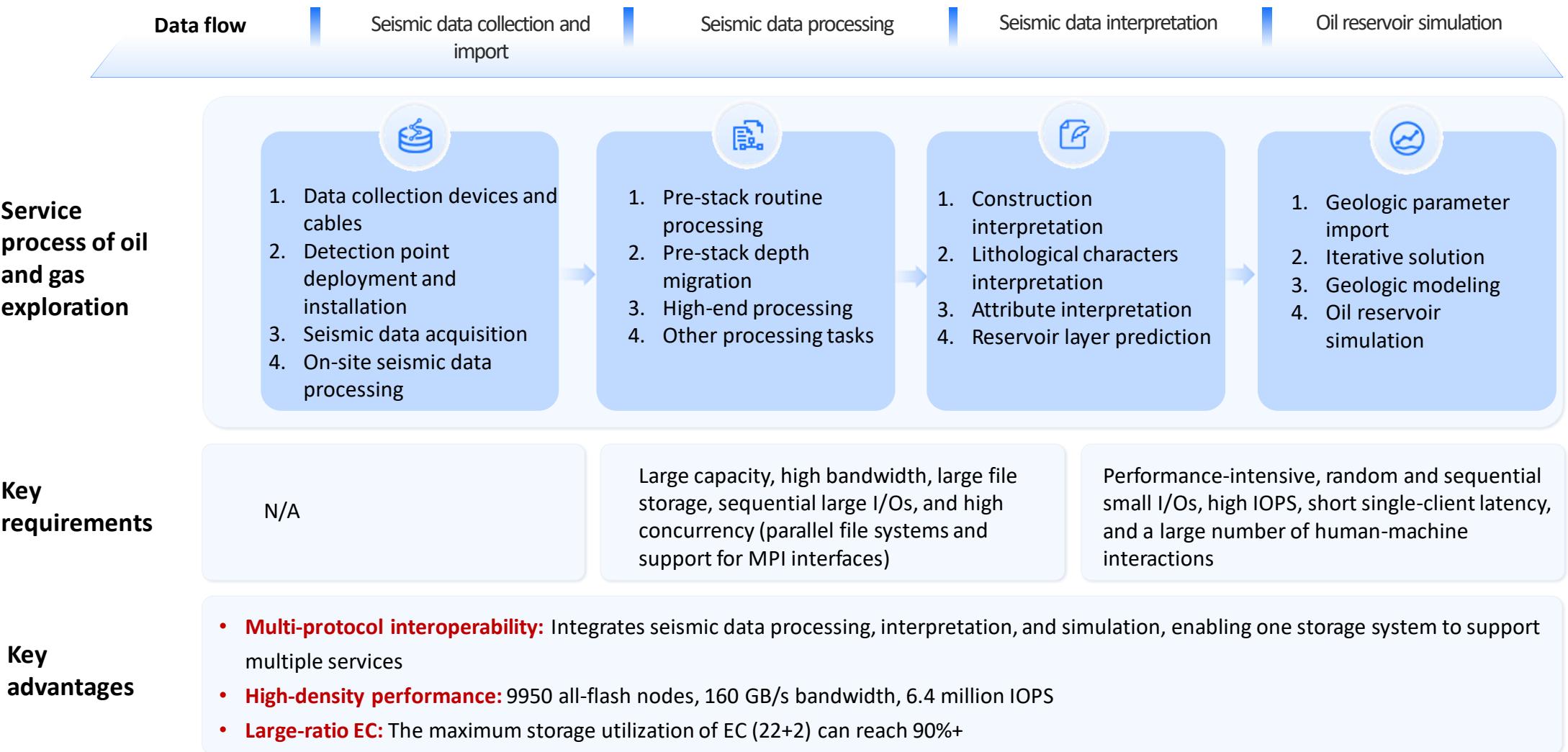
POSIX NFS/CIFS

Distributed parallel client (DPC)

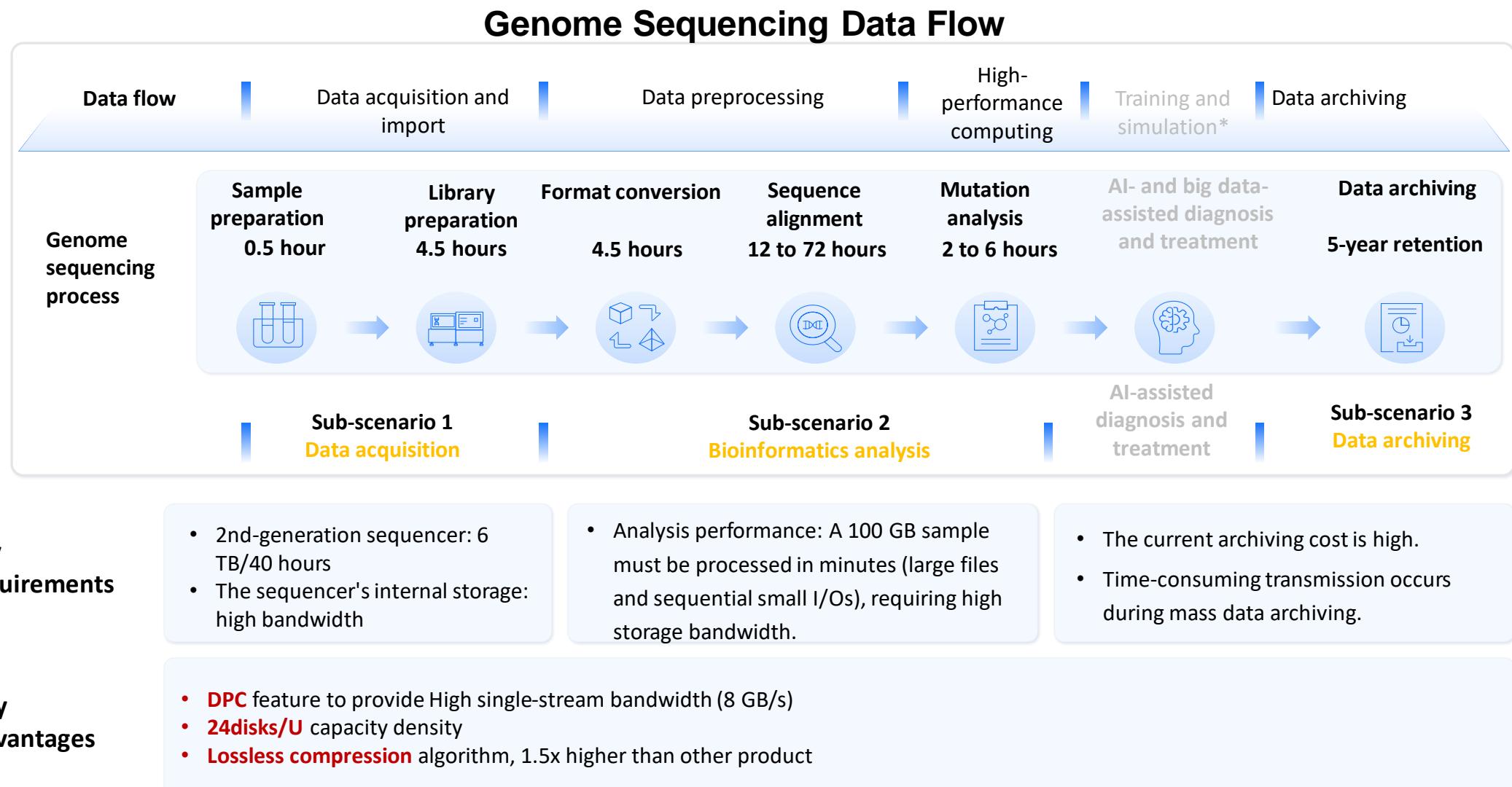
RDMA

Simple management:
Multiple devices -> 1 device

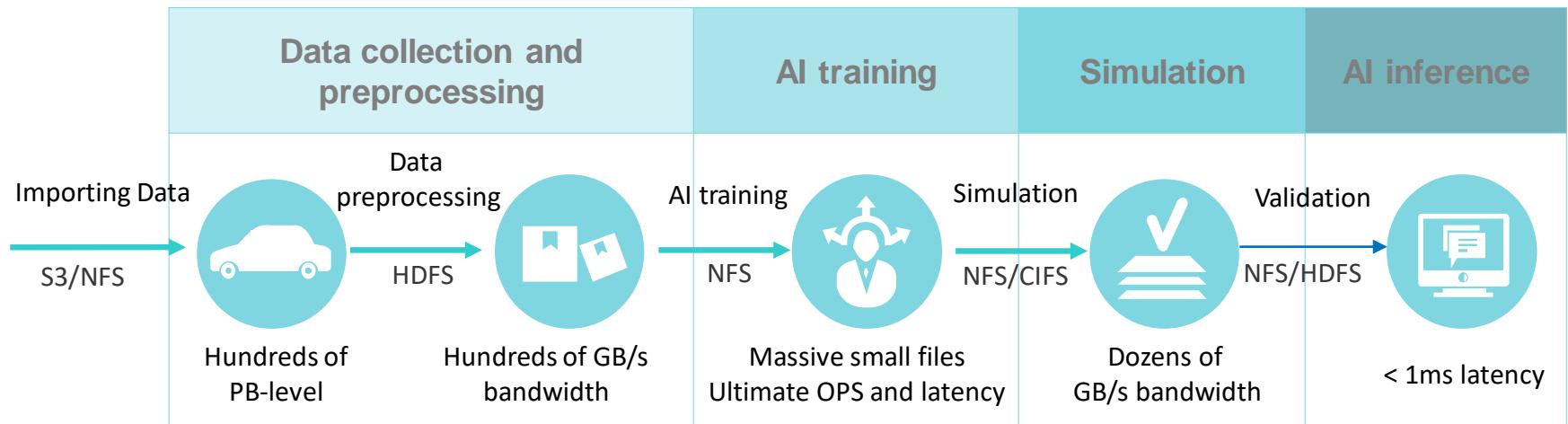
HPDA Scenario 1: Oil & Gas Exploration



HPDA Scenario 2: Life science-Genome Sequencing



HPDA Scenario 3: Autonomous driving

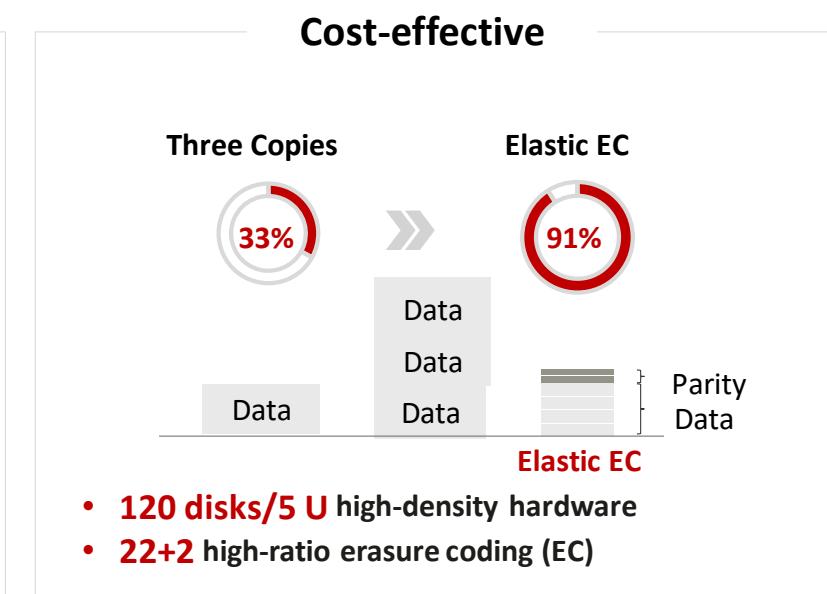
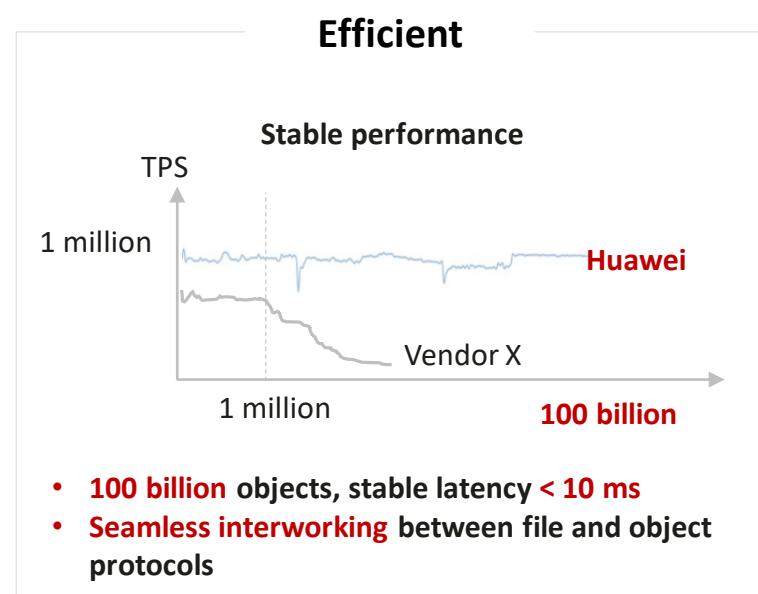
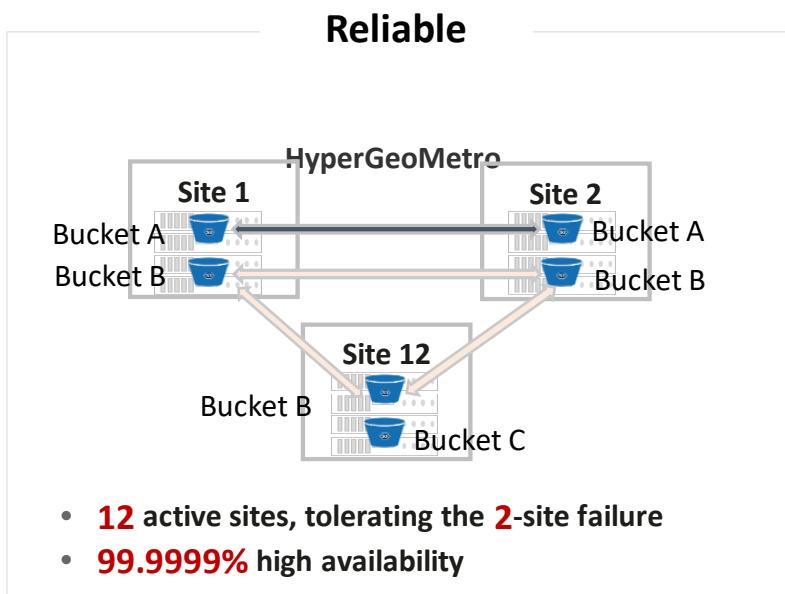
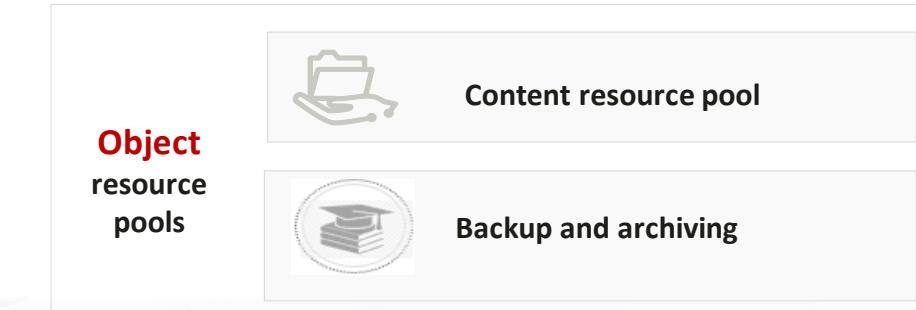


Evolution from L3 to L4
50-fold mileage increase
and 50-fold data growth

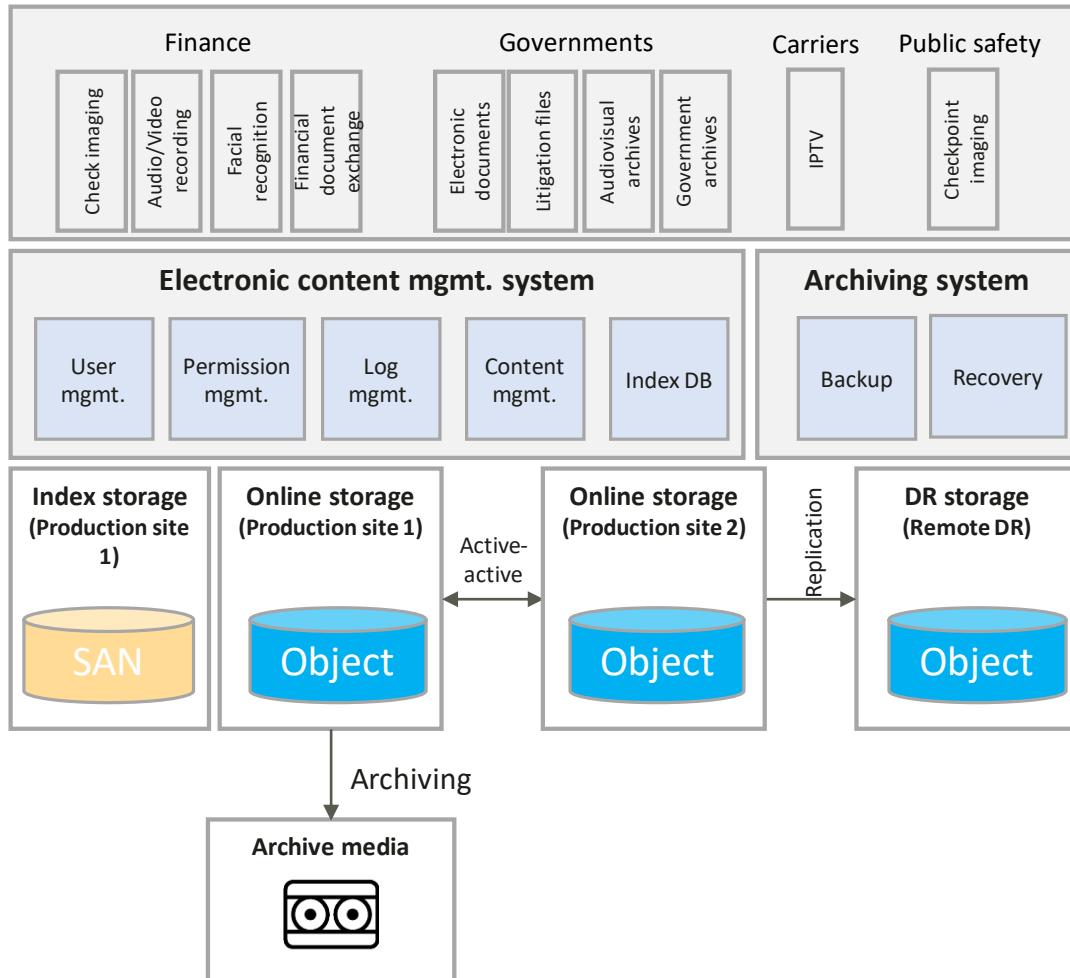
- 64TB / vehicle / day
- Hundreds of PB-level data storage requirements
- S3, NFS, and HDFS interfaces
- Massive data copy and low analysis efficiency
- High bandwidth and high OPS
- Ultra-low latency and high performance challenges

- One storage system serves the entire process and implements **multi-protocol interworking** with zero data copying for 25% higher analytical efficiency
- One storage system offers up to **32GB/s bandwidth and 400K IOPS per U** for 30%+ higher performance than the industry's next best player
- Intelligent tiering of hot and cold data, **24 disk slots/U high-density hardware**, and full-lifecycle data management reduce TCO by 20%

Object Storage Solution for Massive Content Resource Pools and Backup & Archiving



Object Resource Pool Scenario 1: Content Resource Pool



Current situation and challenges

- Poor service experience:** The CAGR of electronic document data (from governments, public security organs, procuratorates, courts, healthcare PACS, and financial document imaging) is **over 20%**. The file quantity increases **from tens of millions to billions**. The current enterprise/distributed NAS is inefficient in retrieval. The latency of **over 200 ms** affects service experience.
- Difficult backup:** It takes **several days** to back up and recover data, which cannot meet service requirements.

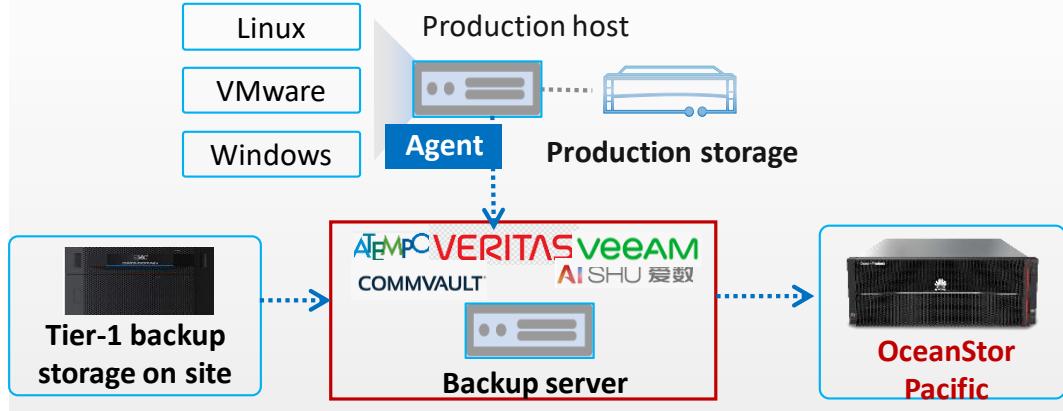
Why object storage

- Efficient retrieval:** NAS uses a tree metadata structure, which requires **multiple I/O interactions** in retrieval of **hundreds of millions** of files. Object storage uses the hash algorithm and a flat structure to ensure **stable performance** with just **one I/O interaction**.
- Free of data backup:** NAS requires data backup. Object storage uses versioning and multi-site DR to recover corrupted data.

Why Huawei

- Stable performance:** Tens of billions of objects with data read/write latency < 50 ms (Sea, Singapore)
- 99.999% availability:** 2 to 3 active sites across regions, ensuring service continuity
- Backup-free:** Versioning, bucket-level snapshot, enterprise-class prevention of accidental deletion, and batch data recovery
- Multi-protocol interworking,** allowing for coexistence of old and new services

Object Resource Pool Scenario 2: Backup



Scenario: OceanStor Pacific acts as tier-2 backup storage (backup software migrates data from tier-1 backup storage to OceanStor Pacific)

Current situation and challenges

- **Siloed construction:** Each application system has its own backup system. Resources cannot be efficiently used.
- **Large amount of data:** Production data keeps increasing, and backup data is retained for a long time.
- **Fast recovery required:** Production services must run 24/7, requiring fast fault recovery.
- **High reliability required:** Backup is the last protection for production, requiring that any data within the retention period can be restored.

Why object storage

- The distributed architecture delivers better scalability than centralized devices Object storage supports cross-site EC, achieving optimal reliability and cost.
- Flat storage of object data without file system fragmentation delivers stable performance in processing mass data.

Why Huawei

- **High performance:** 2 GB/s read/write bandwidth per node, which increases linearly as more nodes are added
- **Stable performance:** Tens of billions of objects with data read/write latency < 50 ms (Sea, Singapore)
- **99.9999% availability:** 2 to 3 active sites across regions, ensuring service continuity
- **High reliability:** Versioning, bucket-level snapshot, enterprise-class prevention of accidental deletion, batch data recovery, multi-site DR, and WORM

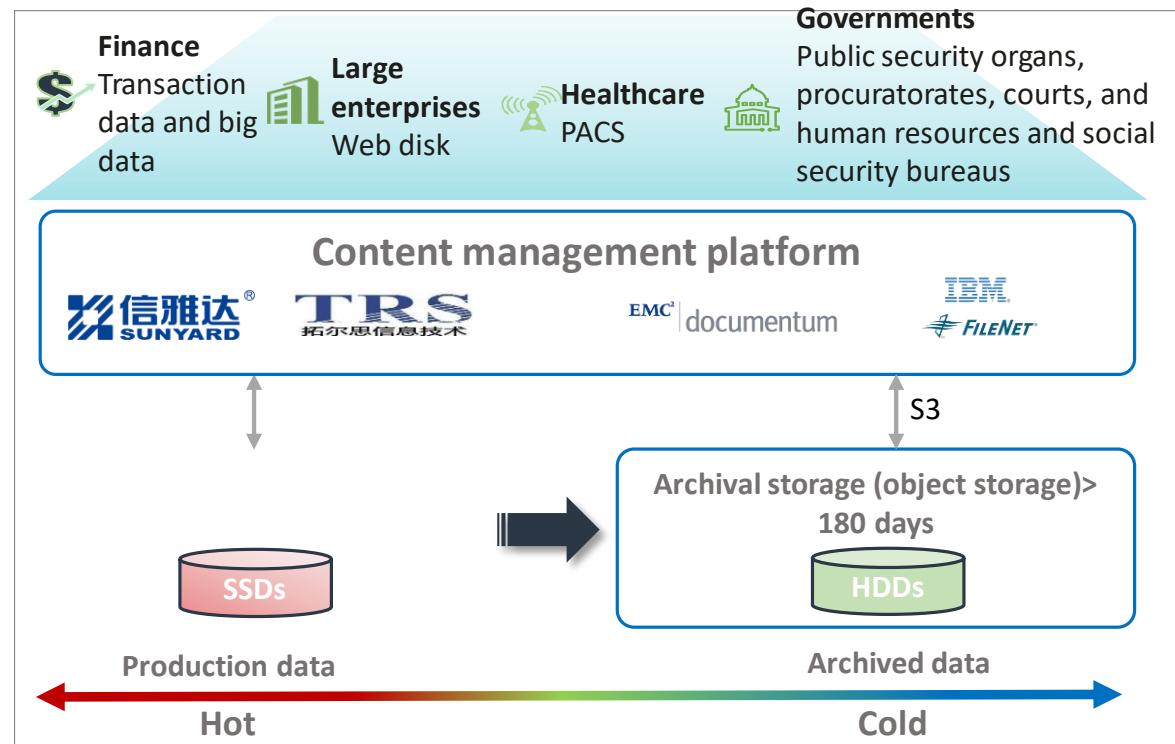
Object Resource Pool Scenario 3: Online Archiving

Scenario

Massive volumes of infrequently accessed data are archived online for a long time to meet regulations.

Industry opportunities

- Finance: Document imaging and audio/video recording
- Healthcare: PACS imaging and digital pathology
- Government: Public security organs, procuratorates, courts, and human resources and social security bureaus



Current situation and challenges

- **Cost:** The file quantity increases from tens of millions to hundreds of millions, and the cost of tier-1 all-flash storage is high.
- **Reliability:** Multi-site DR and anti-tampering are required.
- **Experience:** With massive volumes of files, the current enterprise/distributed NAS is inefficient in retrieval. The latency of over 200 ms affects service experience.

Why object storage

- NAS uses a tree metadata structure, which requires multiple I/O interactions in retrieval of hundreds of millions of files. Object storage uses the hash algorithm and a flat structure to ensure efficient retrieval with just one I/O interaction.
- Object storage's versioning and WORM work with multi-site DR to recover corrupted data, eliminating the need for data backup.

Why Huawei

- **Stable performance:** Tens of billions of objects with data read/write latency < 50 ms (Sea, Singapore)
- **99.9999% availability:** 2 to 3 active sites across regions, ensuring service continuity
- **High reliability:** Versioning, bucket-level snapshot, enterprise-class prevention of accidental deletion, batch data recovery, multi-site DR, and WORM

Quiz

1. (Single-choice) Which one is not the advantage of OceanStor Pacific in HPDA scenario?
 - A. High-density hardware
 - B. Seamless multi-protocol interworking
 - C. Smart Card
 - D. Hybrid-workloads Oriented
2. (True or False) OceanStor Pacific Object Storage Solution focus on content resource pool and Backup & archiving scenario.

Summary

- This chapter describes the hardware architecture, software architecture, performance, reliability, efficiency, and security features and principles of the OceanStor Pacific distributed storage system. It also describes the requirements and advantages of the OceanStor Pacific distributed storage system in major application scenarios.

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 OceanProtect Backup Storage Key Technologies and Best Practice



Foreword

- Huawei OceanProtect Backup Storage is a robust storage system with the fastest recovery capability, which leads the dedicated backup storage industry into the flash-to-flash-to-anything (F2F2X) era. It helps you quickly and reliably back up and restore data at a low TCO, using E2E acceleration and an active-active architecture. It is also a trusted choice for demanding scenarios in government, finance, carrier, healthcare, and manufacturing sectors.
- This chapter describes the hardware and software architecture, functions and features, and the best practice of OceanProtect backup storage .

Objectives

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

- Understand the hardware and software architecture
- Understand the technical features of data reduction
- Understand the technical features of high performance and reliability
- List the centralized backup solution best practice

Contents

- 1. Product Overview and Hardware & Software Architecture**
2. High Data Reduction Ratio
3. High Performance and Reliability
4. Centralized Backup Solution Best Practice
 - Centralized Backup System Architecture
 - OceanProtect Backup Storage Solution Best Practice with Commvault
 - OceanProtect Backup Storage Solution Best Practice with Veeam

Overview and Objectives

- This section describes the overall introduction and hardware & software architecture of Huawei OceanProtect backup storage.
- On completion of this section, you will be able to:
 - List the specific models and specifications
 - Understand the hardware structure and components, software architecture

Product Overview

Mid-range OceanProtect X6000



High-end OceanProtect X8000



High-end OceanProtect X9000



	OceanProtect X6000 (all-flash mode)	OceanProtect X6000 (HDD mode)	OceanProtect X8000 (all-flash mode)	OceanProtect X8000 (HDD mode)	OceanProtect X9000 (all-flash mode)	OceanProtect X9000 (HDD mode)
Single-node specifications	2 U, 2 controllers	2 U, 2 controllers	2 U, 2 controllers	2 U, 2 controllers	4 U, 4 controllers	4 U, 4 controllers
Maximum number of nodes*	1	1	2	2	2	2
Data disk type	SAS SSD	NL-SAS HDD	SAS SSD	NL-SAS HDD	SAS SSD	NL-SAS HDD
Capacity per data disk**	3.84 TB/7.68 TB	4 TB	7.68 TB	4 TB/8 TB	7.68 TB	8 TB
System usable capacity	16 TB to 300 TB		150 TB to 2.0 PB		480 TB to 3.6 PB	
System backup bandwidth	Up to 19 TB/hour	Up to 19 TB/hour	Up to 55 TB/hour	Up to 55 TB/hour	Up to 155 TB/hour	Up to 155 TB/hour
System restore bandwidth	Up to 22 TB/hour	Up to 8 TB/hour	Up to 57 TB/hour	Up to 24 TB/hour	Up to 172 TB/hour	Up to 48 TB/hour
Front-end port type	8/16/32 Gbit/s FC, 10/25/40/100GE					
Back-end port type	SAS 3.0					

Note: *: A node corresponds to a controller enclosure.

**: If disks of other capacity specifications are required, contact Huawei sales personnel for evaluation.

Hardware Platform

OceanProtect X9000 storage controller enclosure



4 U, 28 shared interface modules

OceanProtect X6000/X8000 storage controller enclosure



2 U, 2 controllers per enclosure, 12 interface modules

SAS SSD enclosure

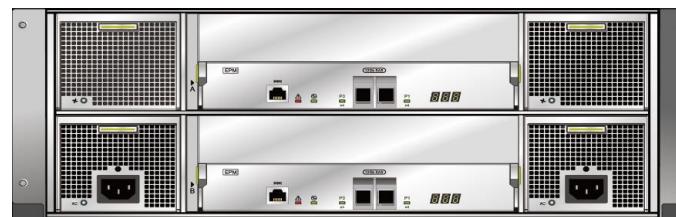


Front panel



4 U, 4 controllers per enclosure

NL-SAS disk enclosure



2 U, 25 SAS SSDs



4 U, 24 NL-SAS disks

Hardware Architecture of OceanProtect X9000 Storage Controller



A/A hardware architecture adopted for redundancy of all components, ensuring no single point of failure:

1. The front panel has four controllers, each of which has an independent BBU and fan, 192 cores, and two 2.5-inch system disks.
2. The rear panel supports up to 28 I/O cards, which are globally shared among four controllers. Four power modules form two power planes.
3. The fans, power modules, BBUs, and I/O cards can be maintained online.

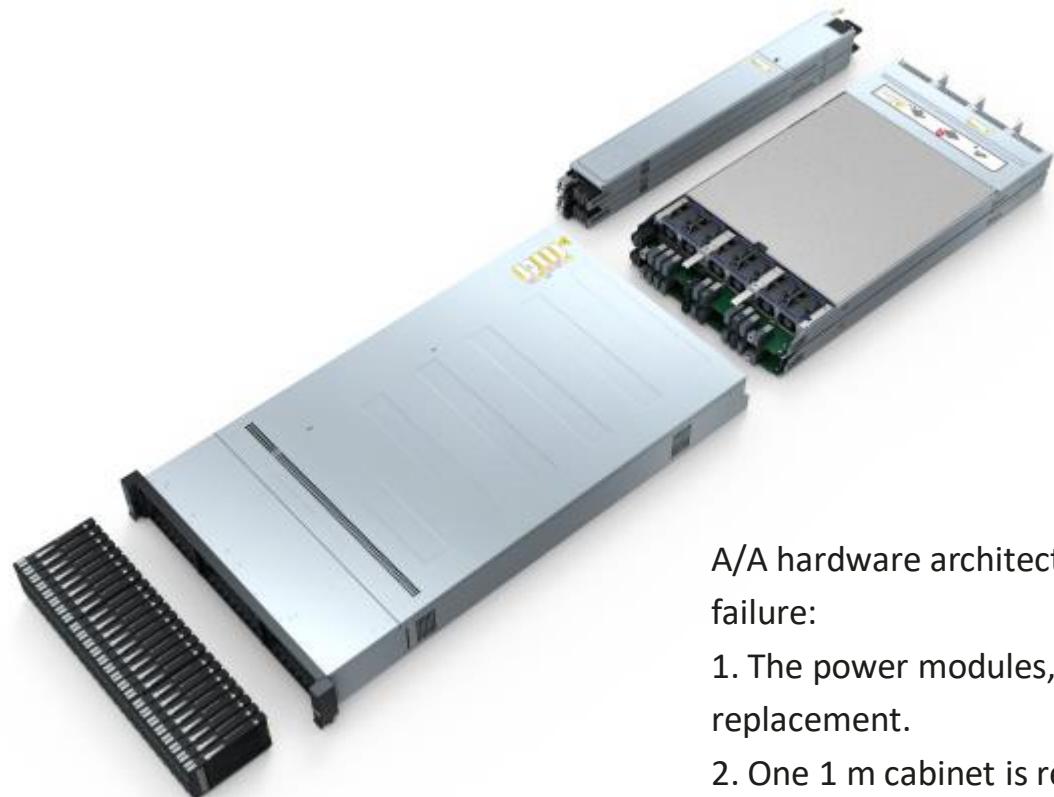
Hardware Architecture of OceanProtect X8000 Storage Controller



A/A hardware architecture adopted for redundancy of all components, ensuring no single point of failure:

1. The power modules, BBUs, and I/O cards are hot swappable. Controllers must be removed for fan replacement.
2. One 1 m cabinet is required for installing a controller enclosure with a length of 0.82 m.
3. The DIMM slots beside the IOB are empty.

Hardware Architecture of OceanProtect X6000 Storage Controller

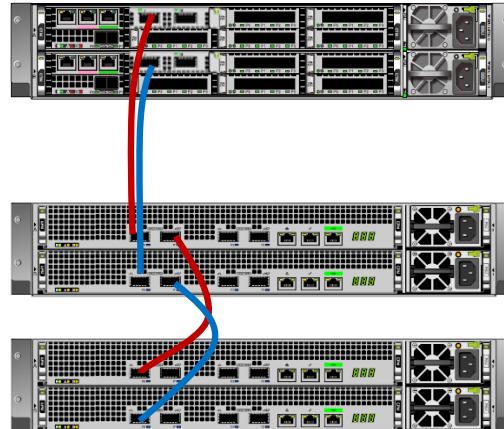


A/A hardware architecture adopted for redundancy of all components, ensuring no single point of failure:

1. The power modules, BBUs, and I/O cards are hot swappable. Controllers must be removed for fan replacement.
2. One 1 m cabinet is required for installing a controller enclosure with a length of 0.82 m.
3. The DIMM slots beside the IOB are empty.

Scale-Up + Scale-Out: On-Demand Performance and Capacity Expansion to Meet Service Growth Needs

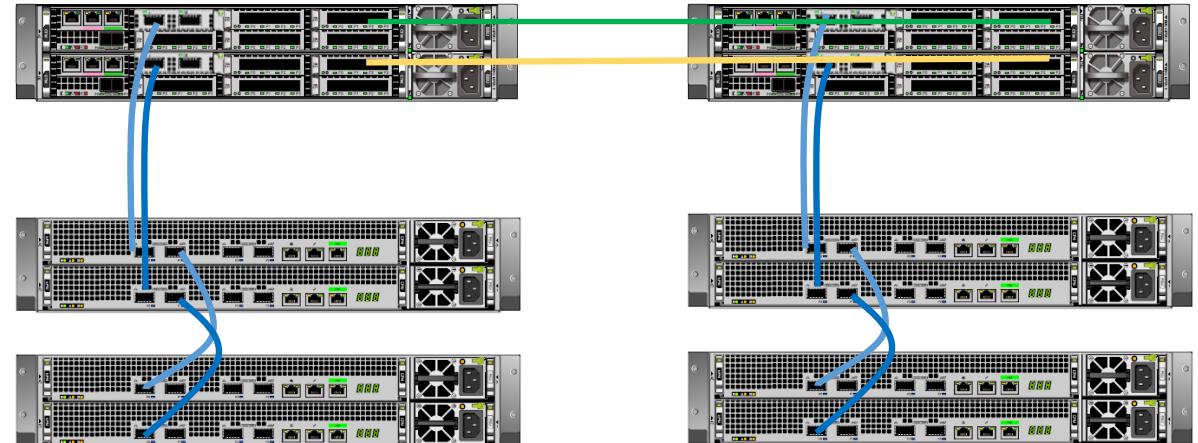
Scale-up



OceanProtect X8000 (all-flash)
(single-node)

A single-node system supports a maximum of 1.0 PB usable capacity and up to 33 TB/hour backup performance.

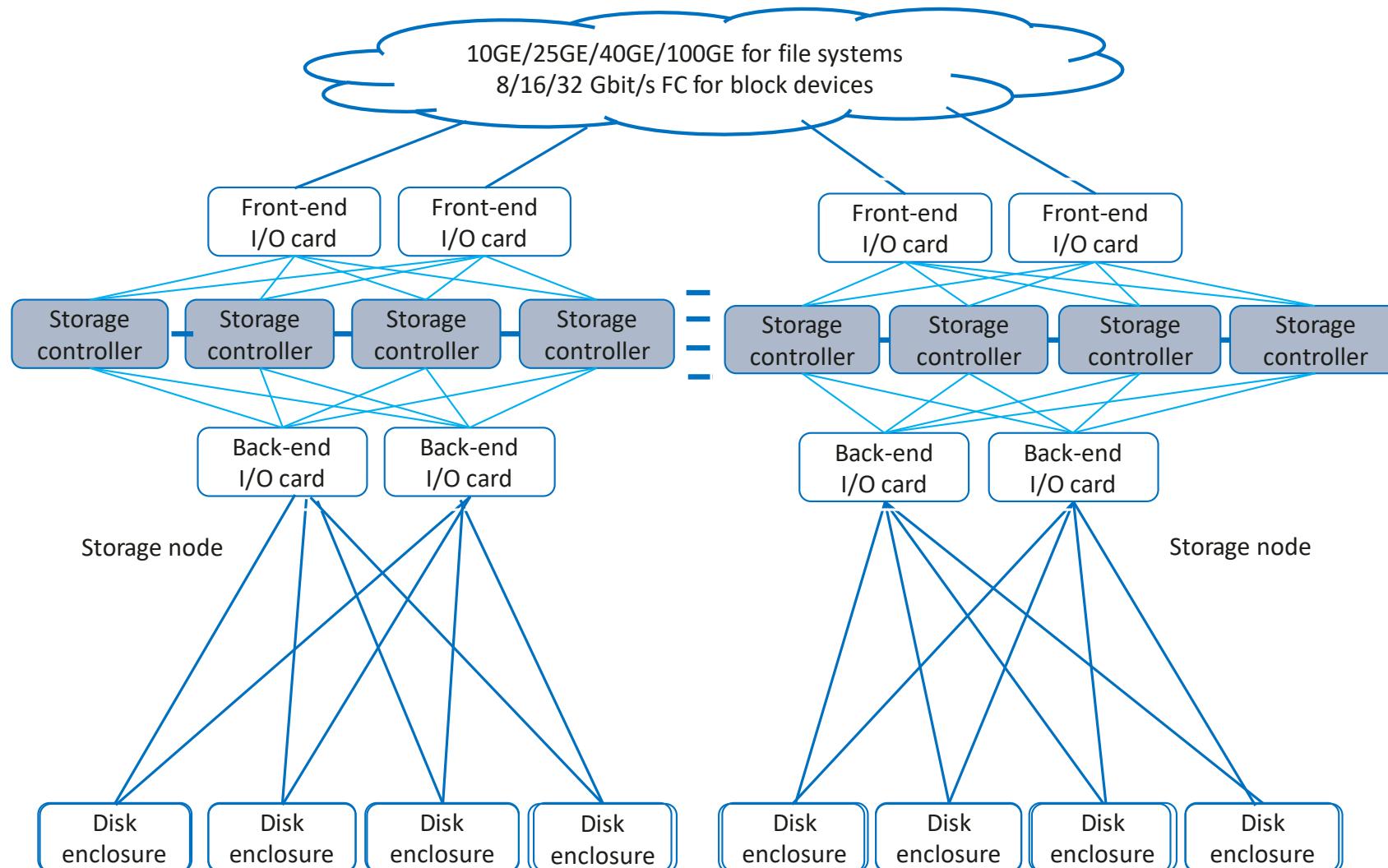
Scale-out



OceanProtect X8000 (all-flash)
(dual-node)

A dual-node system supports a maximum of 2.0 PB usable capacity and up to 55 TB/hour backup performance.

OceanProtect Backup Storage Design Principle - Distributed Architecture



Distributed architecture

- The entire series adopts the distributed system architecture. Failover is performed within seconds, so that services are not affected. Backup tasks are not redone, which ensures that each backup task is completed within the specified time window.
- Symmetric client access balances data to all controllers, providing ultimate performance.
- Controller load balancing and automatic rebalancing of scale-out, failover, and fallback services are supported, achieving on-demand expansion and simple O&M.

Global resource sharing

- Cache and pool resources are globally shared, and all resources can be automatically scheduled to maximize the system performance and eliminate performance silos.

Quiz

1. (Single-choice) Which of the following models does not support expansion of controller nodes?
 - A. X6000(all-flash mode)
 - B. X8000(all-flash mode)
 - C. X9000(all-flash mode)
 - D. X8000(HDD mode)
2. (True or False) OceanProtect X series adopts the Active-Active distributed system architecture.

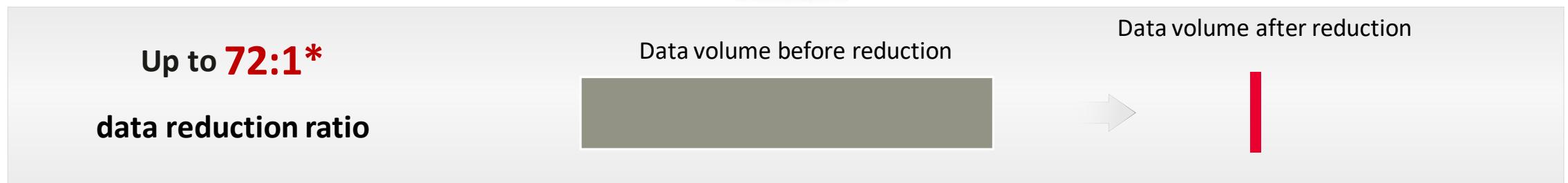
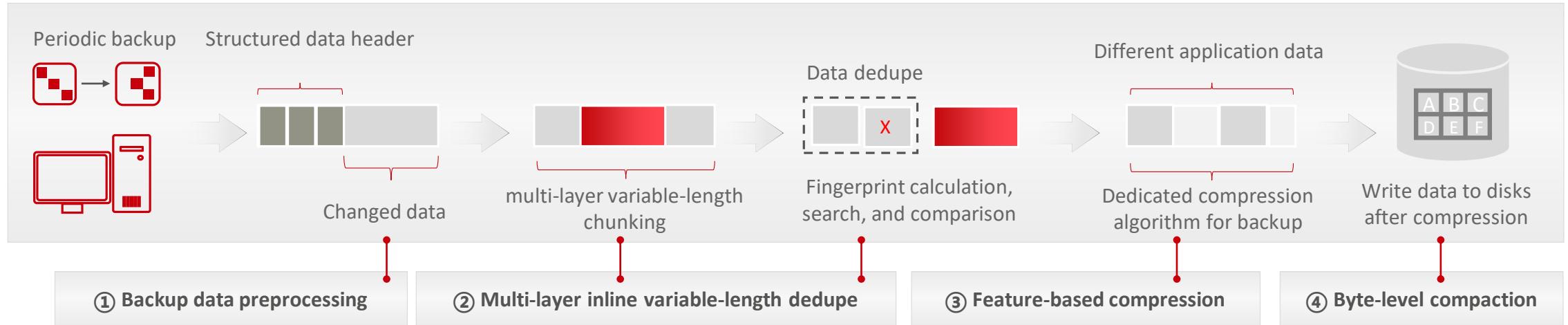
Contents

1. Product Overview and Hardware & Software Architecture
- 2. High Data Reduction Ratio**
3. High Performance and Reliability
4. Centralized Backup Solution Best Practice
 - Centralized Backup System Architecture
 - OceanProtect Backup Storage Solution Best Practice with Commvault
 - OceanProtect Backup Storage Solution Best Practice with Veeam

Overview and Objectives

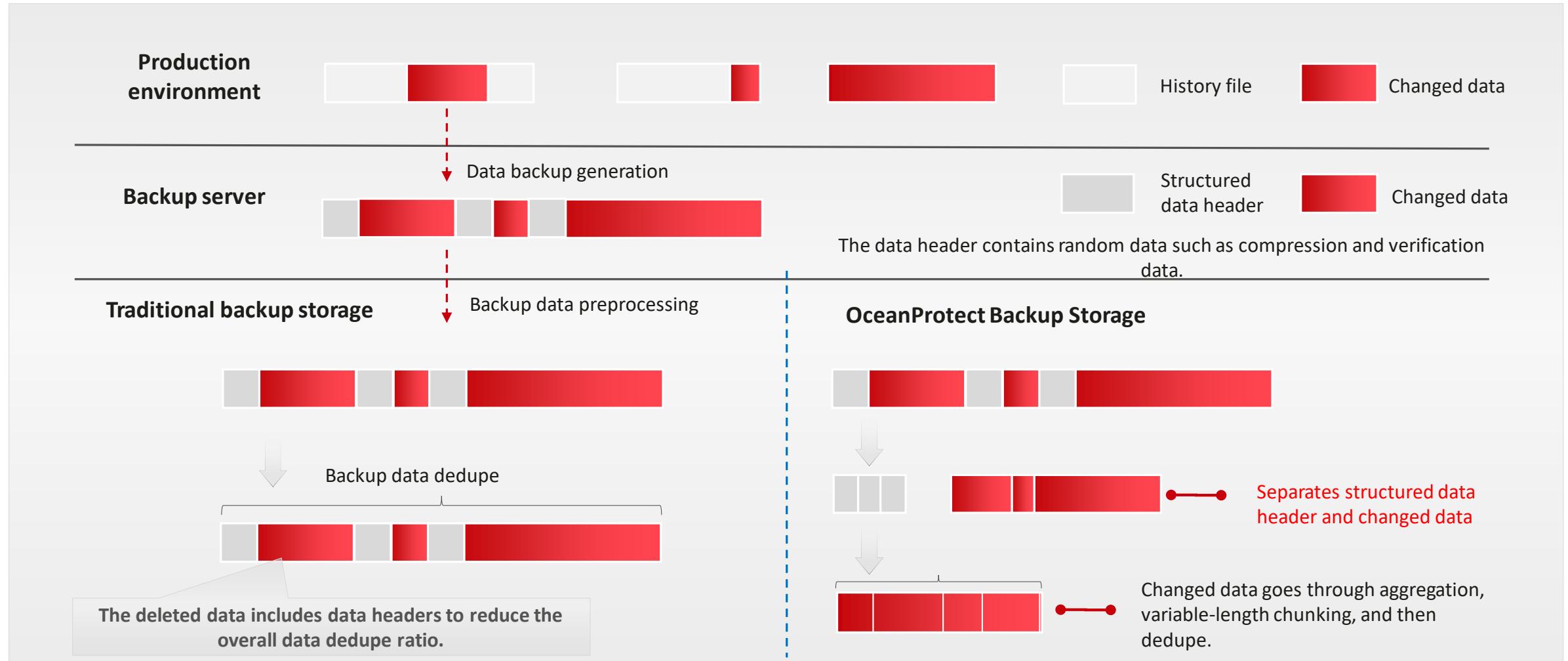
- This section describes the data reduction design of Huawei OceanProtect backup storage.
- On completion of this section, you will be able to:
 - Understand the technical principles and features of deduplication and compression.

4-Step Advanced Dedupe and Compression for Optimal Data Reduction

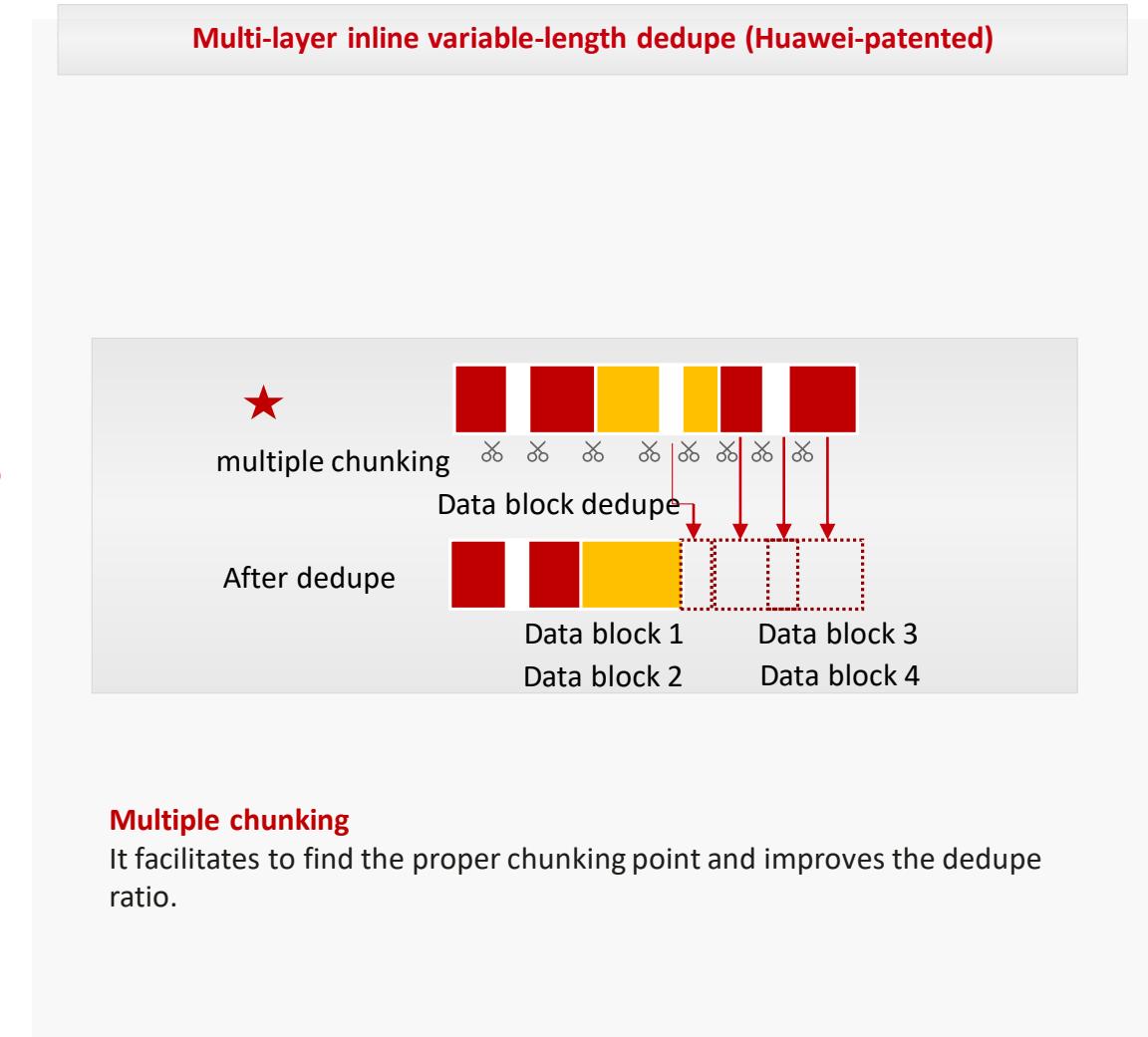
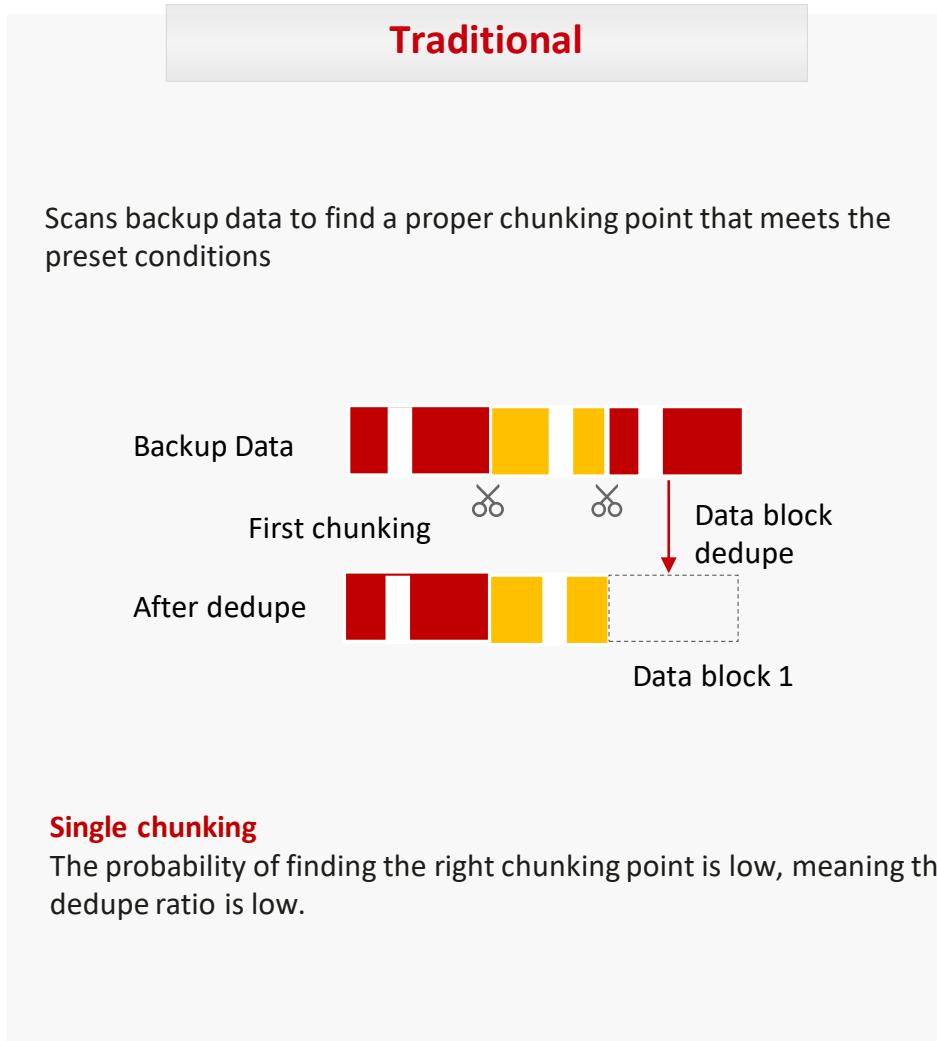


*Test case: 200 VMs in a virtualization scenario, daily full backup for 28 days, 4% modifications per day, 0.125% new data per day.

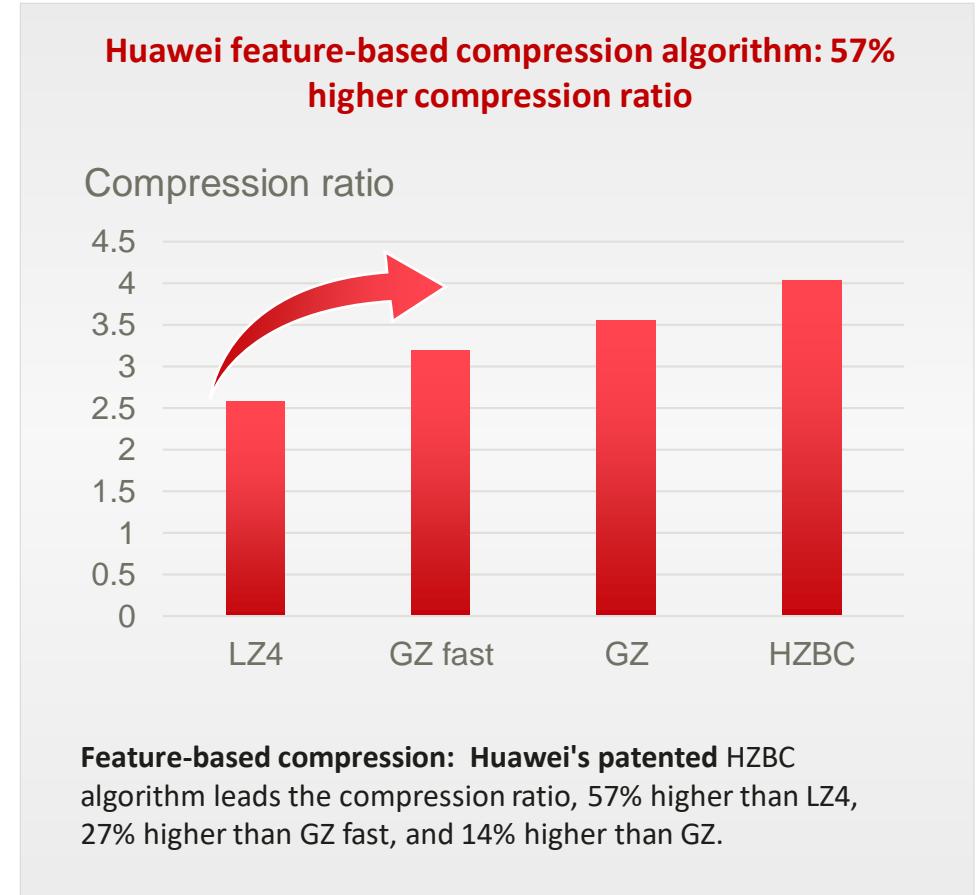
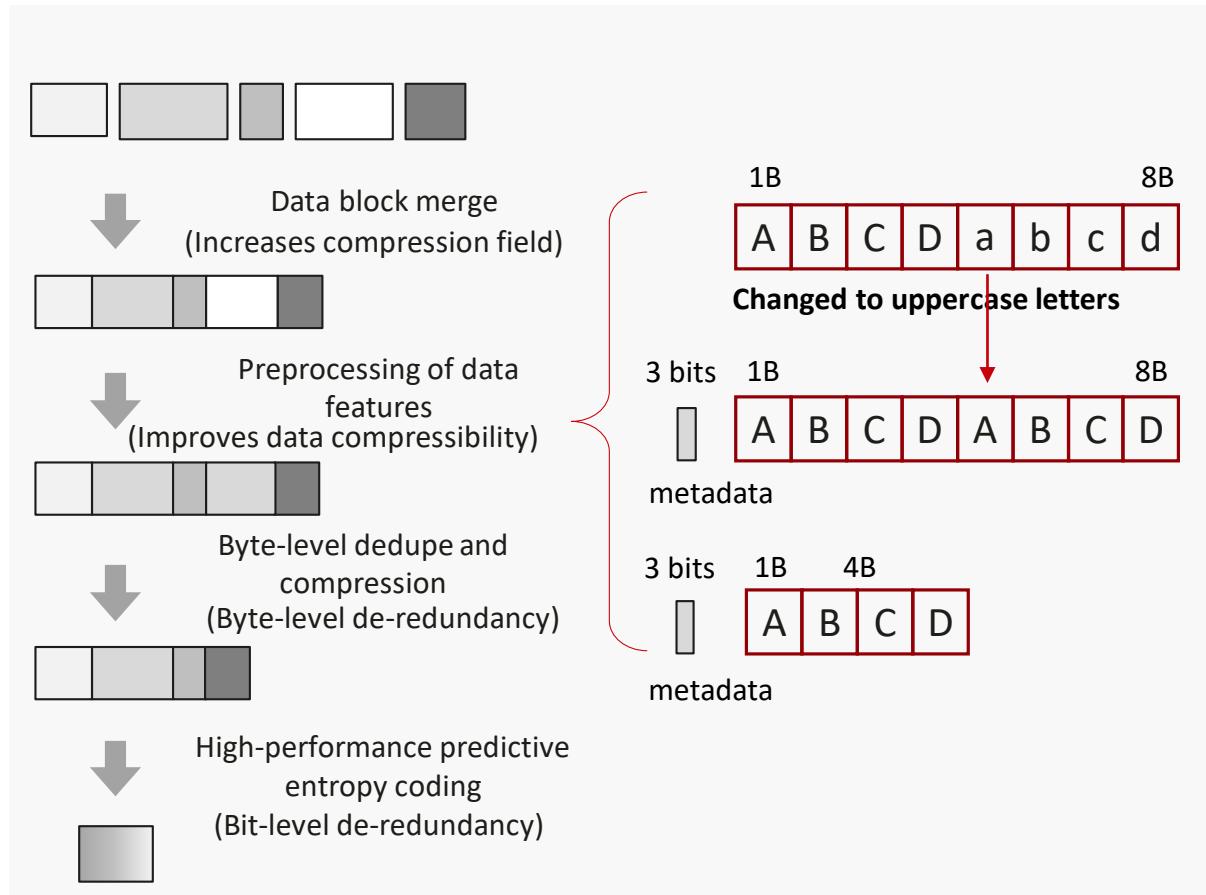
1. Backup Data Preprocessing—Separating Data Headers from Changed Data



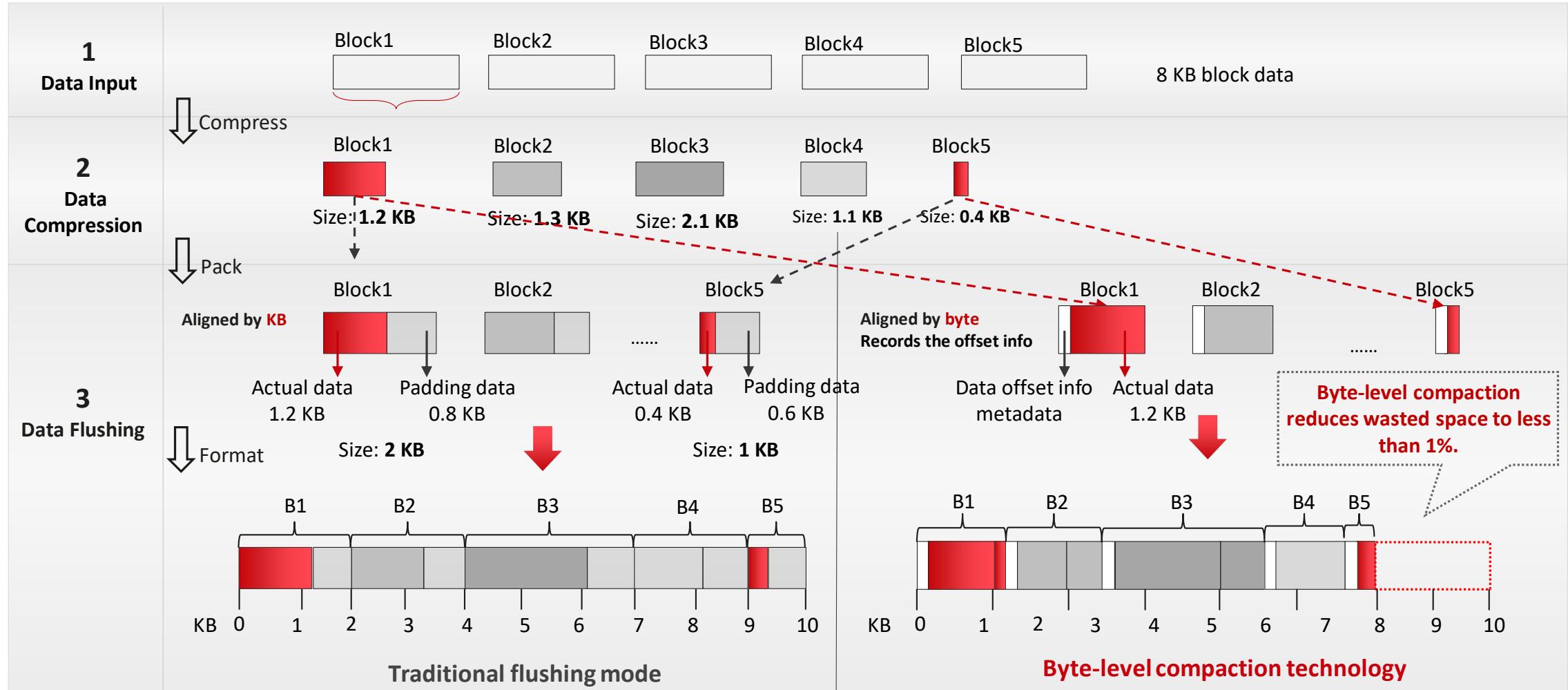
2. Multi-Layer Inline Variable-Length Dedupe of Backup Data



3. Feature-Based Compression



4. Byte-Level Compaction



Quiz

1. (Multiple-choice) What deduplication and compression technologies does OceanProtect X series support?
 - A. Backup data preprocessing—separating data headers from changed data
 - B. Multi-layer inline fixed-length dedupe
 - C. Feature-based Compression
 - D. Byte-Level Compaction
2. (Single-choice) What is the maximum reduction ratio of the OceanProtect X series?
 - A. 45:1
 - B. 65:1
 - C. 70:1
 - D. 72:1

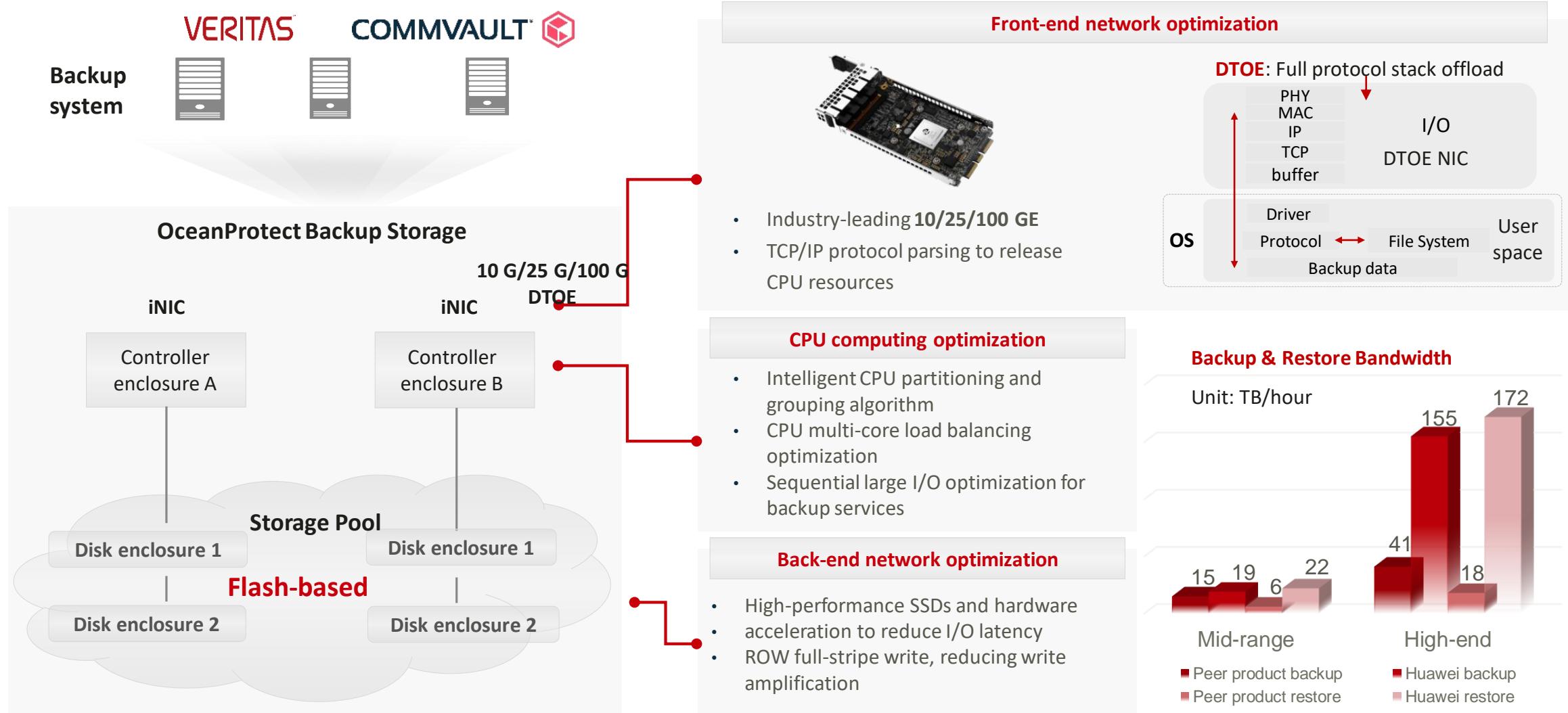
Contents

1. Product Overview and Hardware & Software Architecture
2. High Data Reduction Ratio
- 3. High Performance and Reliability**
4. Centralized Backup Solution Best Practice
 - Centralized Backup System Architecture
 - OceanProtect Backup Storage Solution Best Practice with Commvault
 - OceanProtect Backup Storage Solution Best Practice with Veeam

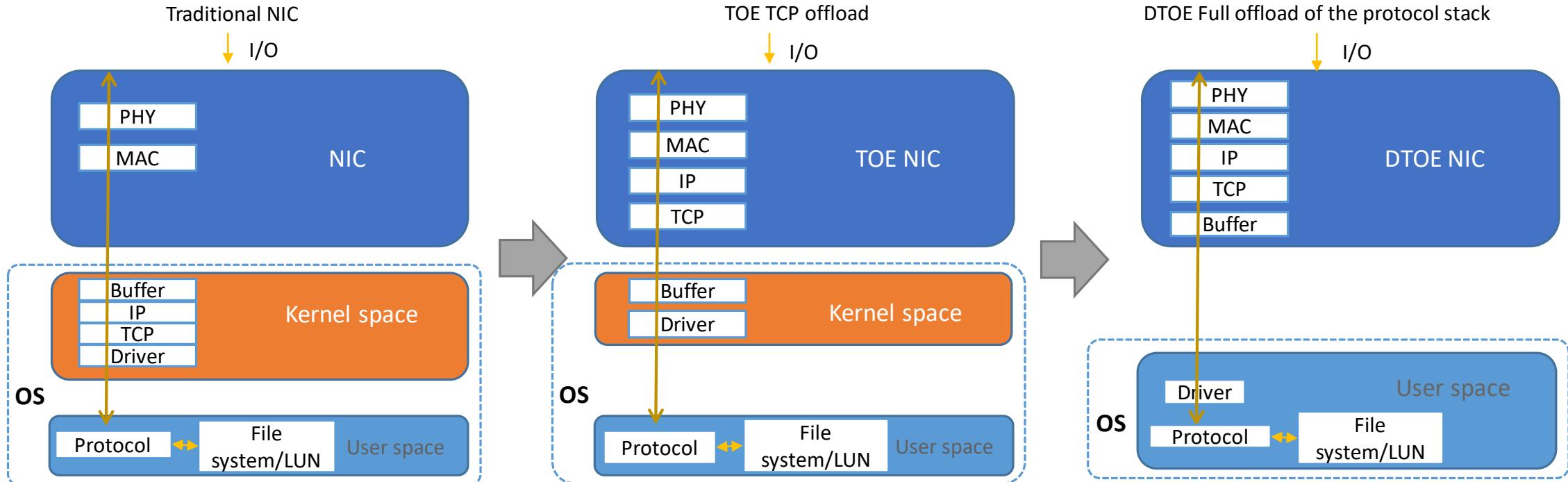
Overview and Objectives

- This section describes the high performance and reliability design of Huawei OceanProtect backup storage.
- On completion of this section, you will be able to:
 - Understand the technical principles and features of performance
 - Understand the technical principles and features of reliability

Flash-Based Acceleration Boosts Backup & Restore Performance



iNIC Optimization: Protocol Offloading Through DTOE, Releasing CPU Resources and Providing Higher Performance



Challenges:

The CPU have to spend many resources on processing each MAC frame and the TCP/IP protocol (checksum and congestion control).

Benefits:

The NIC offloads the TCP/IP protocol. The system only processes the actual TCP data flow.

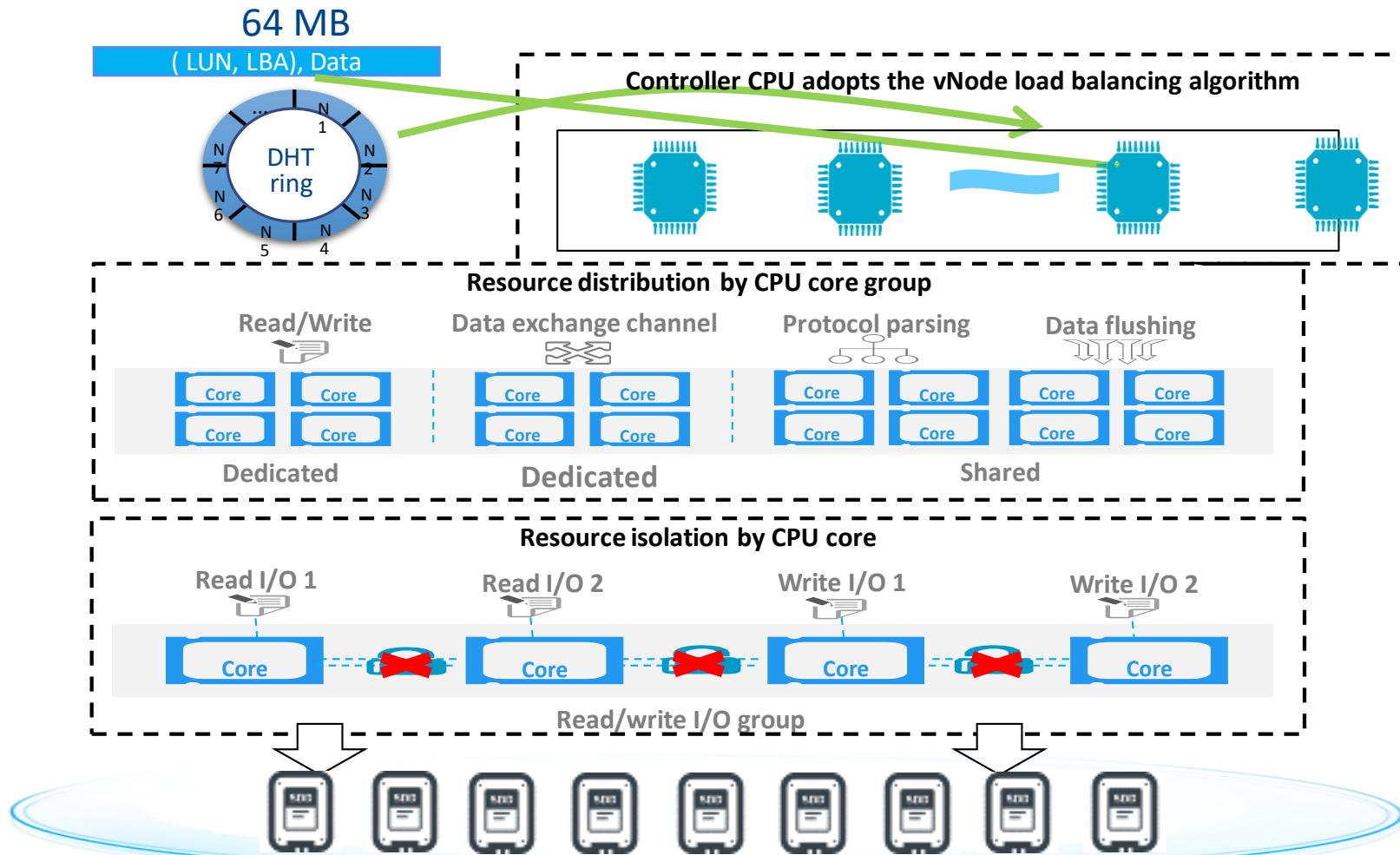
Challenges:

High latency overhead still exists in kernel mode interrupts, locks, system calls, and thread switching.

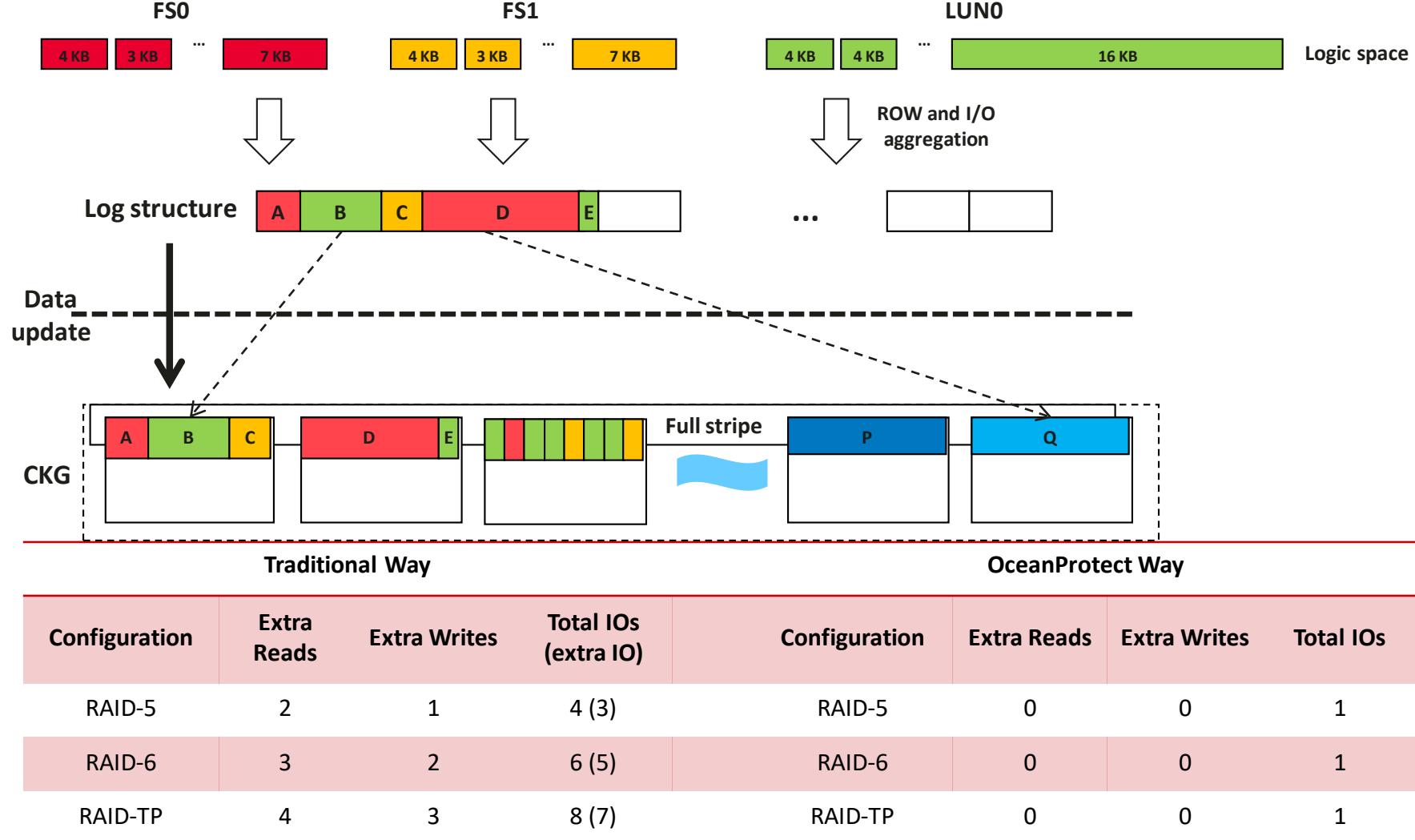
Benefits:

1. Each TCP connection has an independent hardware queue to avoid the lock overhead.
2. The hardware queue is operated in user mode to avoid the context switching overhead.
3. The polling mode reduces the latency.
4. Better performance and reliability

Service Load Balancing: Intelligent CPU Partitioning and Grouping Algorithms, Improving CPU Processing Efficiency by 30%

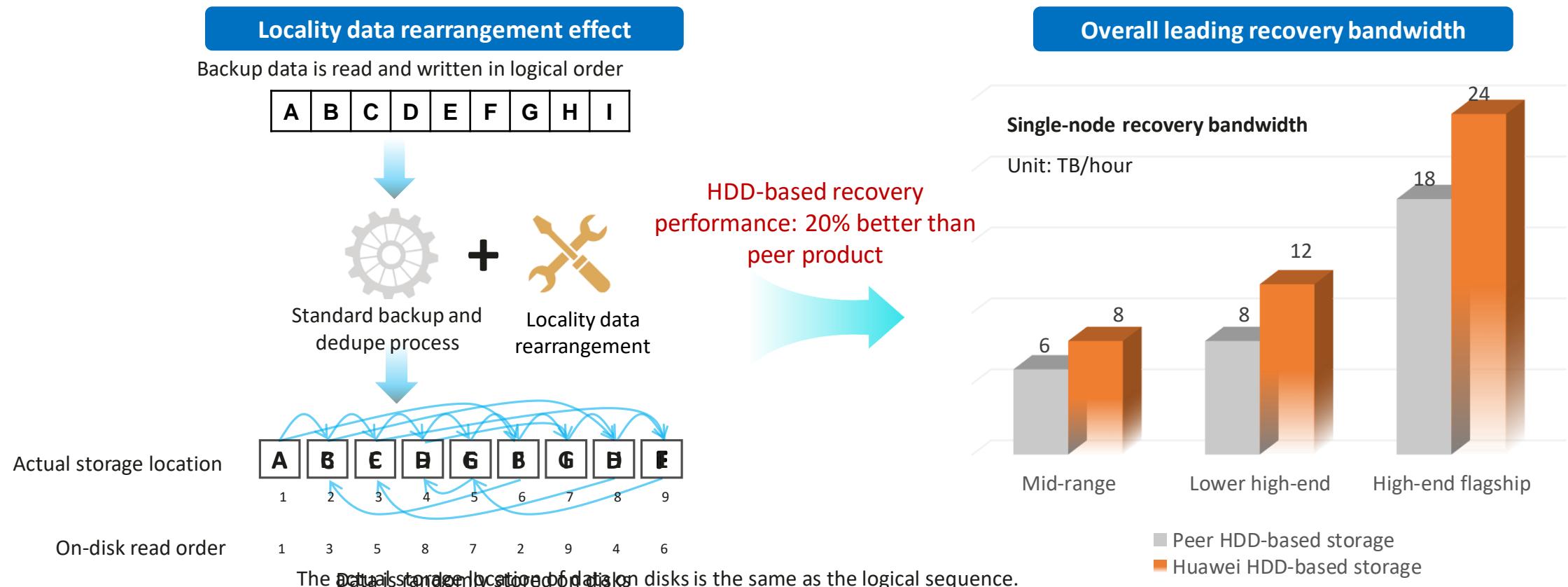


ROW-based Full-stripe Write: Reducing Write Penalty and Providing Similar Performance for Different RAID Levels

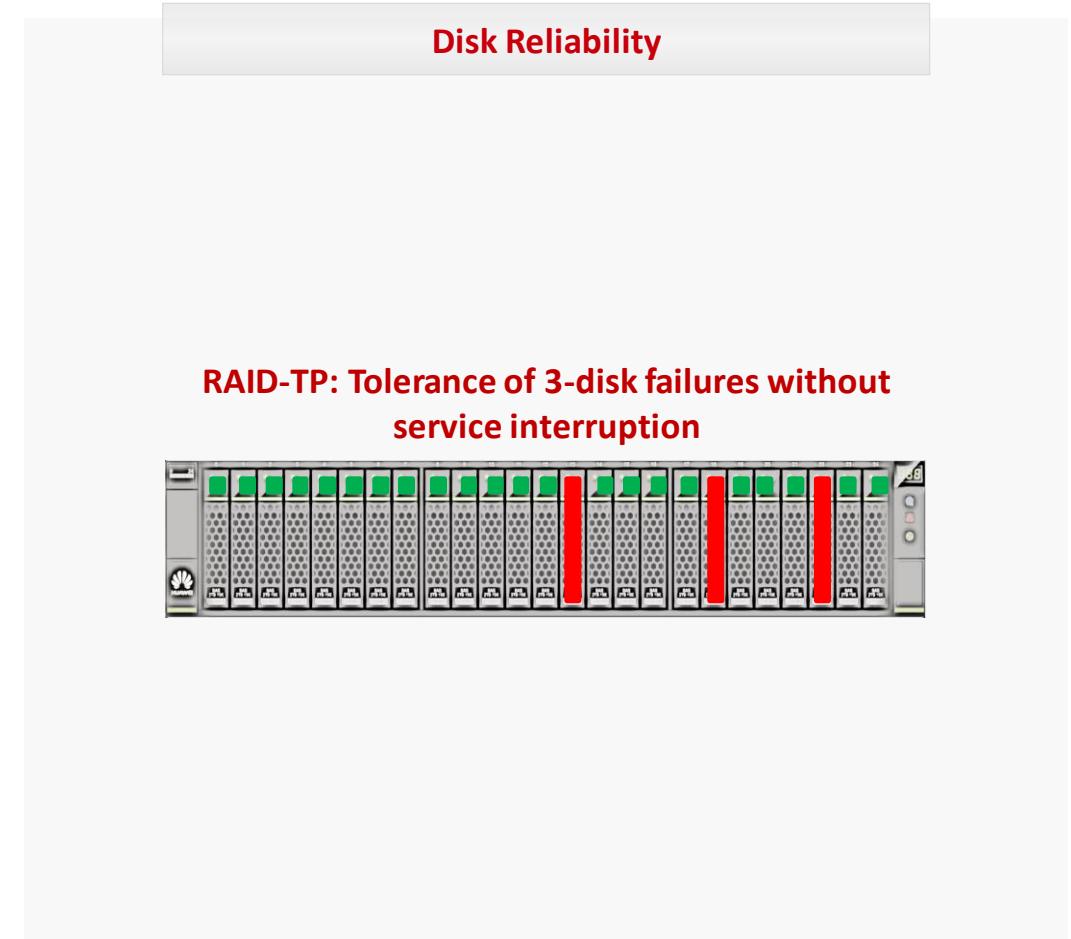
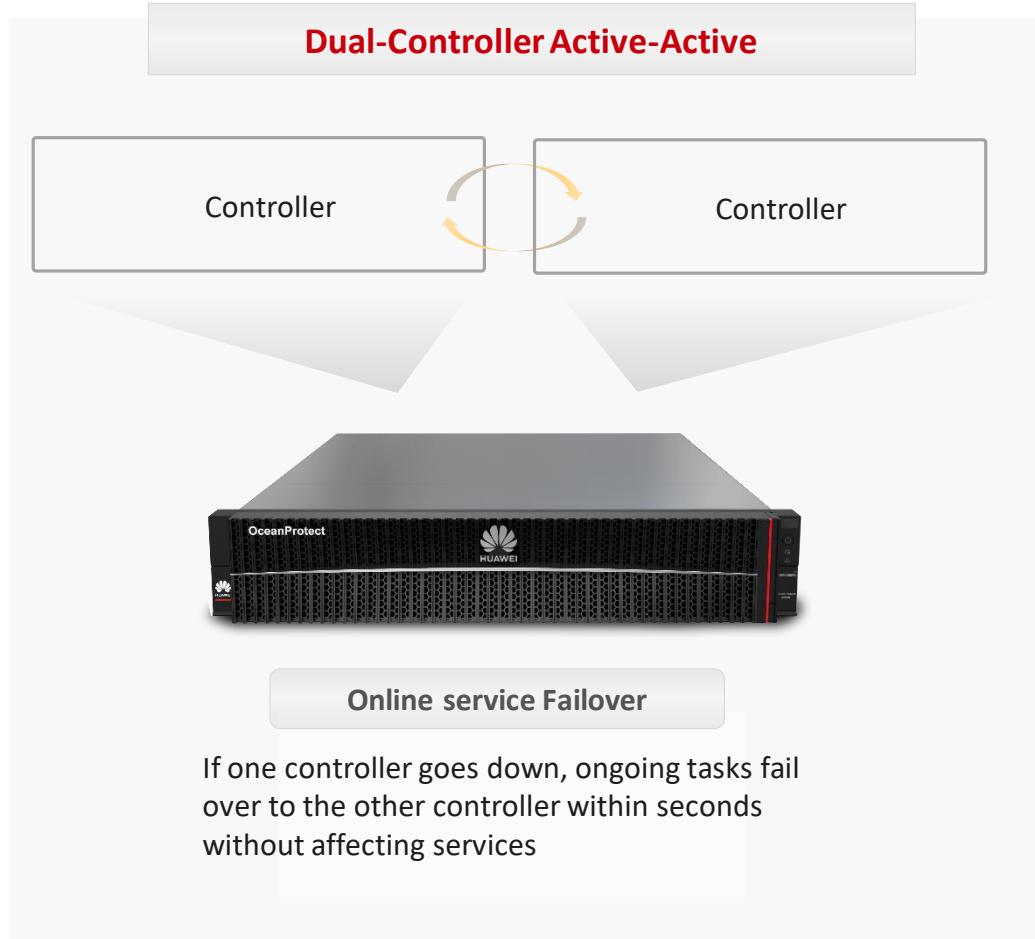


Adaptive locality data rearrangement on write optimizes HDD-based recovery performance for a leading recovery bandwidth

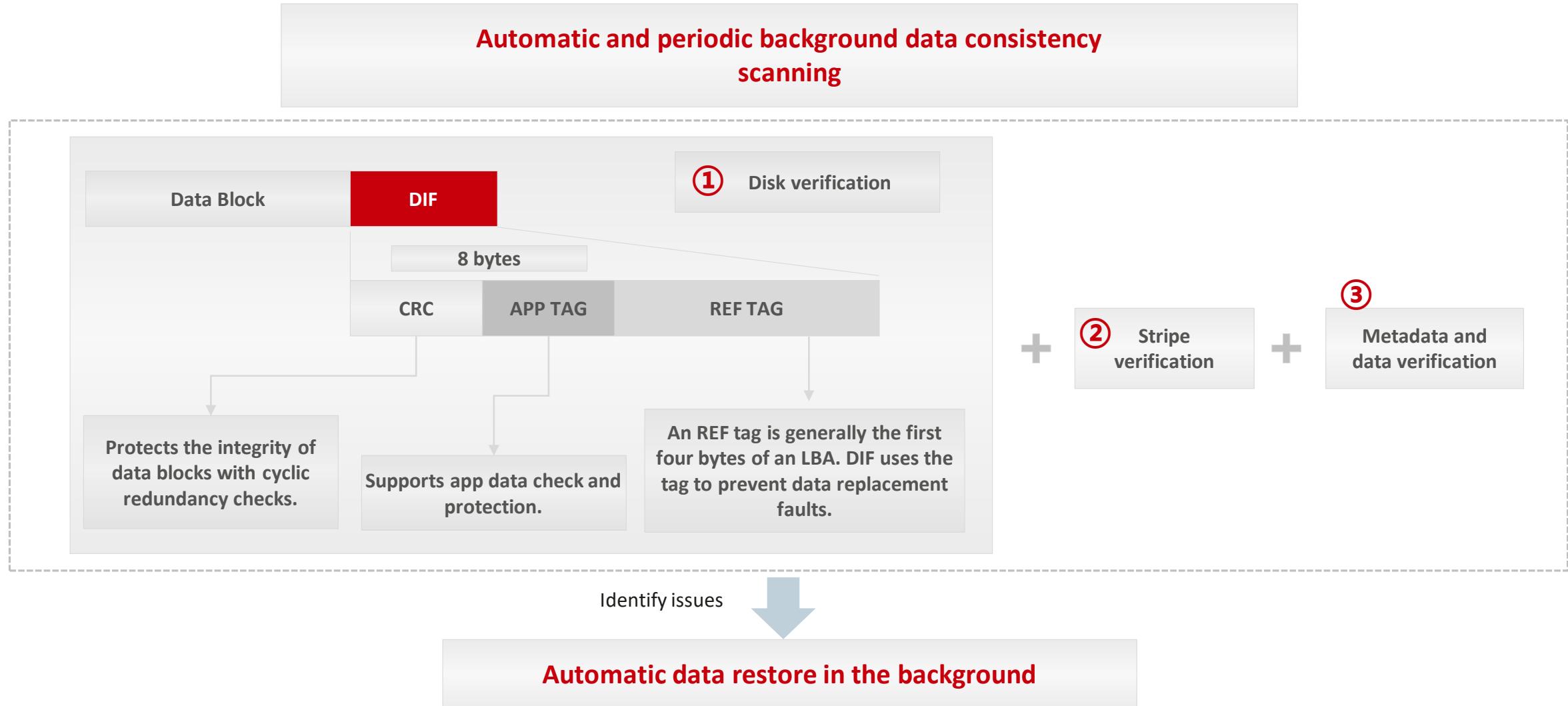
- Calculates data dispersion in real time during backup data write and adjusts data layout on disks to reduce random read operations during data restoration.
- During backup data restoration, only a few sequential reads are required to improve HDD-based recovery performance.



Dual-Controller Active-Active Architecture & RAID-TP Technology Ensures System-Level Reliability



Silent Data Consistency Check for Reliable Data Lifecycle Protection



Quiz

1. (Multiple-choice) What flash-based acceleration technologies can help to boost backup & restore performance?
 - A. Front-end network optimization
 - B. CPU computing optimization
 - C. Backup job scheduling optimization
 - D. Back-end network optimization
2. (Multiple-choice) OceanProtect system level reliability includes:
 - A. Wear leveling and anti wear leveling
 - B. Active-active controller
 - C. RAID-TP
 - D. Sync and async remote replication

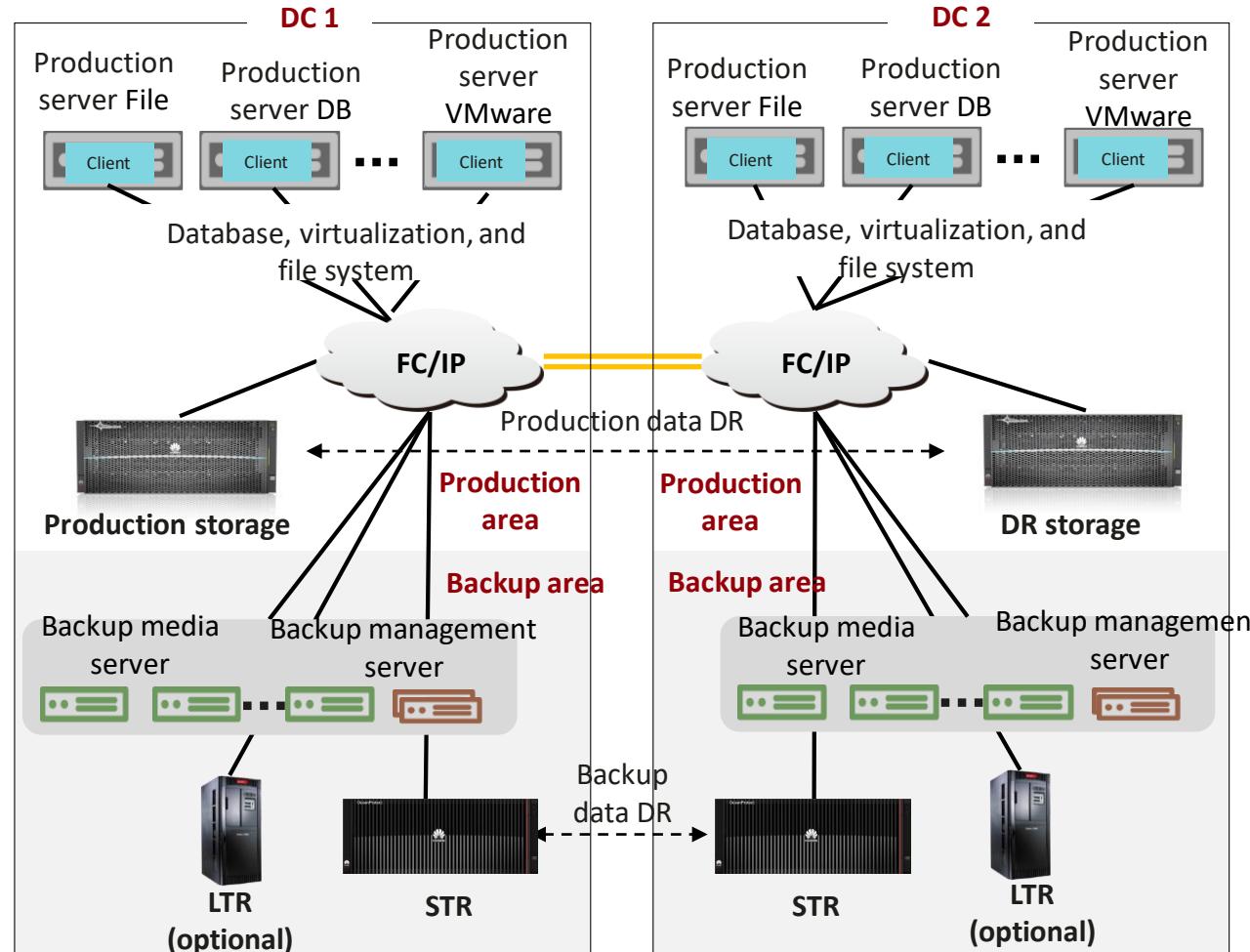
Overview and Objectives

- This section describes the centralized backup solution best practice of Huawei OceanProtect backup storage.
- On completion of this section, you will be able to:
 - Understand the centralized backup system architecture
 - Understand the best practice with Commvault
 - Understand the best practice with Veeam

Contents

1. Product Overview and Hardware & Software Architecture
2. High Data Reduction Ratio
3. High Performance and Reliability
4. **Centralized Backup Solution Best Practice**
 - **Centralized Backup System Architecture**
 - OceanProtect Backup Storage Solution Best Practice with Commvault
 - OceanProtect Backup Storage Solution Best Practice with Veeam

Centralized Backup System Architecture



- It is recommended that the 3-2-1 Rule be used, that is, there should be 3 copies of data on 2 different backup storage media with 1 copy being off site.

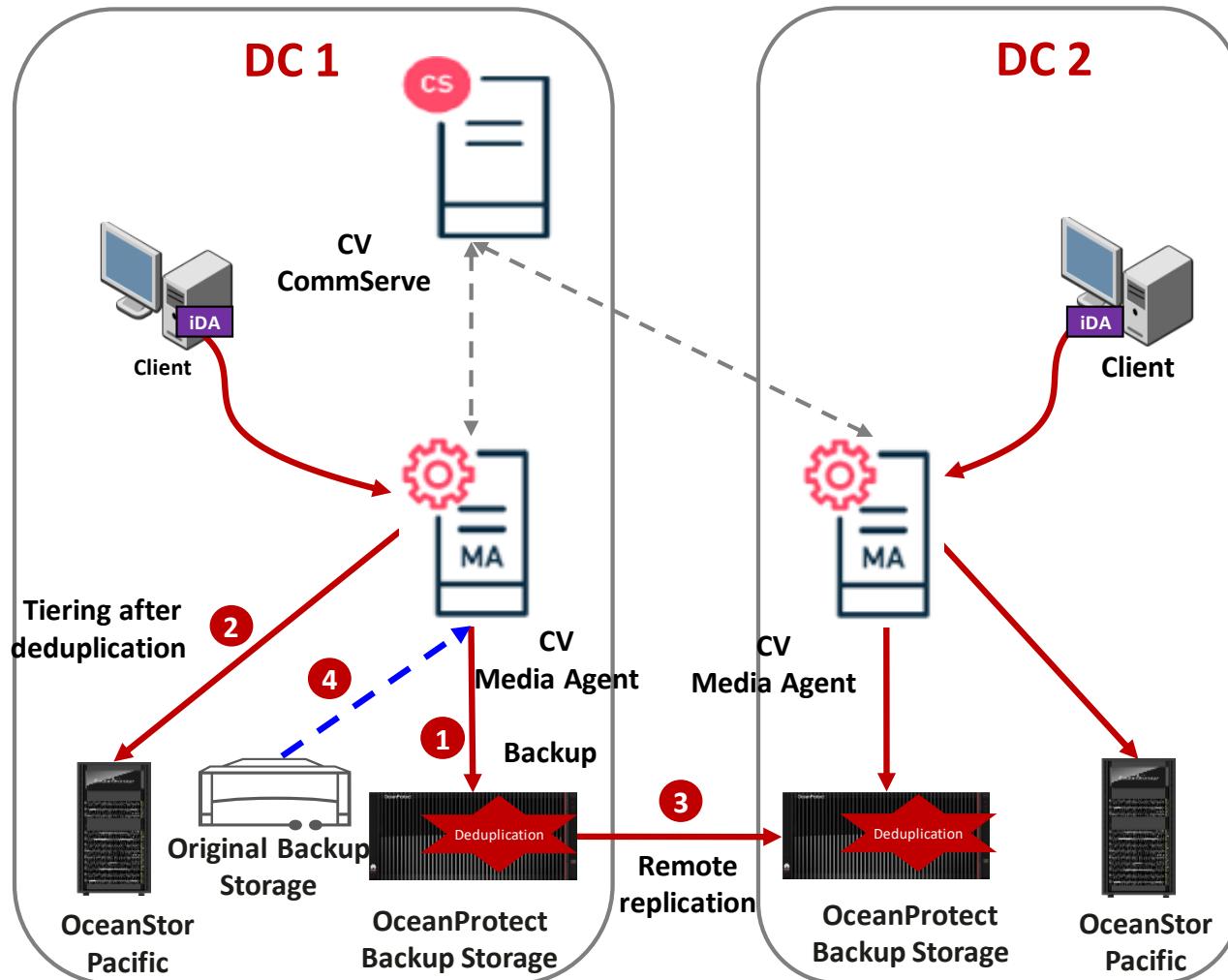
Architecture overview

- A **backup system** consists of the backup software, backup media server, backup management server and backup storage.
- Mainstream backup software:** includes NetBackup (Veritas), Veeam, Commvault, NetWorker (EMC), and TSM (IBM).
- Backup device:**
 - Short-Term Retention(STR):** can be dedicated backup storage, general storage, and VTL. It provides quick and efficient backup and recovery capabilities to obtain optimal backup and recovery performance. The cost of First level backup device is high, and data deduplication and compression are important.
 - Long-Term Retention(LTR) :** can be tape library or object storage. It stores data for a long term at a low cost. Backup copies are migrated to LTR and be stored for more than a certain period of time, achieving the optimal comprehensive retention cost.
- Backup data tiering:** migrate data that has been retained for a long time from STR to LTR based on policies. This can be implemented by backup software or the tiering feature of STR device.
- Backup data replication:** Based on compliance and data security requirements, remote disaster recovery (DR) is also required for backup data. During remote DR of backup data, data can be transmitted through the media server or backup storage.

Contents

1. Product Overview and Hardware & Software Architecture
2. High Data Reduction Ratio
3. High Performance and Reliability
4. **Centralized Backup Solution Best Practice**
 - Centralized Backup System Architecture
 - **OceanProtect Backup Storage Solution Best Practice with Commvault**
 - OceanProtect Backup Storage Solution Best Practice with Veeam

OceanProtect Backup Storage Best Practice- Connection with Commvault



Scenario and solution description

1 Backup

- OceanProtect supports NFS, CIFS, Fibre Channel, and iSCSI. **NFS and CIFS are recommended.**
- When CV uses the **Front End Terabyte (FET) capacity** license, **CV deduplication is disabled**, and backup data is deduplicated and compressed on OceanProtect after it is written by MediaAgent (MA) to OceanProtect, delivering optimal data reduction ratio.
- When CV uses the **Back End Terabyte (BET) capacity** license, CV deduplication is enabled to lower CV License costs, backup data is deduplicated and compressed on OceanProtect again after it is written by MediaAgent (MA) to OceanProtect. In this case, the gain of data reduction is not obvious. The scenario exists in only some existing Huawei re-sell sites.

2 Long-term retention (LTR) tiering

- After the MA deduplicates and tiers the data, LTR copies are stored in the object storage or public cloud storage.
Note: The copies in the object storage or public cloud storage can be **directly used for recovery**.

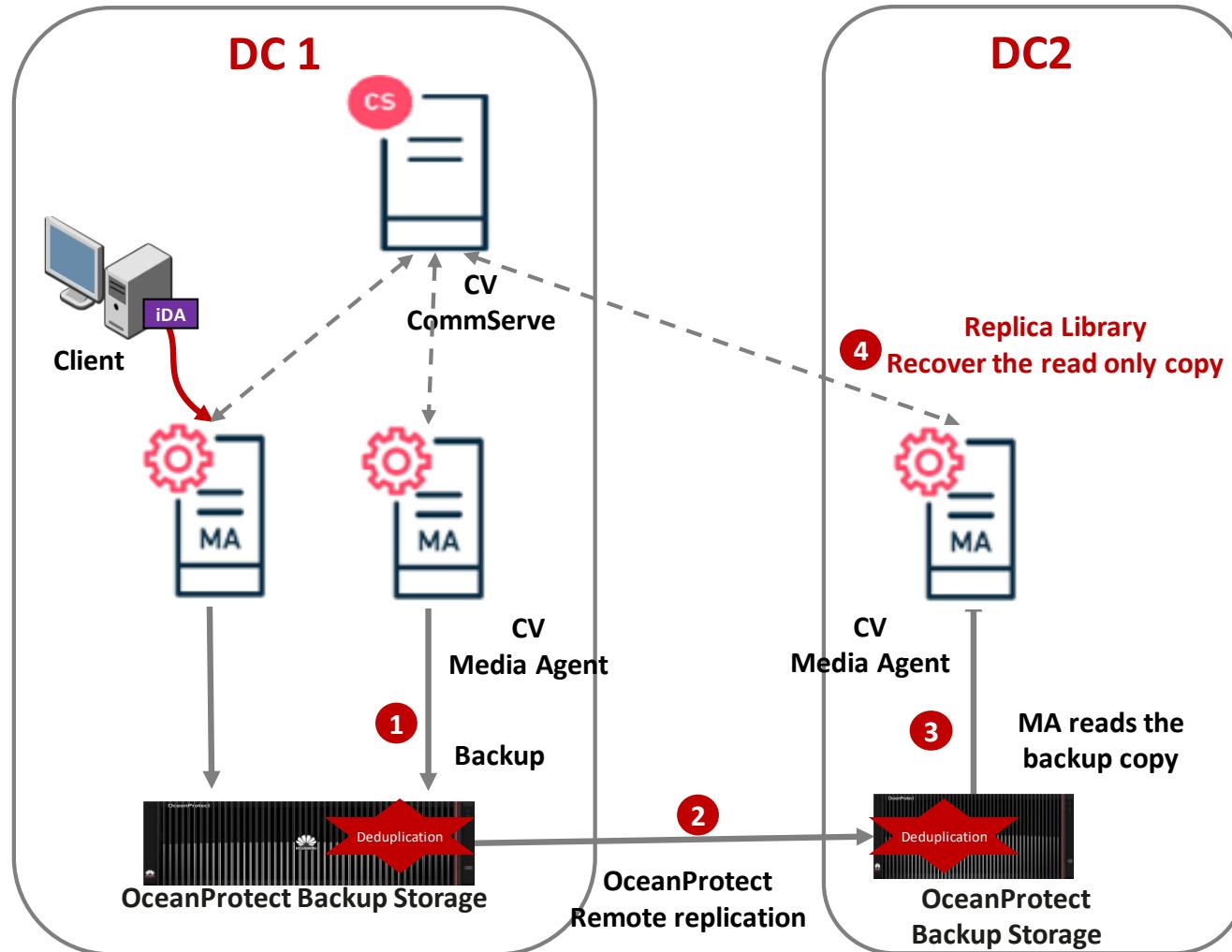
3 Remote replication

- OceanProtect is used to implement remote replication of backup data. A Replica Library created on CV can manage the backup copies.

Data dump

- Application scenario: Data on the original backup storage does not expire and needs to be dumped or migrated to a new backup storage.
- Recommended solution: Use the Auxiliary Copy function of CV to dump data from the original backup storage to OceanProtect.

Replication Solution 1 (Based on OceanProtect, Recommended)



Scenario and solution description

OceanProtect for remote replication

Scenario: In a CV backup domain, if backup copies in one MA need to be replicated to another MA, the backup copies in the primary DC are remotely replicated to the DR DC. The solution uses the HyperReplication function of the OceanProtect backup storage to implement efficient remote replication.

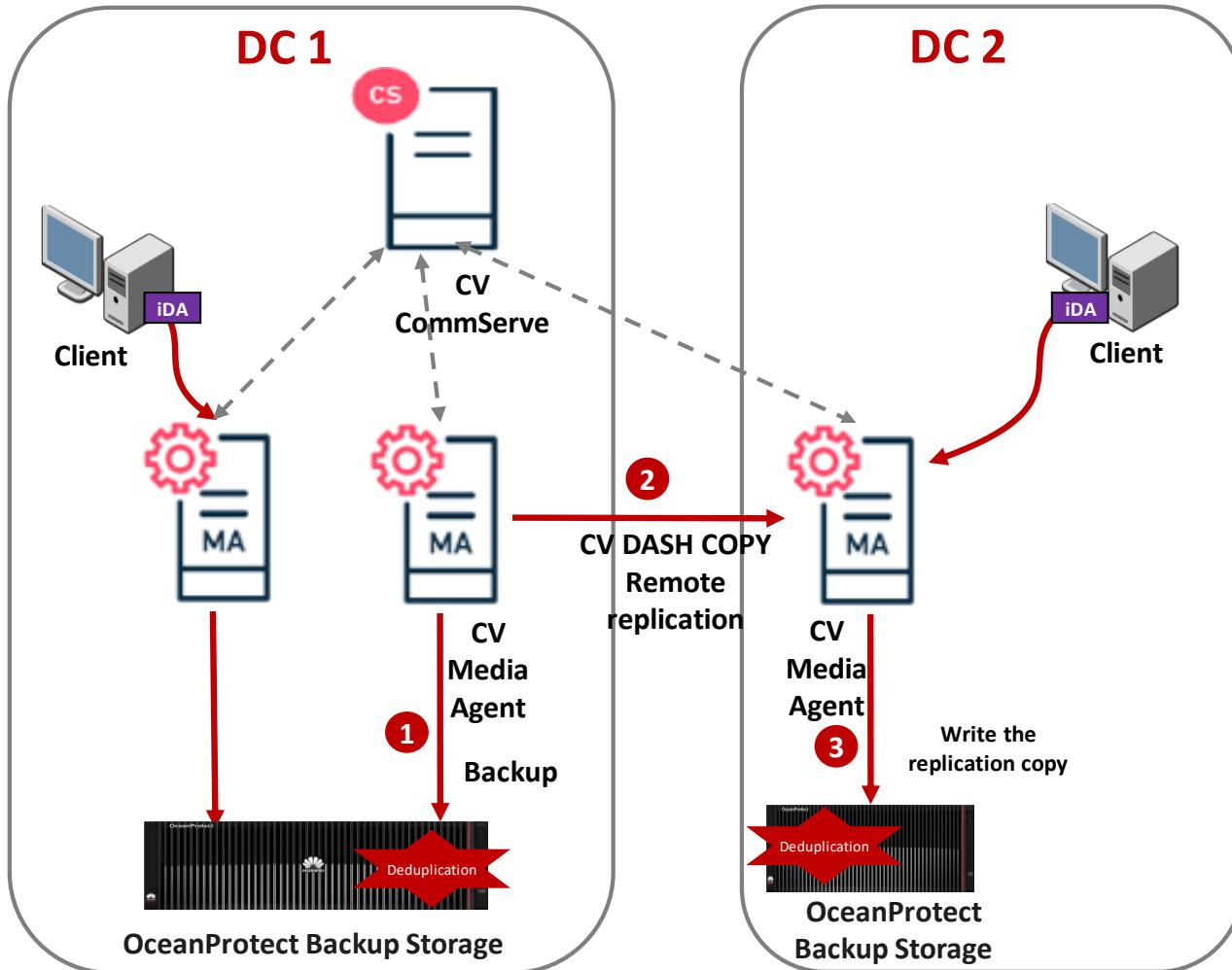
Replication model:

- one-to-one and one-to-many

Replication description:

1. OceanProtect implements remote replication for the backup file system using HyperReplication.
2. The file system replicated by the target-end OceanProtect is mounted to the target-end MA, and used to create a Replica Library in read-only mode.
3. During recovery, after selecting the target-end MA on the CS, the MA automatically finds and recovers the required data in the replicated file system.

Replication Solution 2 (Based on Commvault)



Scenario and solution description

Backup software for remote replication (Auxiliary Copy)

Scenario: In a CV backup domain, if backup copies in one MA need to be replicated to another MA, the backup copies in the primary DC are remotely replicated to the DR DC. The solution uses the DASH Copy function of CV to implement remote replication for backup data.

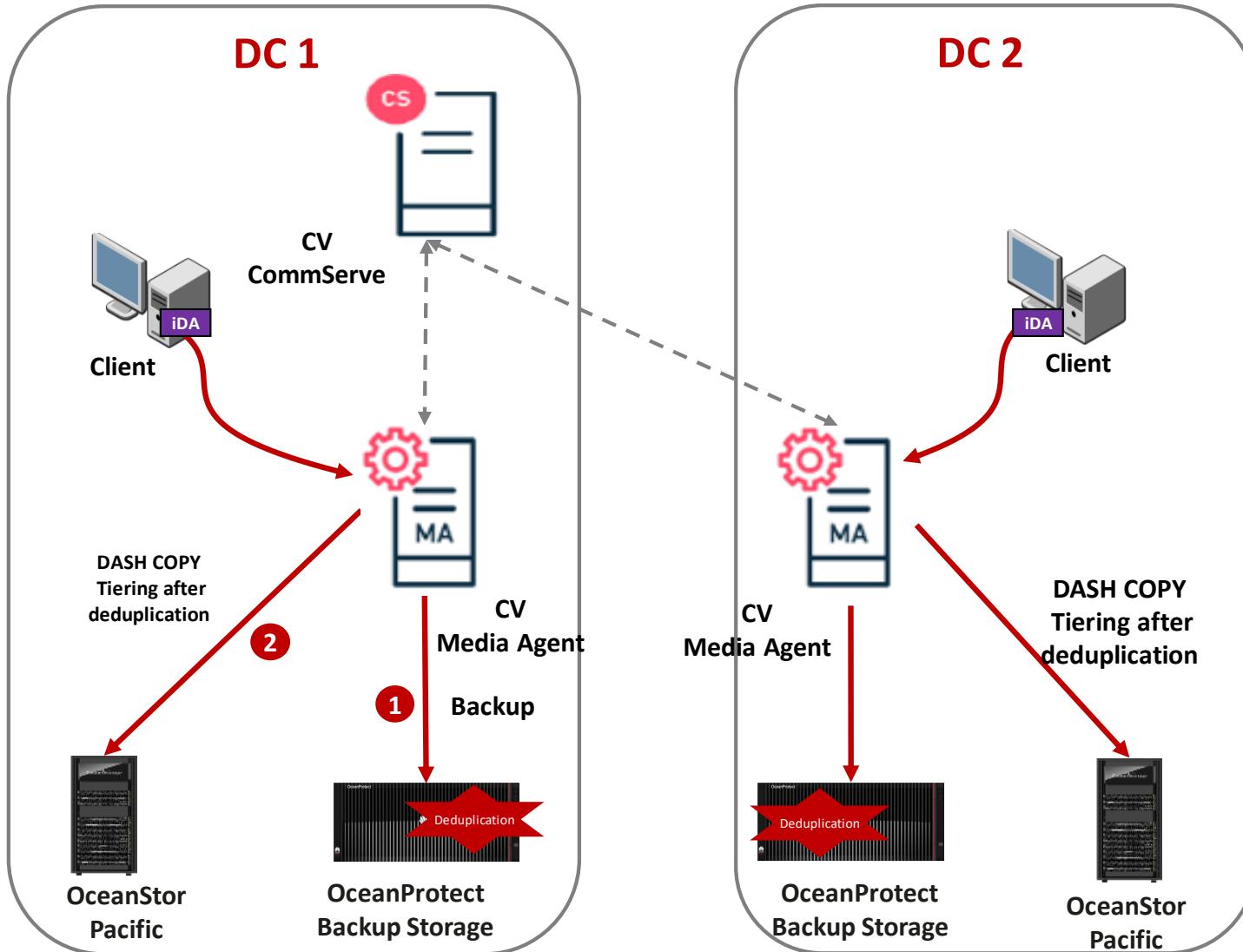
Replication model:

- One-to-one unidirectional replication
Data in a single production data center can be backed up to a DR site.
- One-to-many
Data in a single production data center can be backed up to multiple DR sites.
- Many-to-one
Remote office data in multiple domains can be backed up to a storage device in a single domain.
- Many-to-many
Remote data centers in multiple domains can back up multiple DR sites.

Replication description:

1. The source-end MA reads backup data from OceanProtect and replicates the data using DASH Copy.
2. With deduplication enabled on the target-end MA, the source-end MA sends data fingerprint to the deduplication database (DDB) on the target-end MA over the network for deduplication. Only deduplicated data is transferred to the storage.
3. The target-end MA writes the deduplicated data to OceanProtect.

Tiered Archiving Solution (Based on the Commvault LTR)

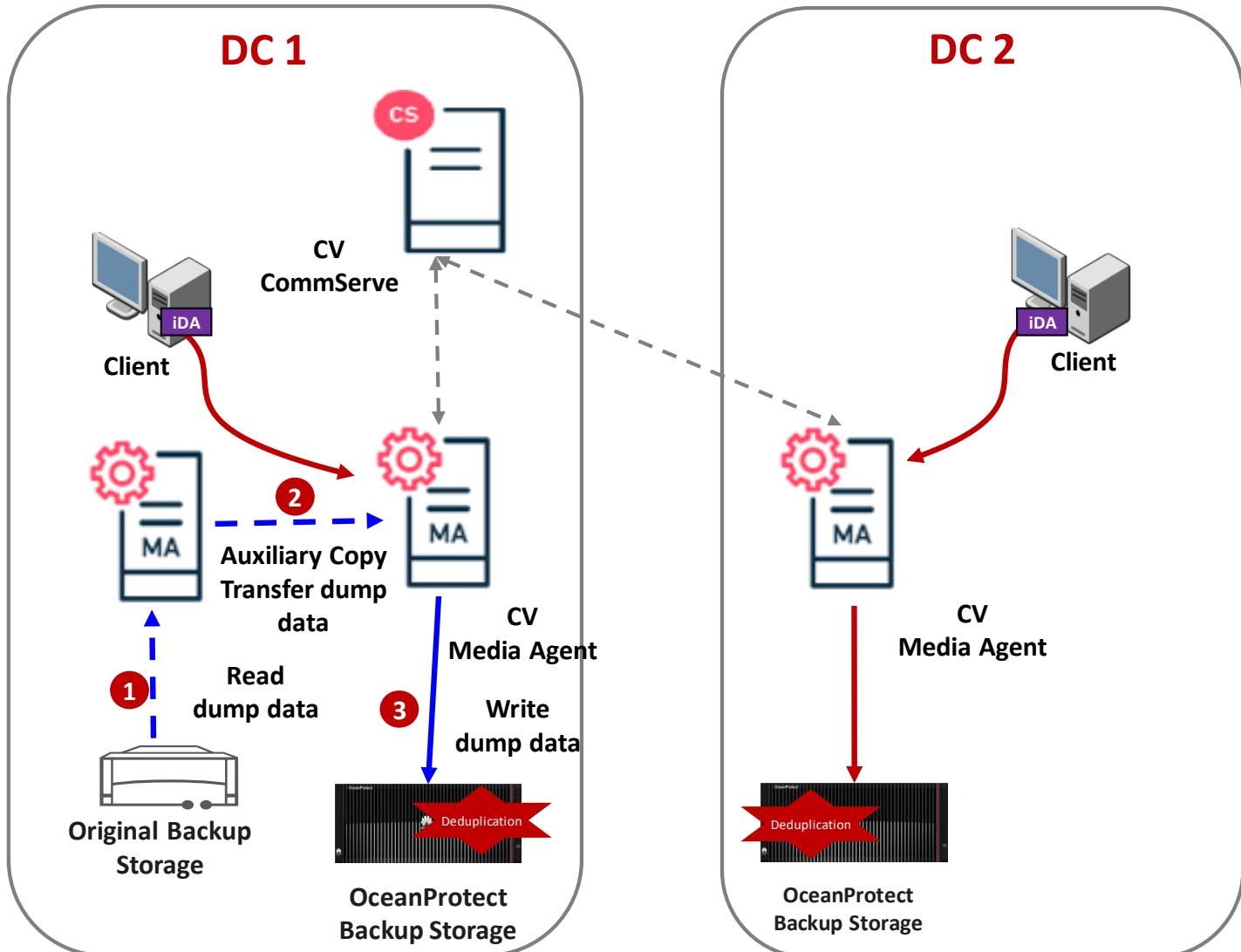


Scenario and solution description

Tiering description:

- Data in the OceanProtect is tiered to the object storage or public cloud storage using CV. Data is replicated to the S3 object storage using the DASH COPY function of CV.
- 1. The client data is backed up to the OceanProtect backup storage for short-term retention.
- 2. DASH COPY is configured to replicate data to the S3 object storage. CV can deduplicate and then archive the data to the S3 object storage for long-term retention.
- 3. The solution allows the retention policy configuration of tiered copies on CV, or the data recovery through the S3 object storage.
- 4. CV supports mainstream S3 object storage and public cloud storage.

Data dump Solution

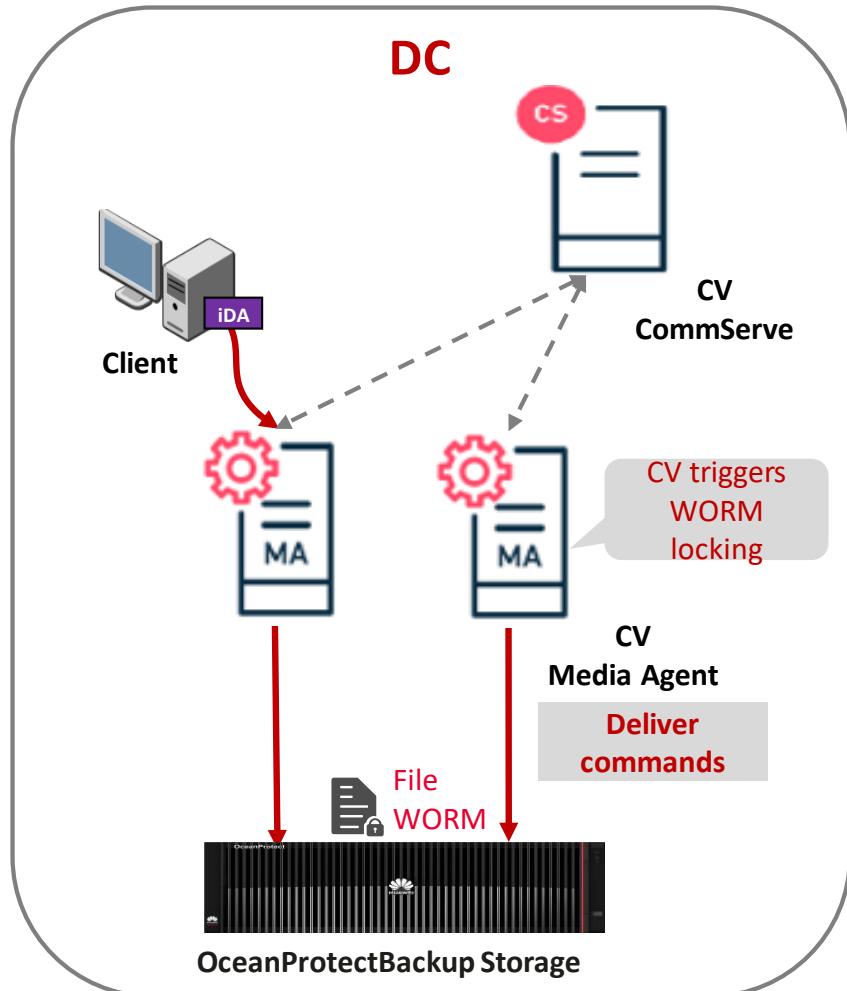


Scenario and solution description

Data dump description:

- The Auxiliary Copy function of CV is used to migrate data from the original backup storage to OceanProtect.
 1. The CV MA reads data to be dumped from the original backup storage.
 2. CV replicates the data to the MA server connected to the OceanProtect using Auxiliary Copy.
 3. The MA writes the data to OceanProtect. The OceanProtect then deduplicates and stores the data.
- Using this solution, the dumped data can be successfully restored.

Commvault + OceanProtect Ransomware Protection Solution – WORM File System integrate with CV



Solution description

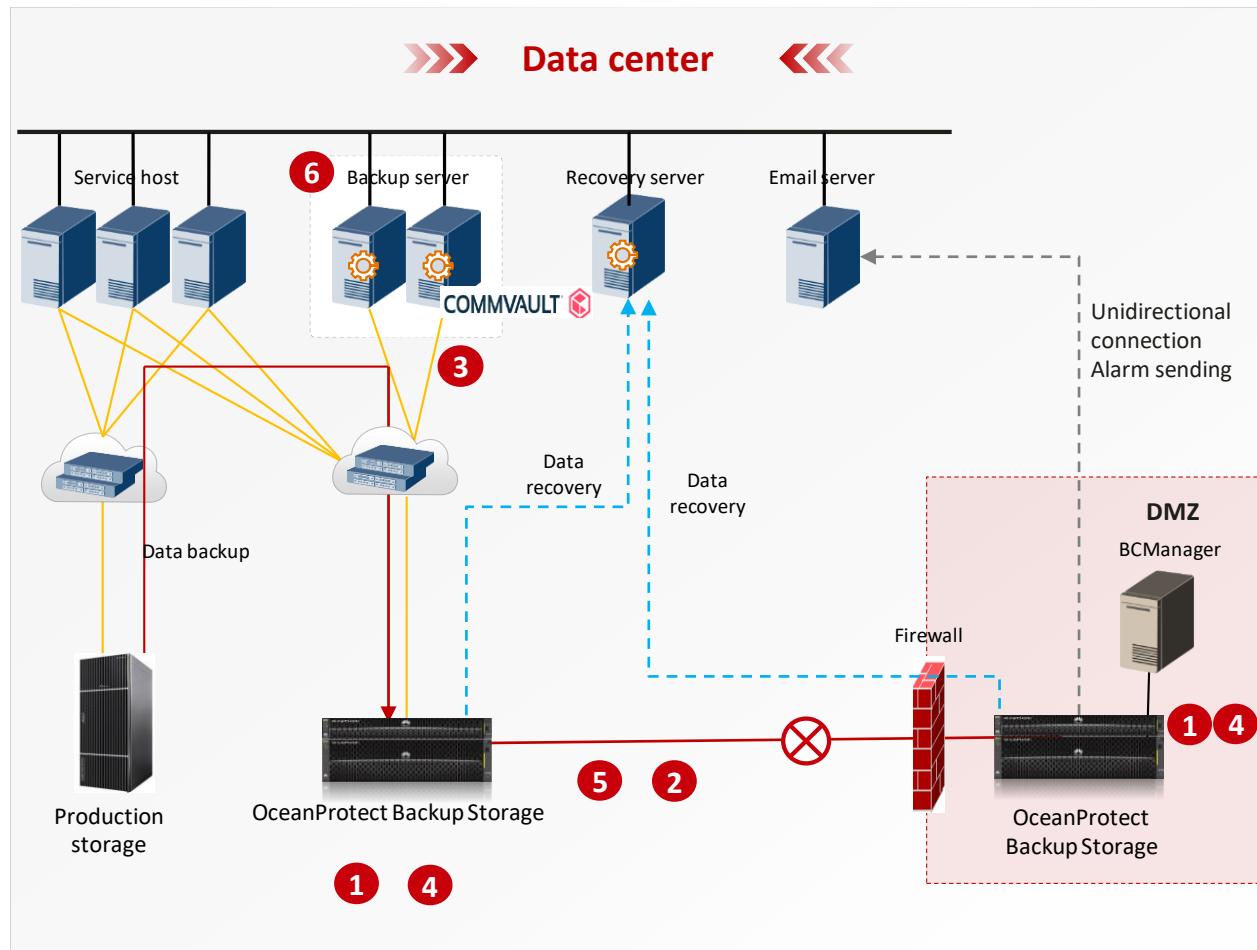
Ransomware often deletes backup data in the backup system, and then encrypts production data for a ransom. If no backup data is available, users have to pay the ransom to regain their data. Thus, the anti-tamper capability of backup copies is necessary.

Solution description:

1. OceanProtect provides a WORM file system at the storage hardware layer. You can enable this function to prevent data from being modified or deleted.
2. After a backup job is complete, CV invokes the OceanProtect interface to lock the backup files and retain them for a certain period.
3. The solution uses CV to implement WORM protection for specified backup copies at the backup storage hardware layer.

Commvault + OceanProtect Ransomware Protection Best Practice

Encryption, isolation, and immutable through the backup storage, which is more secure and reliable



Key technologies:

1	Storage encryption	Software-based encryption, preventing data leakage	Implemented by OceanProtect
2	Replication encryption	The replication links transmit data in ciphertext. The offloading of data encryption to NICs improves performance.	
3	SMB encryption	SMB packets are transmitted in ciphertext to prevent data from being stolen.	
4	File system WORM	Recommended when CV serves as the backup software. It prevents data written in the file system from being modified or deleted, or encrypted by ransomware.	
5	Air Gap	Automatically disables replication links to copy data to an isolation area for higher security.	
6	Detection and analysis	Implements proactive detection using backup software detection technologies (file system exception detection and honeypot detection) on the production system.	

Features in red are roadmap features and expected to be generally available on Sep. 30, 2022.

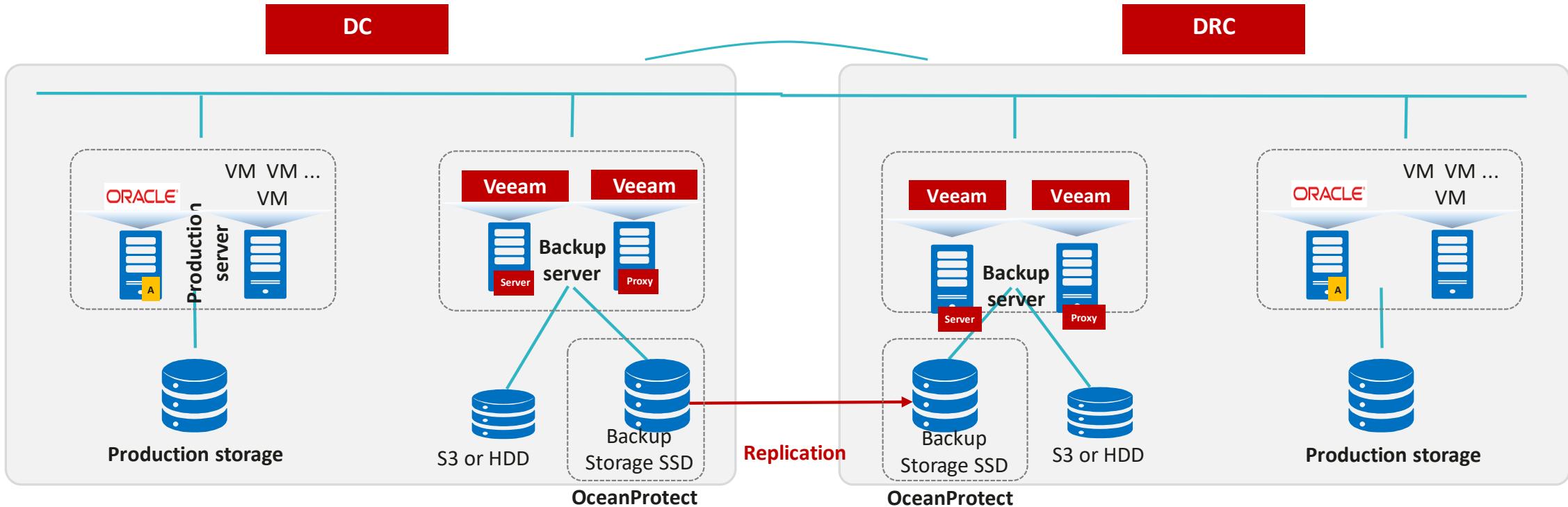
Main components:

Backup software	Commvault backup software
Backup storage	OceanProtect Backup Storage
Firewall(Optional)	Entry-level firewalls such as USG6100 are recommended.
Isolation zone	OceanProtect Backup Storage
BCManager server	-

Contents

1. Product Overview and Hardware & Software Architecture
2. High Data Reduction Ratio
3. High Performance and Reliability
4. **Centralized Backup Solution Best Practice**
 - Centralized Backup System Architecture
 - OceanProtect Backup Storage Solution Best Practice with Commvault
 - **OceanProtect Backup Storage Solution Best Practice with Veeam**

OceanProtect Backup Storage best practice with Veeam*



Backup solution: OceanProtect backup storage + Veeam

1. Backup and tiering: OceanProtect backup storage serves as level-1 backup storage. HDDs or S3 object storage can be used for long-term backup retention. Veeam is used for tiering.
2. Instant recovery: is mainly provided by Veeam. OceanProtect backup storage provides high bandwidth and temporary buffer (the temporary storage area of recovered instances).
3. Remote replication: OceanProtect backup storage can be used for remote replication, and Veeam for recovery.
4. Data dump: Data stored in the original backup storage can be migrated to the new OceanProtect backup storage using backup software.

*Remark: It is based on customer best practices, which is using the industry general NAS protocol capabilities.

Backup - Connecting to Veeam Using the NAS Protocol

1. Veeam Description

The Veeam backup software consists of the following components:

The backup server is a mandatory component for starting Veeam. It must run on 64-bit Windows OS. The backup proxy is located between the backup server and other components. The backup server schedules and manages backup jobs, and the backup proxy executes the backup jobs and transmits data. The backup proxy is responsible for:

- (1). Retrieval of VM data from the production storage
- (2) Compression, replication, and encryption
- (3). Data transmission to the backup repository

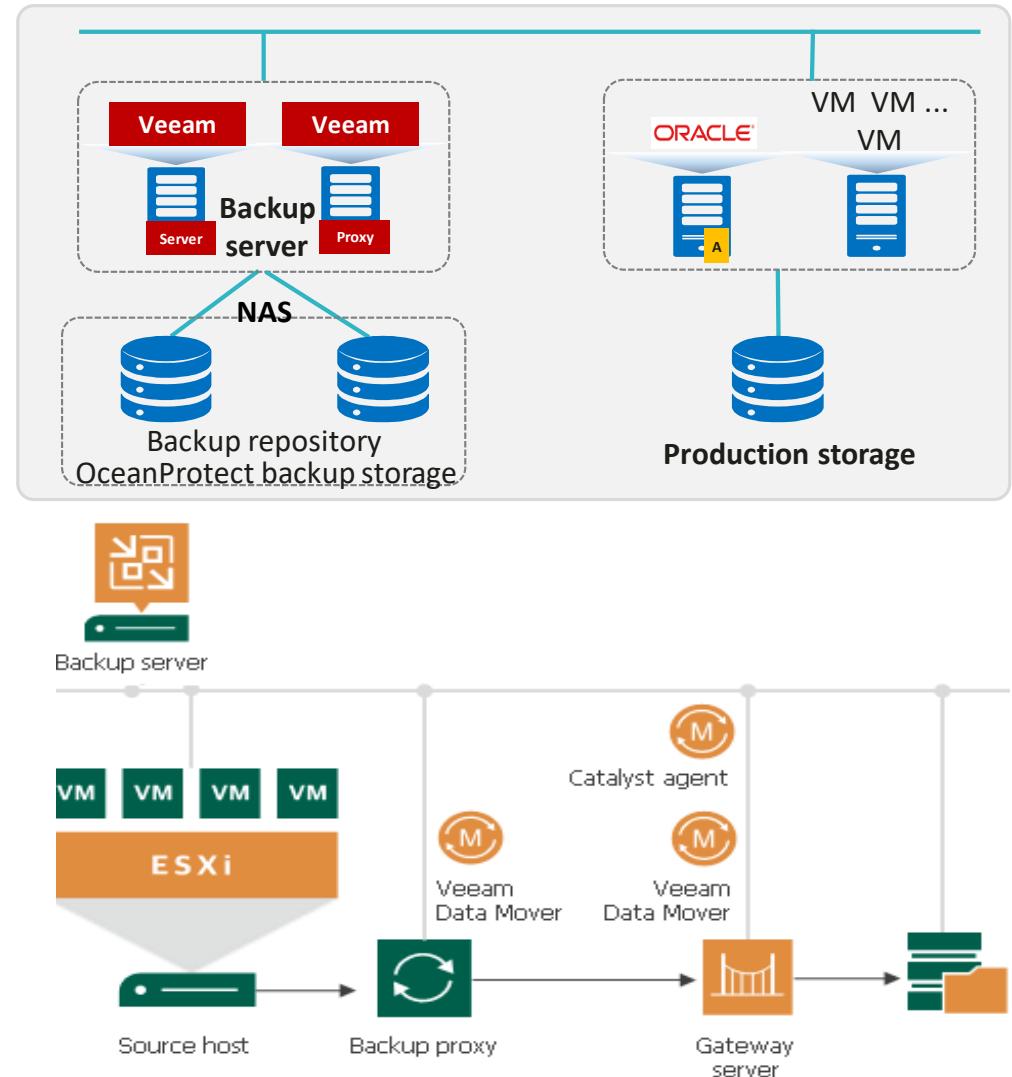
The gateway server connects the backup repository and backup server. Generally, the backup proxy also functions as the gateway server. In Hot-Add transmission mode, the backup proxy is a VM. In such a case, a physical host is used as the gateway server.

The backup repository is a storage unit for storing backup files and supports various storage systems.

2. Backup Storage Connection Description

- (1) On the Veeam backup server, add OceanProtect X8000 as the backup storage (backup repository).
- (2) OceanProtect backup storage deduplication function is recommended.

*Remark: It is based on customer best practices, which is using the industry general NAS protocol capabilities.



Instant Recovery

Function Description:

1. The Veeam Backup & Replication server functions as an NFS server and connects to the ESXi host through the NFS client built in the host. In this way, the backup VMs are started immediately.
2. The Storage vMotion function is used to migrate data from the OceanProtect backup storage to the production storage, ensuring service continuity.

Prerequisites: 1. You have enabled the vPower NFS mount service on the mount server. 2. The NFS server has sufficient storage space.

Features of Instant Recovery

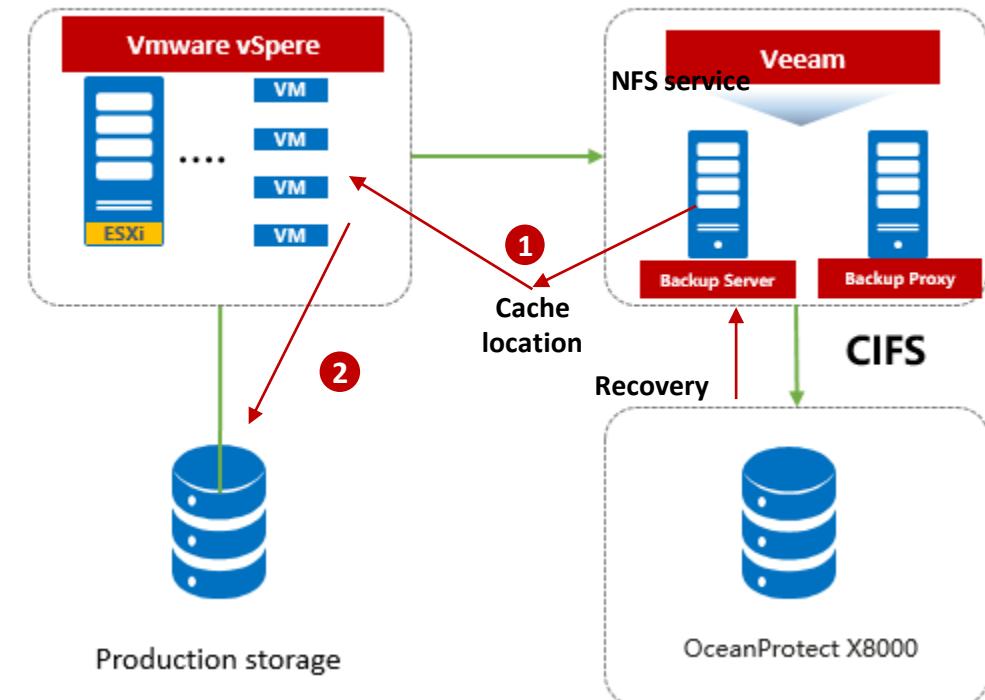
The instant recovery enables the startup of VMs from backup files without modifying the backup files.

The system status of the instantly recovered VMs is consistent with the status at the backup restoration time point, and the VMs can be read and written.

The instantly recovered VMs can be used for function tests and can be restored to the production system using the Storage vMotion function.

Application Scenario: Instant recovery is a high-speed recovery technology that enables data in the backup infrastructure to be available immediately instead of waiting for data retrieval for recovery.

*Remark: It is based on customer best practices, which is using the industry general NAS protocol capabilities.

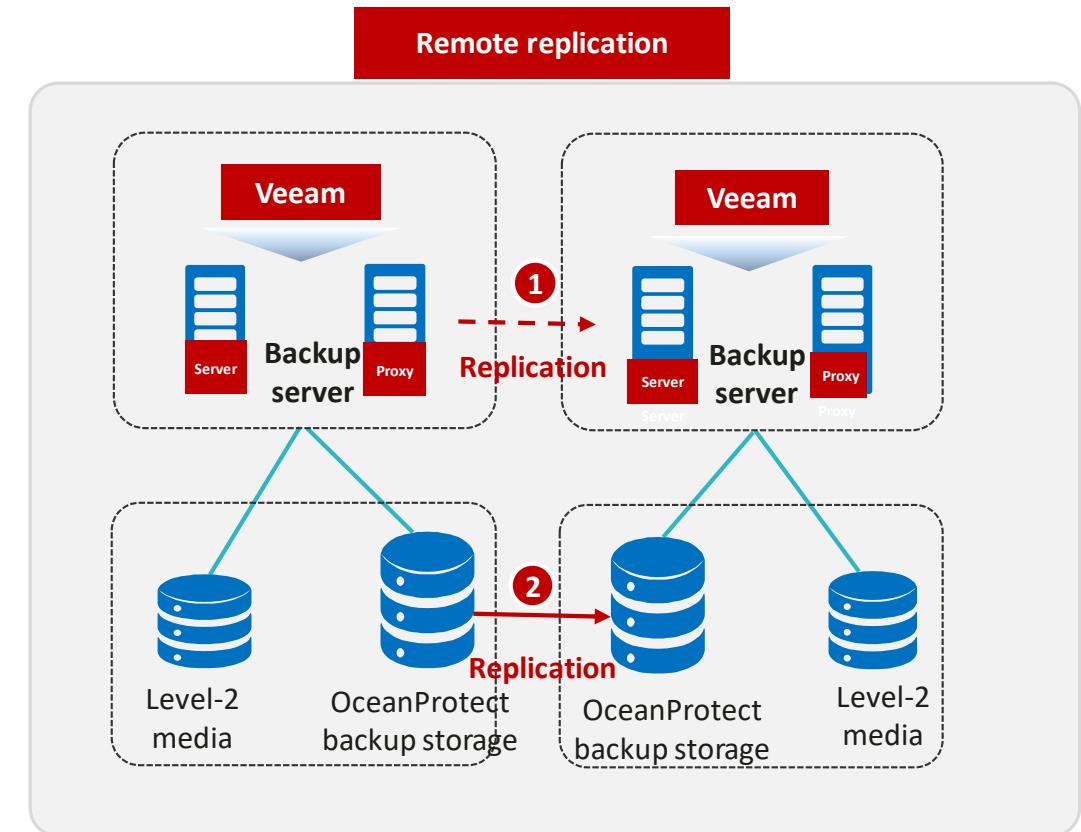


Remote Replication

Function Description

1. Replication by using the backup software: Use the Backup Copy function of Veeam for remote replication of backup jobs. Snapshot replication is supported for some applications. Replication after deduplication and compression is supported.
2. Use the OceanProtect backup storage for remote replication (HyperReplication). After the remote replication is created, use the backup software to import the data for recovery.
 - Create a file system replication pair. The synchronization period is 30 seconds.
 - Perform a full backup on the Veeam GUI.
 - After a full backup is performed, wait until the data synchronization is complete.
 - On the Veeam backup software page, scan for the snapshot copy of the backup image. Select the backup repository of the target end and choose **Rescan** from the shortcut menu.
 - Mount the file system from the secondary system to the Veeam server through SMB.
 - Choose **Home > Backups > Disks (Imports)**. Then right-click and choose **Import Backup**. On the displayed page, click **Browse** to select the corresponding mount. Then, enter the credential.
 - Select the desired VM for recovery.

Suggestion: Use the replication capability of the OceanProtect backup storage to save network bandwidth.



Note: 1. The start, stop, and process of OceanProtect backup storage replication are not perceived by the backup software. But policies can be set on the backup storage.

*Remark: It is based on customer best practices, which is using the industry general NAS protocol capabilities.

Tiering

Tiering:

1. The Veeam backup software supports multiple types of repositories, such as NAS and S3. Data can be stored in different repositories based on policies. The **Cloud Tier** function classifies copies that need to be archived for a long time to level-2 media.

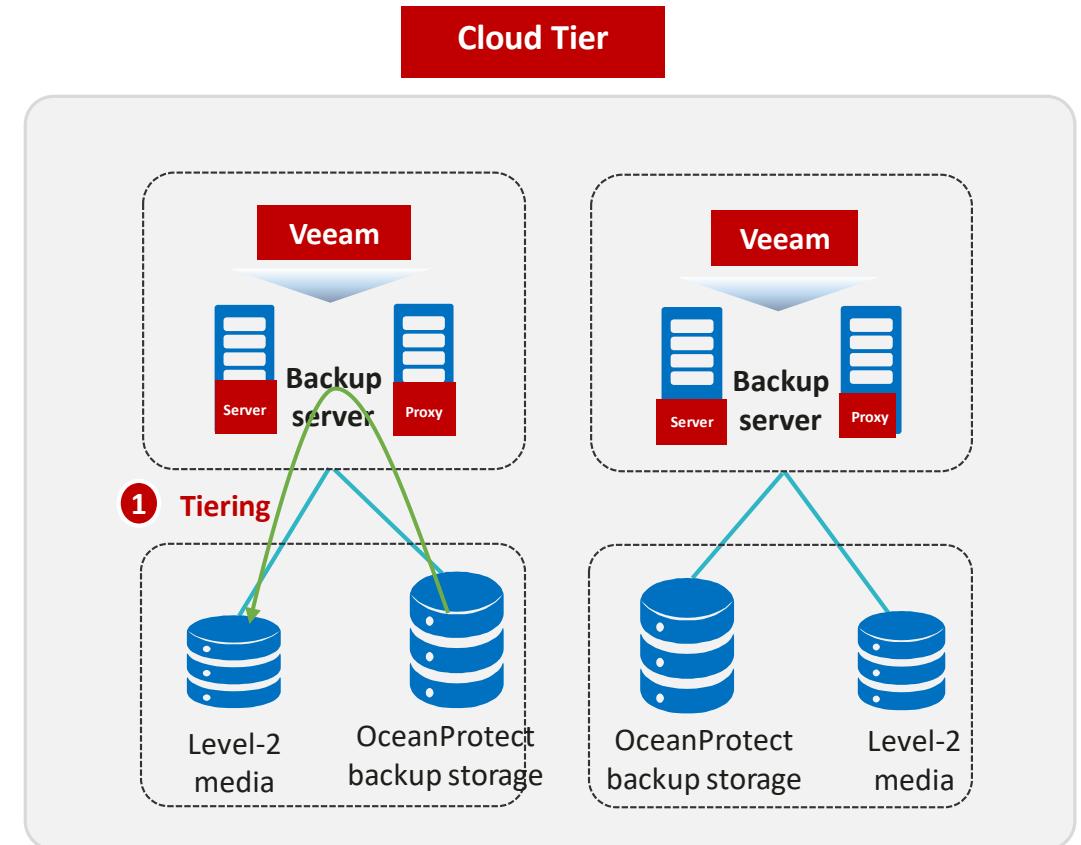
Application Scenarios:

1. Different tiers correspond to different long-term retention policies of backup data. When a large amount of application data must be retained for a long time, customers usually use capacity tier as the level-2 backup device to reduce costs.
2. The 3-2-1 Rule of the backup policy is used.

Advantages and Restrictions:

1. Policies are configured and managed on the backup software, and the tiering is controllable and manageable.
2. The backup software can directly recover long-term retained copies.
3. Backup server resources are occupied.

*Remark: It is based on customer best practices, which is using the industry general NAS protocol capabilities.



Data Dump

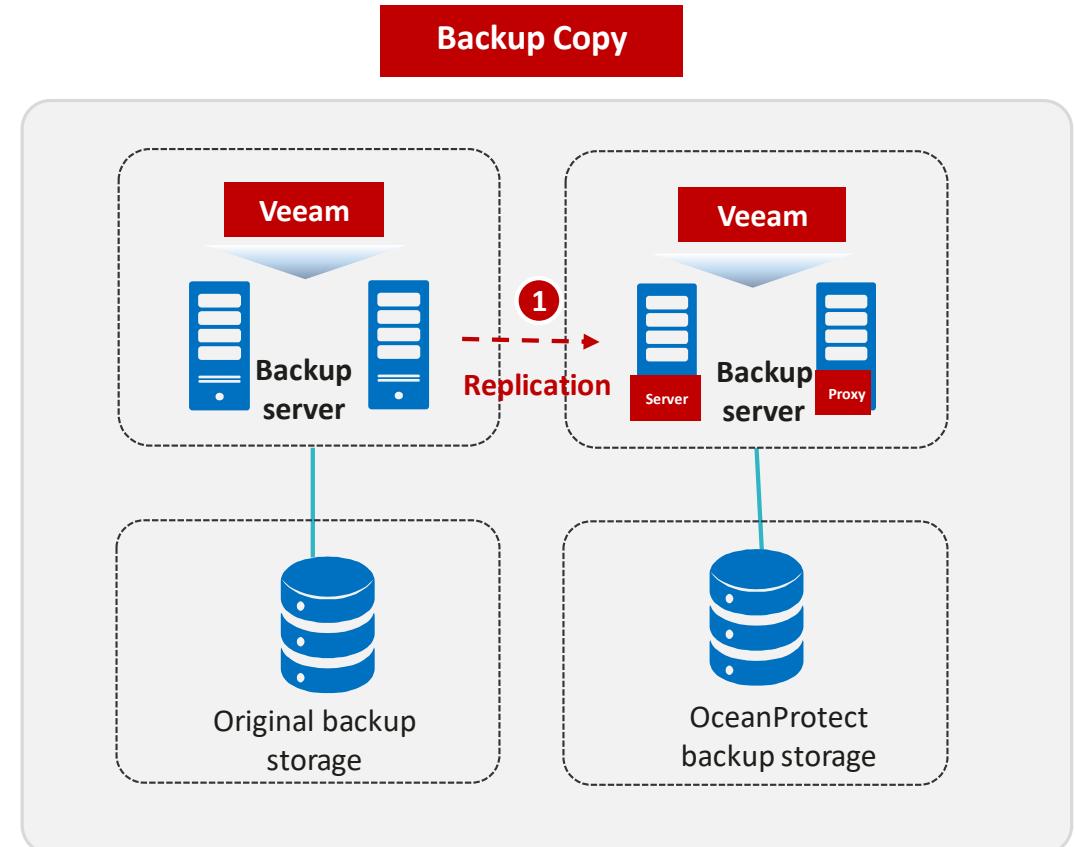
Solution description: The Backup Copy function of Veeam is used to copy data from the original backup storage to the OceanProtect backup storage.

Restrictions:

The Backup Copy function of Veeam consumes backup resources. Therefore, the function must be executed separately or at different time.

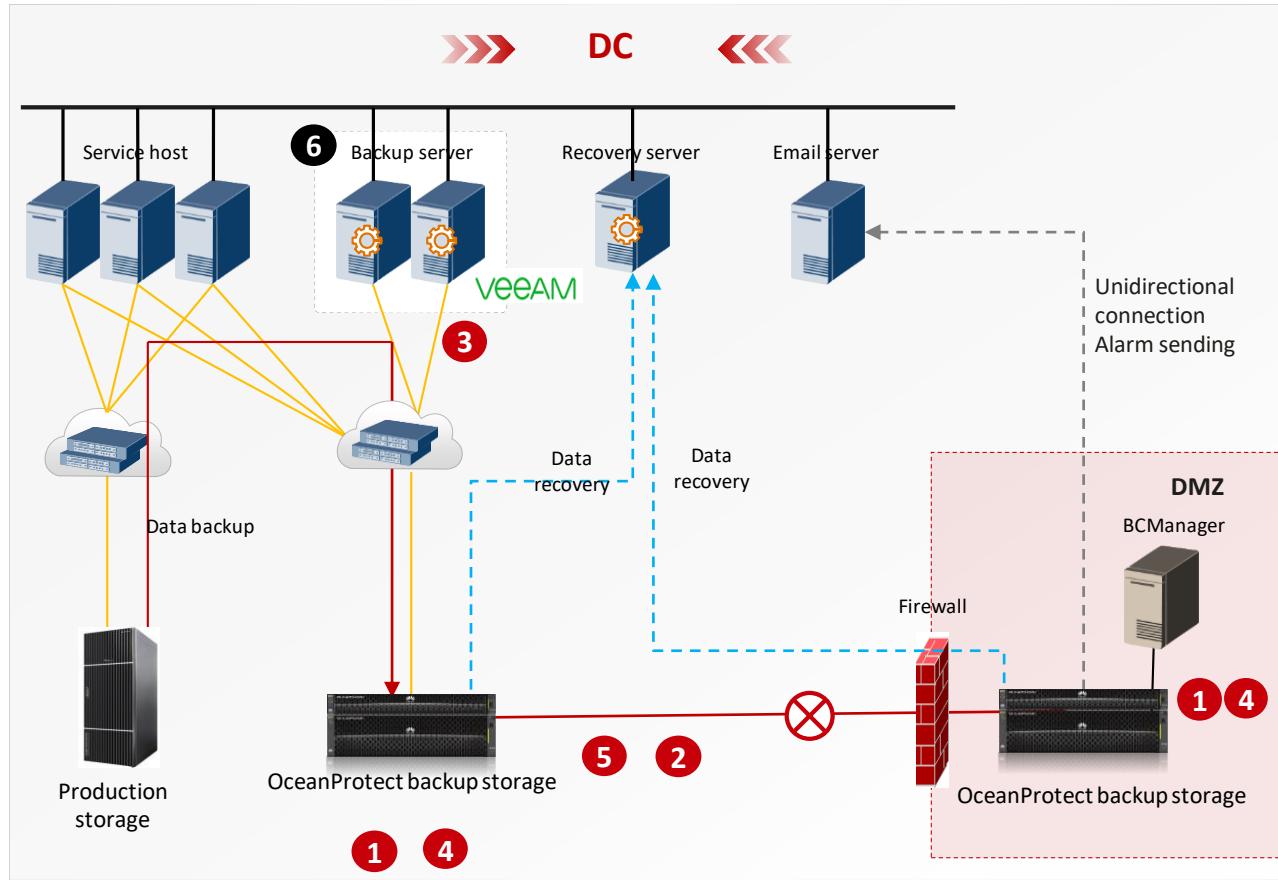
Application scenario: In backup storage replacement, data in the original backup storage needs to be retained and imported to the new backup storage.

*Remark: It is based on customer best practices, which is using the industry general NAS protocol capabilities.



OceanProtect Backup Storage Ransomware Protection Best Practice with Veeam*

Encryption, isolation, and anti-tampering through the backup storage, which is more secure and reliable



*Remark: It is based on customer best practices, which is using the industry general NAS protocol capabilities.

Key technologies:

1	Storage encryption	Software-based encryption, preventing data leakage	Implemented by the backup storage
2	Replication link encryption	The replication links transmit data in ciphertext. The offloading of data encryption to NICs improves performance (only for products in China region).	
3	SMB encryption	SMB packets are transmitted in ciphertext to prevent data from being stolen.	
4	Secure snapshot	The backup storage in the production center and security isolation area uses the secure snapshot technology to ensure that data is read-only and cannot be modified or deleted within a specified period.	
5	Air Gap	Automatically disconnects the replication link to copy data to an isolation area for higher security.	
6	Secure recovery	Before the recovery, start the antivirus software for scanning and restore the data to the production environment. Start the VMs in the isolated environment (Virtual Lab) to check whether they are normal.	

Main components:

★ Features of OceanProtect in red are roadmap features and expected to be generally available on Sep. 30, 2022.

Backup software	Veeam backup software
Backup storage	OceanProtect backup storage
Firewall(Optional)	Entry-level firewalls such as USG6100 are recommended.
Isolation end	OceanProtect backup storage
BCManager server	-

Quiz

1. (Single-choice) By now, which is not recommended when OceanProtect is working with Commvault?
 - A. Backup using NFS and CIFS
 - B. Back end terabyte (BET) license for all use cases
 - C. Backup sets replication by OceanProtect
 - D. Auxiliary copy should be used for data migration
2. (Multiple-choice) Veeam backup proxy is responsible for:
 - A. Retrieval of VM data from the production storage
 - B. Compression, replication, and encryption
 - C. Store backup files and supports various storage systems
 - D. Data transmission to the backup repository

Summary

- This chapter describes the hardware architecture, software architecture, data reduction, performance, reliability, features and principles of the OceanProtect backup storage system. It also describes the centralized backup solution best practice with Commvault and Veeam.

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 Data Management Solution



Foreword

- Huawei DMS(Data Management Solution) consists of three layers:
- The top layer is the cloud-based DME IQ which is a management platform deployed on Huawei Hi-Cloud. Huawei storage devices can connect to DME IQ for unified monitoring as long as they can access the network.
- The middle layer is DME STORAGE for data center management. The DME STORAGE can configure and manage multiple sets of storage devices and connect to existing third-party platforms to implement automatic resource provisioning, improving management efficiency.
- The bottom layer is DeviceManager which is used for single-device management.
- Through this three-layer management architecture, Huawei helps partners and end customers to simplify device management and improve management efficiency.

Objectives

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

- Describe the Huawei DMS 3-layer management architecture
- Describe the technical features and highlights of DME IQ
- Describe the technical features and highlights of DME STORAGE
- Understand the typical business scenarios of DME IQ and DME STORAGE

Contents

- 1. Brief Introduction to DMS**
2. DME IQ and DME STORAGE Features
3. Customer Cases

Overview and Objectives

- This section describes the overall introduction to Huawei DMS(Data Management Solution) and the DME IQ&DME STORAGE.
- On completion of this section, you will be able to:
 - Brief understand the three-layer management architecture
 - Brief understand the positioning of DME IQ and DME STORAGE

What do you think are the storage management needs
that administrators are most concerned about?



Key Issues to Be Resolved in Storage O&M Management

Secure, controllable, and traceable automation capabilities;
Comprehensive performance monitoring and analysis, report statistics, and centralized alarm management.

Discovery of Right Resources

- ✓ Integration of O&M and automation



Secure and Controllable Change Operations

- ✓ GUI operations
- ✓ Resource conflict check
- ✓ Automatic change execution

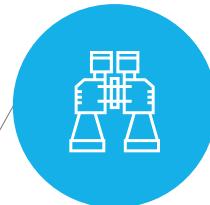


Audit and Tracing After the Change

- ✓ Visible change process
- ✓ Change result records



Plan & Deployment



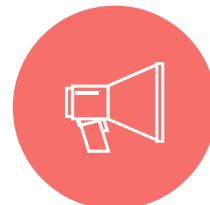
Problem Detection

- ✓ Centralized alarm management and monitoring
- ✓ Threshold definition and intelligent prediction



Problem Cause Analysis

- ✓ Storage network topology
- ✓ Storage performance monitoring and analysis
- ✓ Report statistics



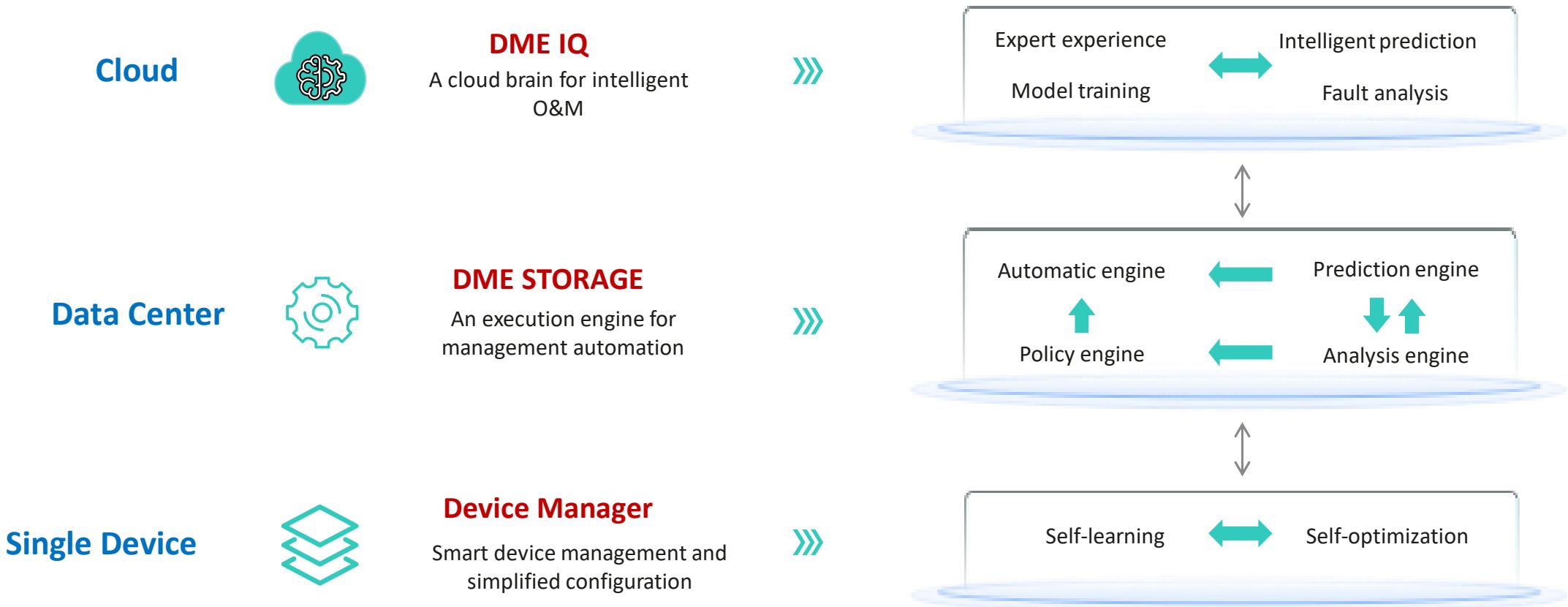
Problem resolution

- ✓ Troubleshooting suggestions
- ✓ Closed-loop automation

O&M & Optimization

DMS Three-Layer Management Architecture: Intelligent, Unified, and Efficient Architecture for Higher Management Efficiency

DMS(Data Management Solution) Three-Layer Management Architecture



DME IQ and DME STORAGE Usage Scenarios for Channel Partners or End Customers

Top 3 Scenarios to Choose DME IQ	VS	Top 3 Scenarios to Choose DME STORAGE
<p>Scenario 1:</p> <p>There are no frequent configuration requirements, mainly for device monitoring and alarm handling.</p>		<p>Scenario 1:</p> <p>Multiple devices need to be configured, including monitoring heterogeneous storage devices.</p>
<p>Scenario 2:</p> <p>Devices are not centrally located in one or two data centers. They are scattered and require remote monitoring.</p>		<p>Scenario 2:</p> <p>A large number of devices are centrally deployed in the data center, requiring management and reporting of multiple devices.</p>
<p>Scenario 3:</p> <p>DME IQ can be used as a tool to record and remind maintenance information, helping partners identify devices that need to be replaced after warranty expiration.</p>		<p>Scenario 3:</p> <p>Integrates into customers' existing management platforms to implement SLA-based automatic storage resource provisioning.</p>

DME IQ and DME Storage Function Comparison

Product Capability		DME IQ	VS	DME STORAGE
Planning Capability	Business Workload Simulation	YES		YES (2022.10.30)
	SLA-based Pooling and Consolidation	NO		YES
Deployment Capability	Multiple Service Provisioning	NO		YES
	To-Do Task	NO		YES
	Task and Log Tracing	NO		YES
O&M Capability	Storage Data Protection	NO		YES
	Centralized Alarm Monitoring	YES		YES
	Capacity Trend & Disk Problem Prediction	YES		YES
	Performance Analysis	YES		YES
	Automatic Problem Resolution	NO		YES
Optimization Capability	Resource Operation Management	NO		YES
	Heterogeneous Management	NO		YES
	Interconnection with Third-Party Platforms	NO		YES
	Callhome	YES		YES
	Intelligent Question and Answer Robot	YES		NO
Mobile O&M App		YES		NO

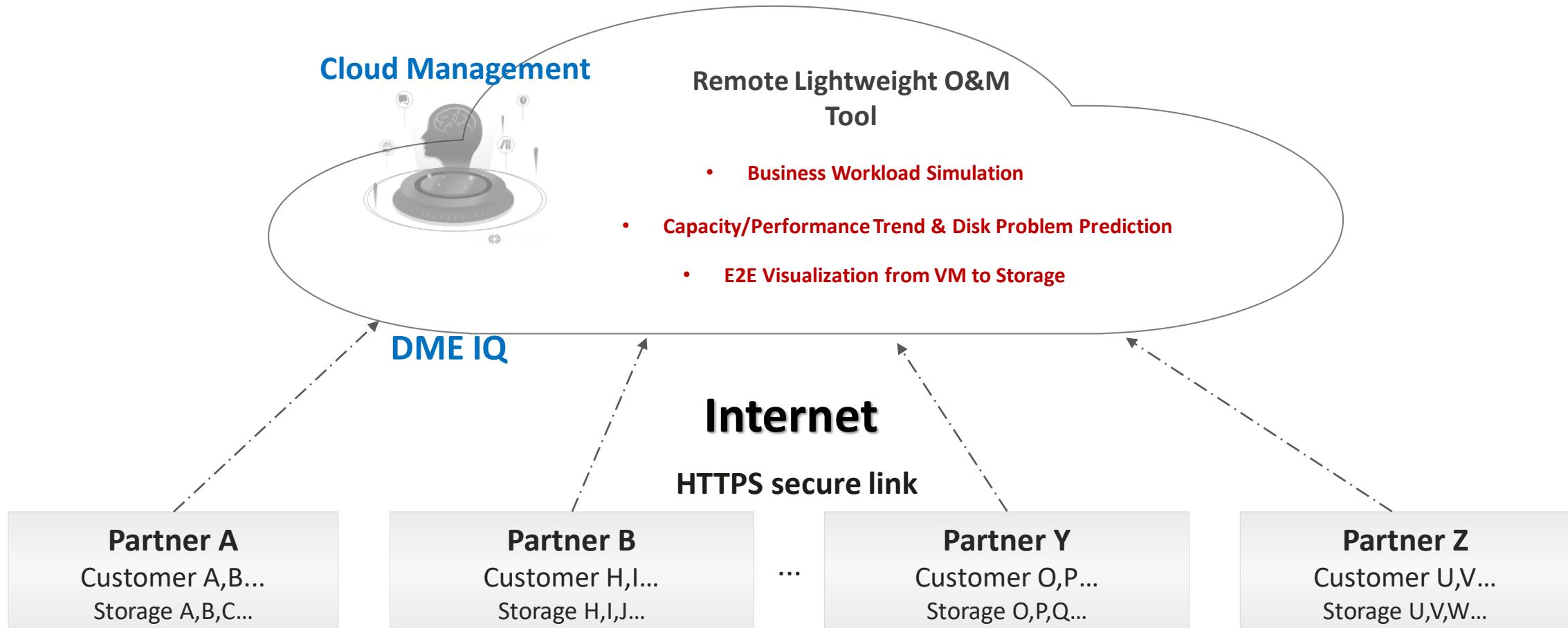
Contents

1. Brief Introduction to DMS
- 2. DME IQ and DME STORAGE Features**
3. Customer Cases

Overview and Objectives

- This part describes the features and highlights of DME IQ
- On completion of this section, you will be able to:
 - Understand the positioning of DME IQ
 - Understand the highlights about planning and remote management of DME IQ
 - Brief understand the advantages of DME IQ compared with other vendors

DME IQ Enables Partners to Remotely Plan and O&M End-customer Storages



Planning Business Workload Simulation to Get Right Storage Resources

As-Is

It's hard to decide which storage to use on service change

»» Service Rollout ««

How to allocate proper resources for new business service?

»» Service Scaling ««

How to assess the workload is fine when the application meet big business activities?

»» Service Migration ««

How to choose the proper storage to migrate services from EOS storage?

To-Be

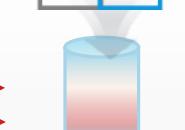
Multiple Applications

> 1000 IOPS/TB
capacity: 150TB

**Input workload
And service type**



Select storage allocated to



workload: 50%
usage: 40%



On-click trigger simulate



workload: 70%
usage: 60%

Get the result



Evaluate whether selecting storage resources can **meet application load requirements**

Evaluate whether selecting storage resources can meet requirements **when the load suddenly increases exponentially**

Evaluate whether **the new storage after data migration** can meet the load requirements



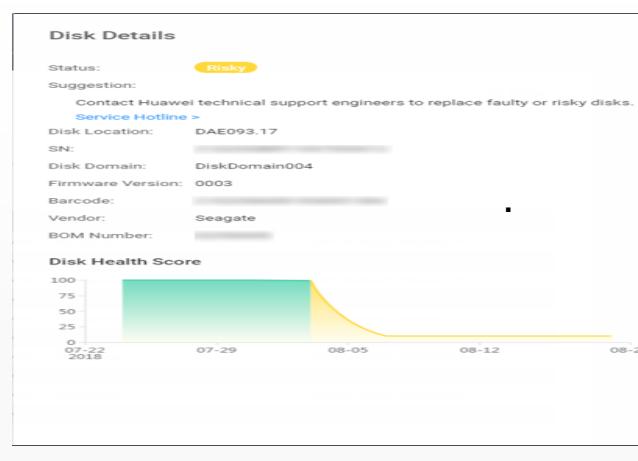
Plan 1 results in Storage overload



Plan X works well

DME Monitor Eliminates Risk Ahead to Ensure Stable Service Experience

✓ Potential Risks Prediction



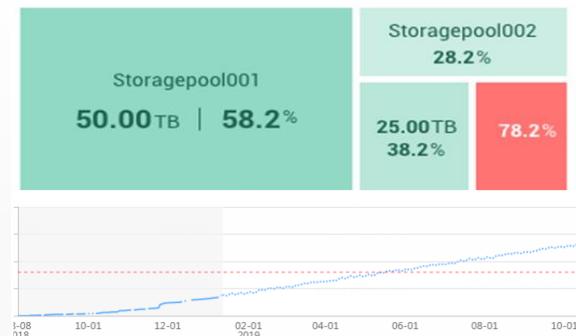
Predict disk risks 14 day ahead

600,000+ disks and 20+ billion feature records for AI training

Identify 80% disk risks with a 0.1% mis-reporting rate

Storage Pools Usage

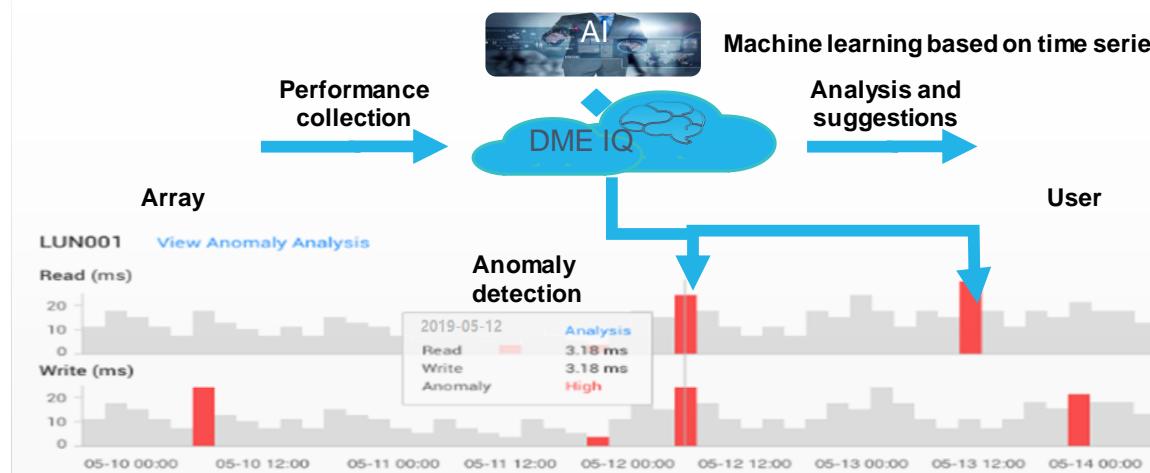
Capacity & Utilization



Predict the capacity trend in the next 12 months and determine the capacity requirements.

Predict the business performance trend 60 days ahead to plan before performance bottleneck occurred.

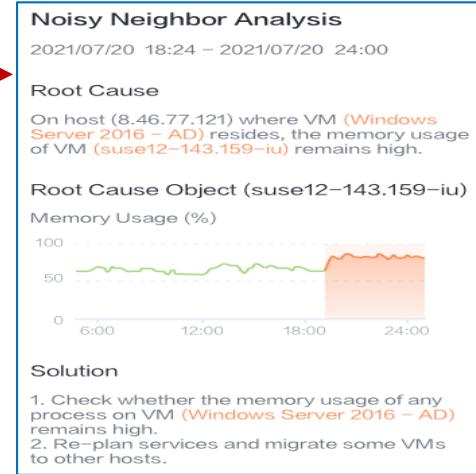
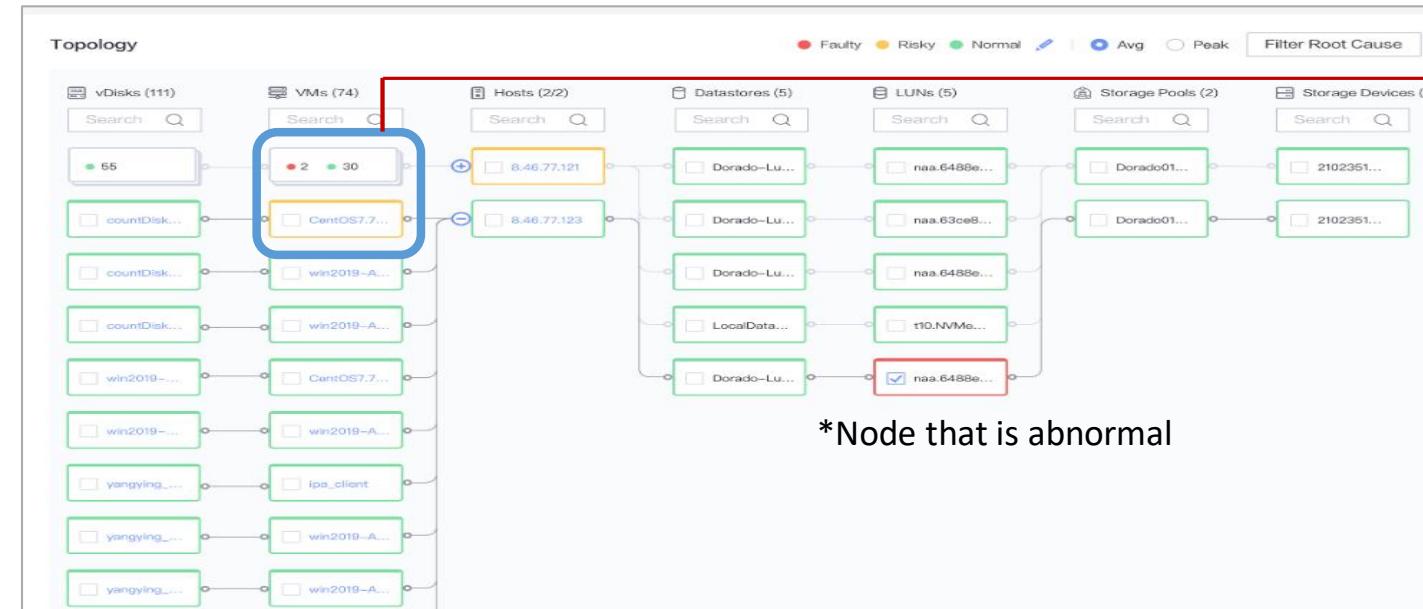
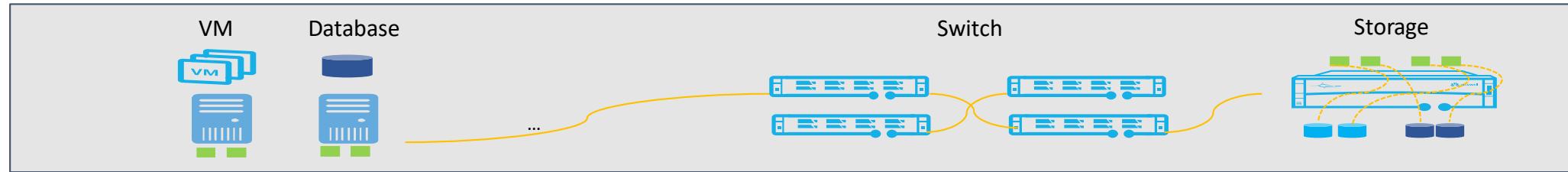
✓ Performance Anomalies Auto Detection



Intelligently identifies device performance anomalies and provides analysis and suggestions to eliminate performance problems in time and ensure stable service running.

E2E Visualization from VM to Storage, Fast Identify Storage Performance Problem

✓ DME IQ Offers E2E Topology Visualization



With multiple nodes share the same infrastructure resource, DME IQ provides intelligent root cause analysis when resource contention occurs within multiple upper VMs

Competitiveness Comparison Between DME IQ and Competitors

The DME IQ has strong problem analysis and diagnosis capabilities, especially leading hardware fault diagnosis and prediction capabilities

Function Comparison	HUAWEI DME IQ	Vendor H InfoSight	Vendor P Pure1	Vendor D CloudIQ	Vendor N ActiveIQ
Full-stack data collection	★★★	★★★★★	★★★	★★★	★★★★★
Risk prediction	★★☆	★★☆	★★☆	★★☆	★★
Problem detection	★★★★★	★★★★☆	★★★	★★☆	★☆
Root cause analysis	★★★★☆	★★★★☆	★★★	★★★	★★★
Issue closure	★★★	★★★★★	★★★★☆	★★	★★★★★
Change Assessment	★★★★☆	★★★★★	★★★★☆	★	★

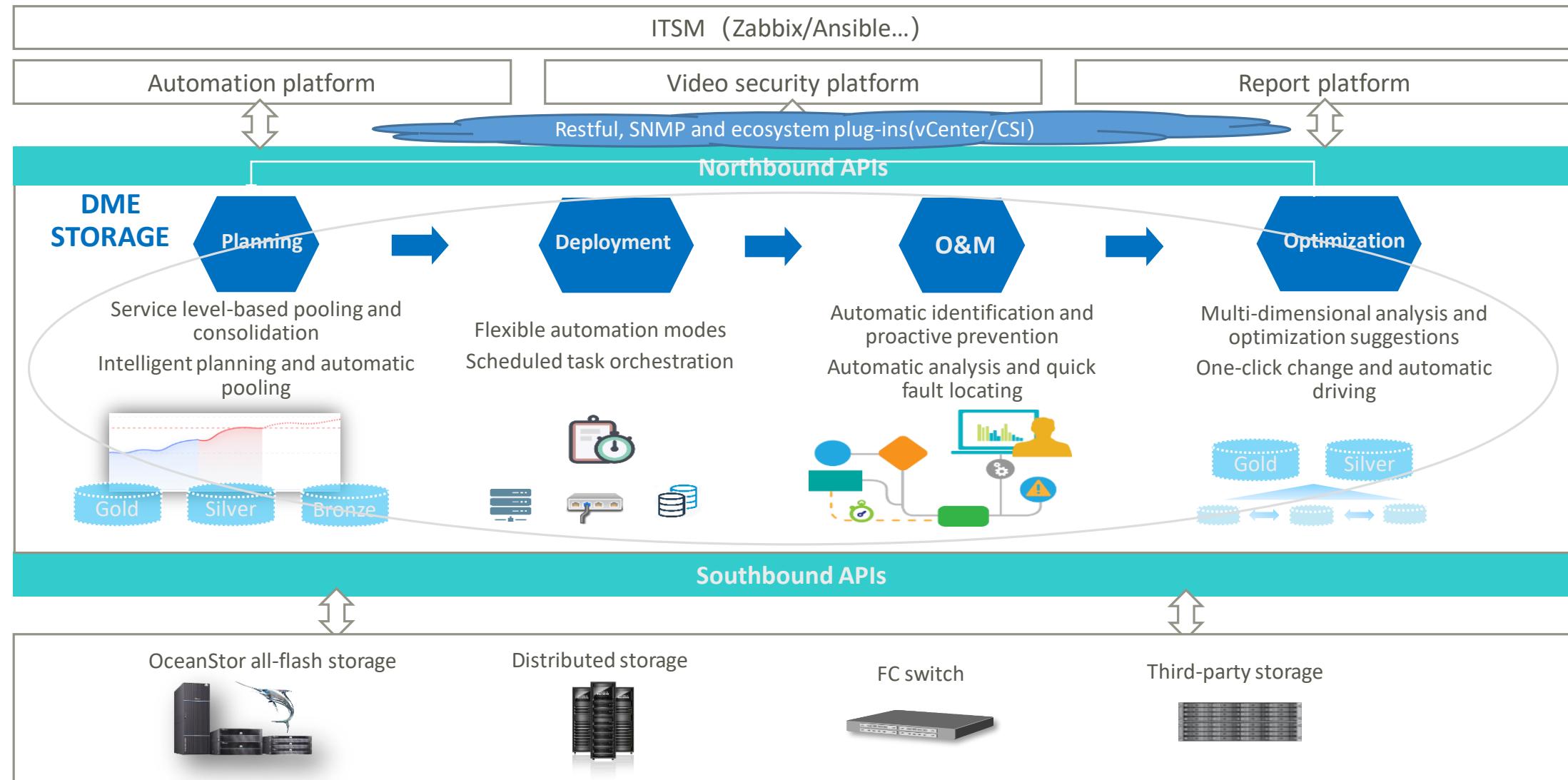
Based on the previous description, what do you think is the greatest value that DME IQ can bring to you?



Overview and Objectives

- This part describes the features and highlights of DME STORAGE
- On completion of this section, you will be able to:
 - Understand the positioning of DME STORAGE
 - Understand the highlights about planning, deployment, O&M and Optimization of DME STORAGE
 - Learn more about multiple automatic provisioning modes of DME STORAGE to improve user management efficiency
 - Brief understand the advantages of DME STORAGE compared with other vendors

DME STORAGE: Full-Lifecycle Automated Management Platform



SLA-based Pooling and Consolidation, Improving Storage Resource Utilization

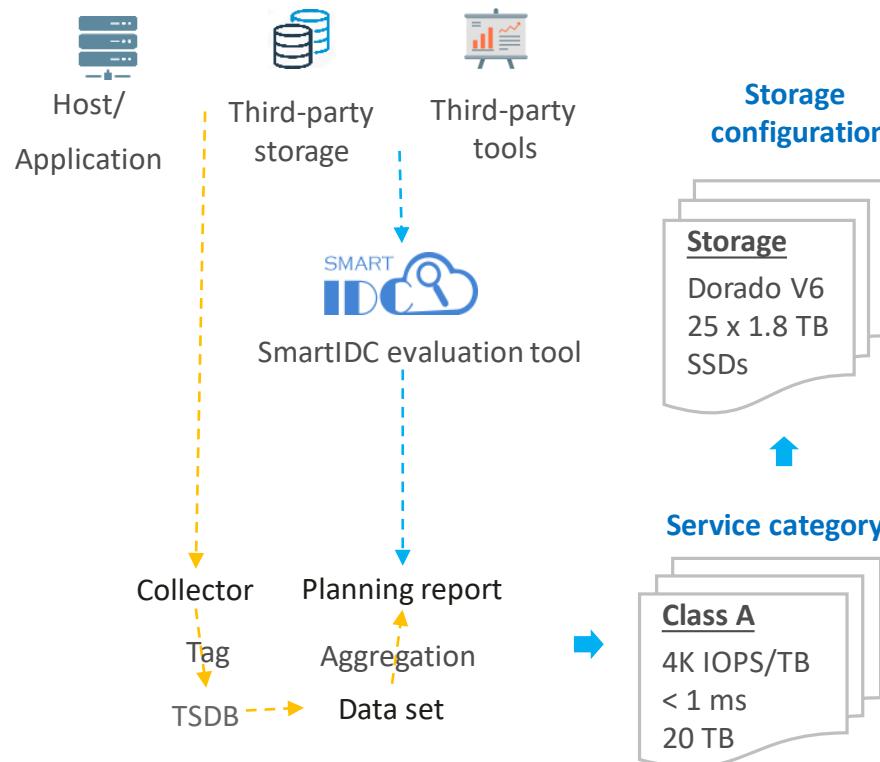
Planning

Deployment

O&M

Optimization

Workload Assessment



SLA definition

Class A applications
ERP, CRM, and ABC

Class B applications
VM and data warehouse

Class C applications
Office and file sharing

QoS: > 4K IOPS/TB, < 1 ms
Protection: active-active/snapshot
Balancing: performance-based
Quota: Project group A: 8 TB
Project group B: 5 TB

QoS: > 2K IOPS/TB, < 5 ms
Protection: synchronous replication
Balancing: performance-based
Quota: Project group A: 20 TB
Project group B: 10 TB

QoS: < 1K IOPS/TB
Protection: asynchronous replication
Balancing: capacity-based
Quota: Project group A: 50 TB
Project group B: 20 TB

Gold

20 TB

Silver

50 TB

Bronze

100 TB

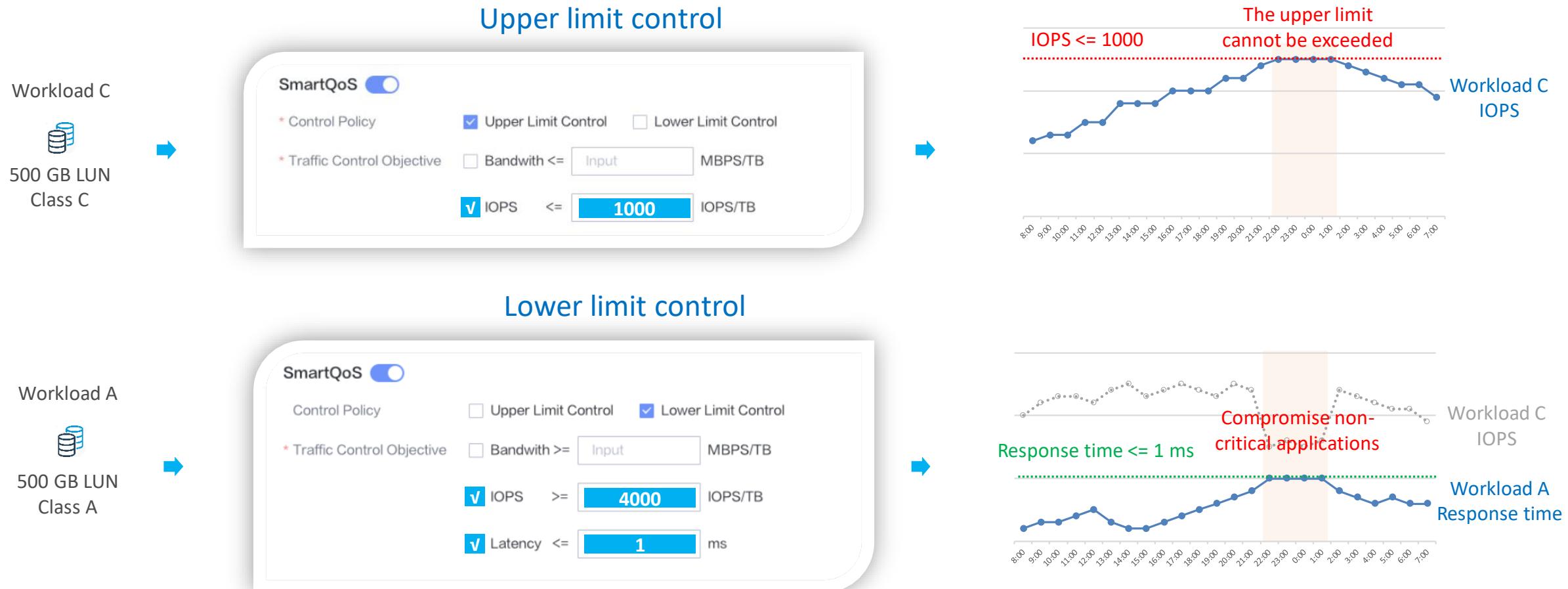
SLA-based Pooling and Consolidation: QoS Policy

Planning

Deployment

O&M

Optimization



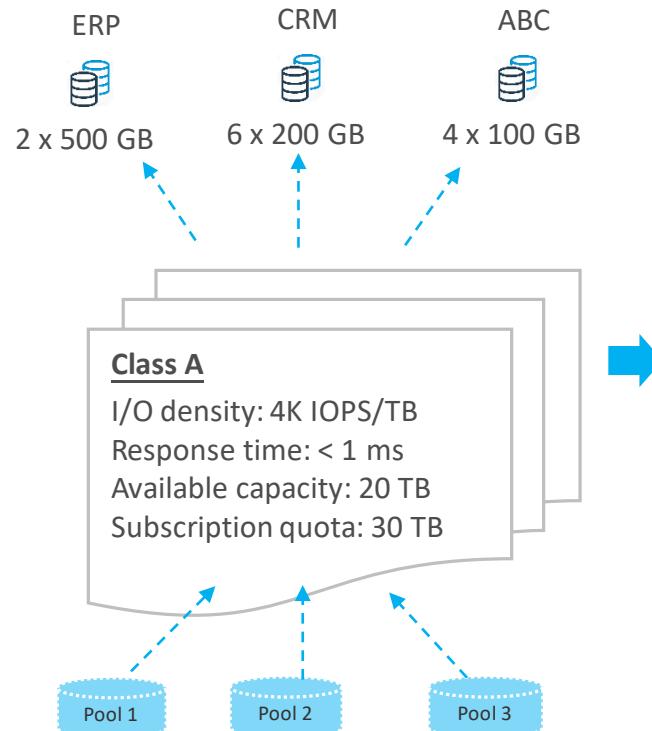
SLA-based Pooling and Consolidation: Threshold Policy

Planning

Deployment

O&M

Optimization



Capacity threshold (Enabled)

An alarm is reported if the objects meet **all** of the following conditions:

- Capacity Usage: $\geq 80\%$
- Free Capacity: $\leq 1000\text{ GB}$

LUN Performance Threshold (Enabled)

An alarm is reported if the objects meet **all** of the following conditions:

- LUN read I/O response t...: $\geq 10\text{ ms}$
- LUN write I/O response t...: $\geq 10\text{ ms}$

Class A

QoS: > 4K IOPS/TB, < 1 ms

Protection: active-active/snapshot

Balancing: performance-based

Class B

QoS: > 2K IOPS/TB, < 5 ms

Protection: synchronous replication

Balancing: performance-based

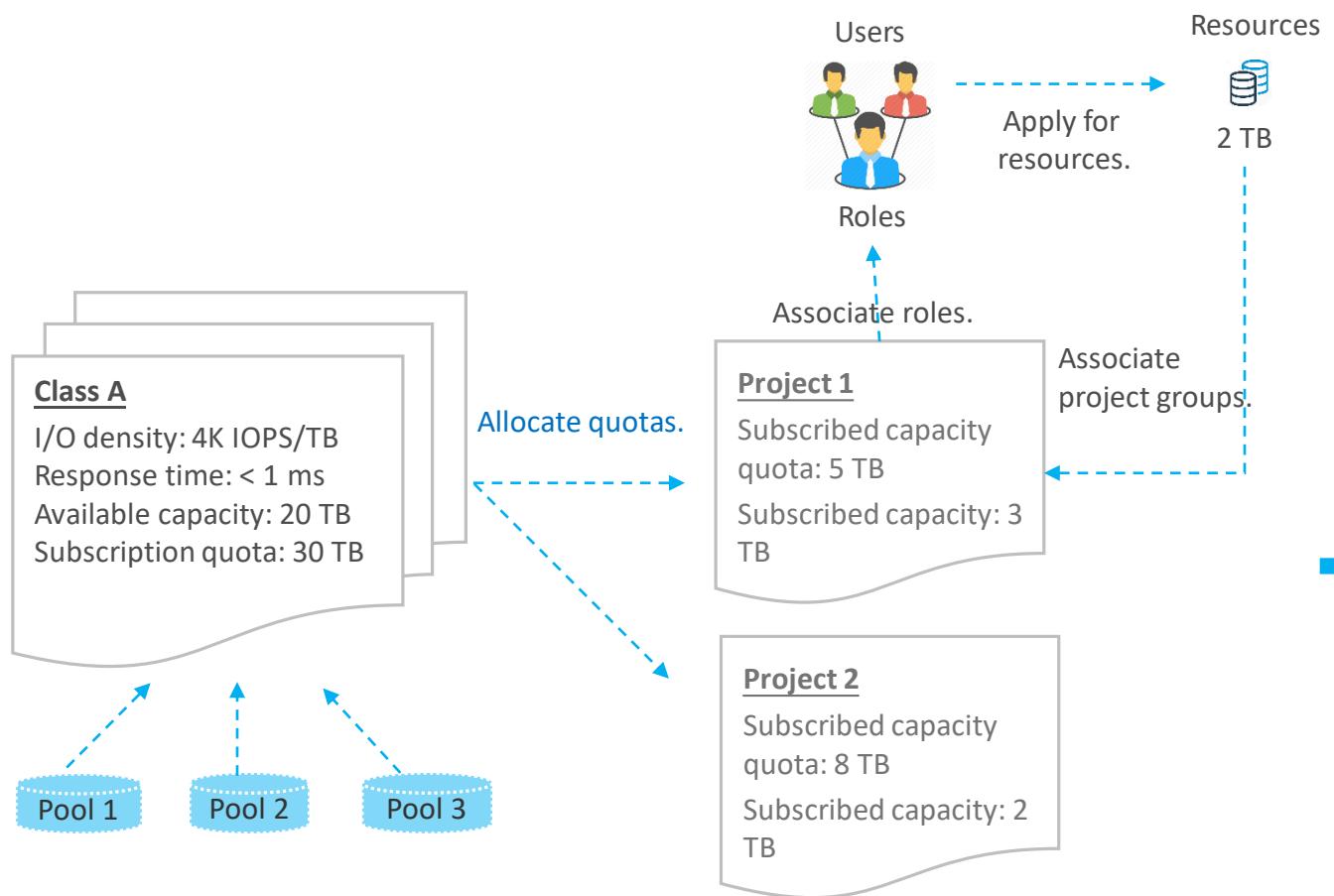
Class C

QoS: < 1K IOPS/TB

Protection: asynchronous replication

Balancing: capacity-based

SLA-based Pooling and Consolidation: Quota Policy



p_2 Project 1

General | Host | Host Group | Volume

Basic Information

Name	p_2
Description	Project 1

Associated Service Levels

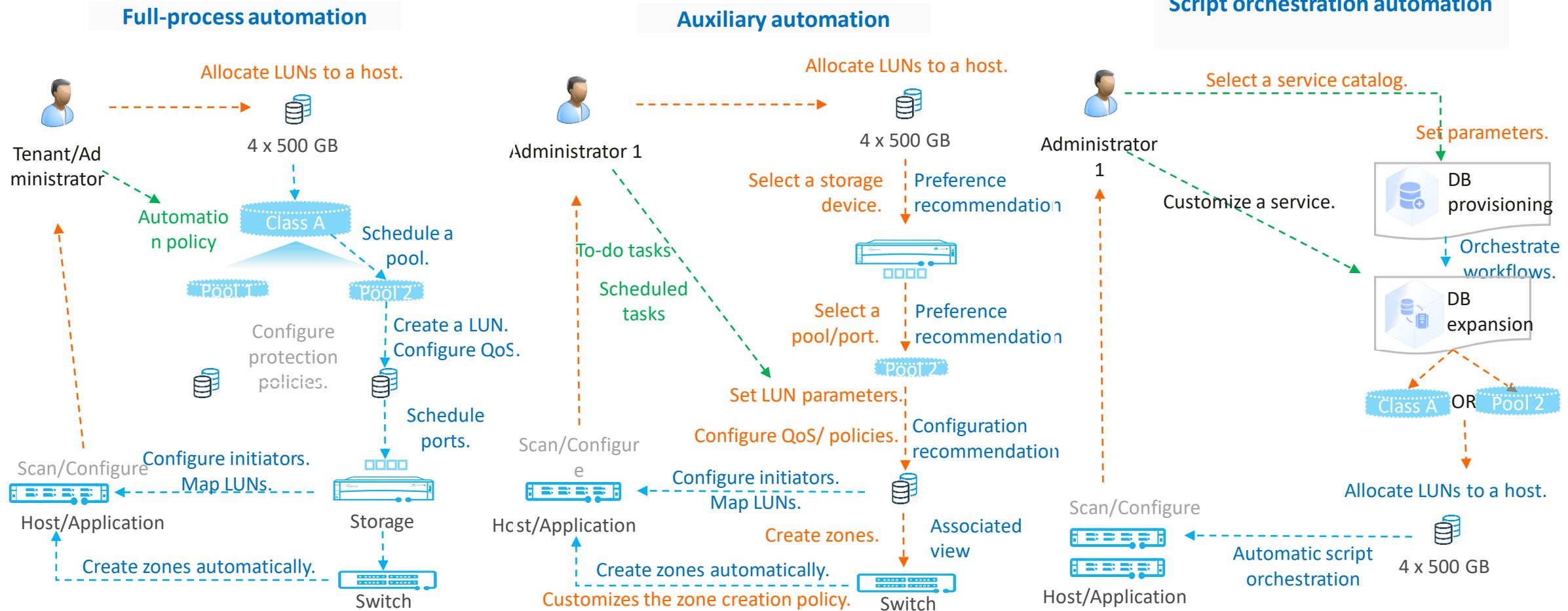
Add Remove

223_sla_1	Modify
Class A	0.00%
Capacity	Subscribed 0.000 GB/5.000 TB

卷基本功能测试	Modify
Class B	0.00%
Capacity	Subscribed 0.000 GB/5.000 TB

Multiple Service Provisioning Capabilities to Meet Different Customer Requirements

Planning Deployment O&M Optimization



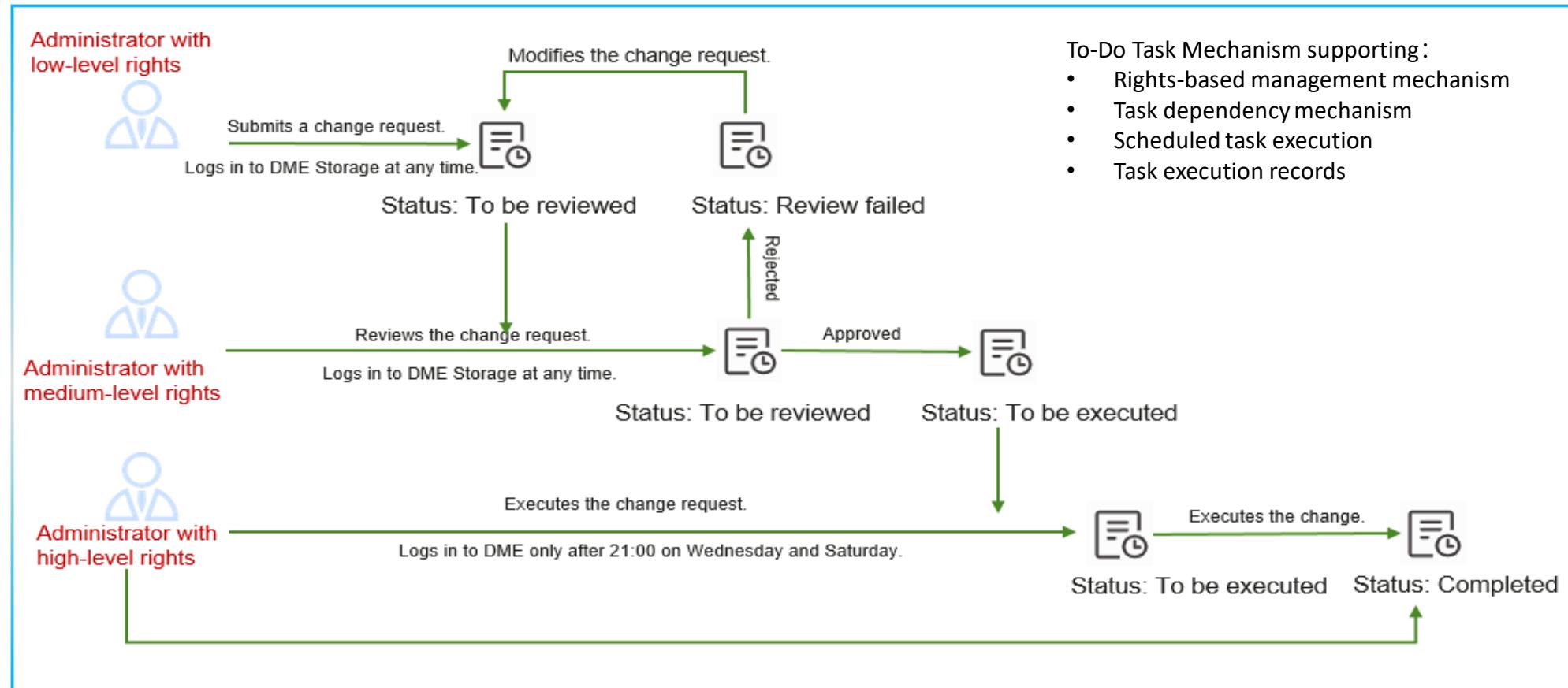
To-Do Task Mechanism Builds Secure, Controllable, and Traceable Automation Process

Planning

Deployment

O&M

Optimization



Automation: Task and Log Tracing

Planning

Deployment

O&M

Optimization

Create volume

Pre-check	Status: 100% Succeeded Start Time: 2019-11-04 22:22:41 Duration: 17s... Details: --			
Subtask	Status	Start Time	Duration	Details
Check volume name	100% Succeeded	2019-11-04 22:22:41	8seconds	--
Check LUN ID	100% Succeeded	2019-11-04 22:22:50	3seconds	--

Basic Information

Check connectivity	
Check advanced feature	
Check capacity	100% Succeeded Start Time: 2019-11-04 22:22:58 Duration: 9s... Details: --

Task

Created By	
End Time	2019-11-04 22:23:45
Description	--

Progress

Task Progress	100% Succeeded
Pre-check	5/5 Succeeded
Create volume	5/5 Succeeded
Create zone	1/1 Succeeded
Map volume	5/5 Succeeded

Create volumes.

Subtask	Status	Start Time	Duration	Details	Device
Create volume cst-1104...	100% Succeeded	2019-11-04 22:22:59	5seconds	--	Storage.5600v3

Create a zone.

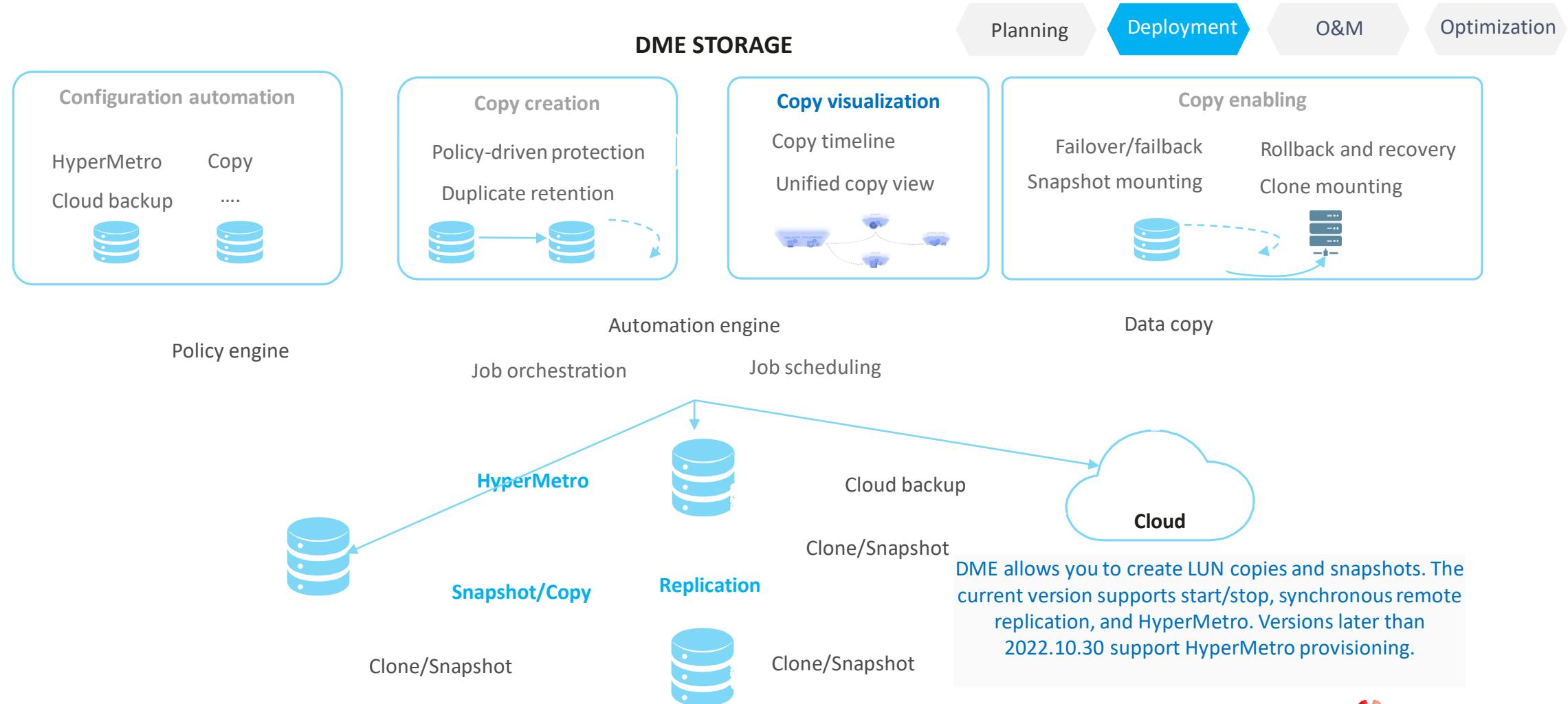
Subtask	Status	Start Time	Duration	Details	Switches
Create zone DJ_154...	100% Succeeded	2019-11-04 22:23:10	15seconds	--	fabric.196.231

Map volumes.

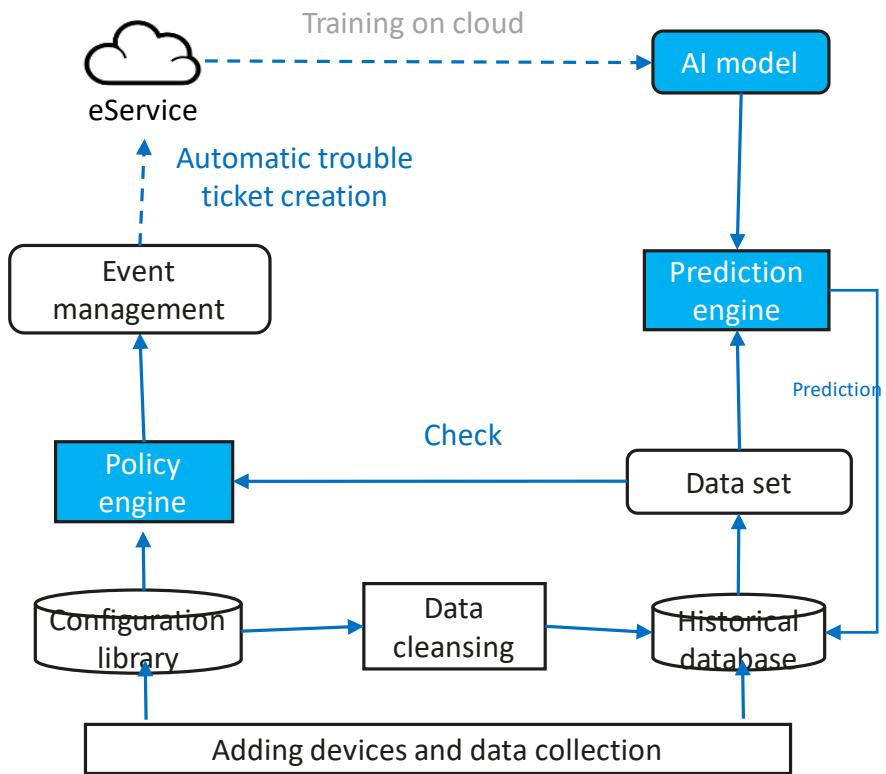
Subtask	Status	Start Time	Duration	Details	Host
Map volume cst-1104...	100% Succeeded	2019-11-04 22:23:27	4seconds	--	localhost.localdomain
Map volume cst-1104...	100% Succeeded	2019-11-04 22:23:32	4seconds	--	localhost.localdomain

Detailed logs are recorded when the operation fails.

Building Storage Data Protection Capabilities with Focus on Protected Objects



Proactive Problem Identification



Fault detection: Manual monitoring --> Automatic analysis

Fault processing: Reactive fault remediation --> Proactive fault prevention

- # Planning Deployment O&M Optimization

Performance check and prediction

- Storage resources/Hosts/Switches
 - Service levels



Configuration check

- Attribute, version, and warranty
 - Changes and associated resources



Availability check and prediction

- Disk health status/service life
 - Hardware and resource status



Capacity check and prediction

- Usage and free capacity
 - Remaining duration of availability



Proactive Problem Identification: Performance Threshold Check

Read I/O response time < 10 ms; Average I/O size <= 8 KB
15 consecutive points (1 point per minute)

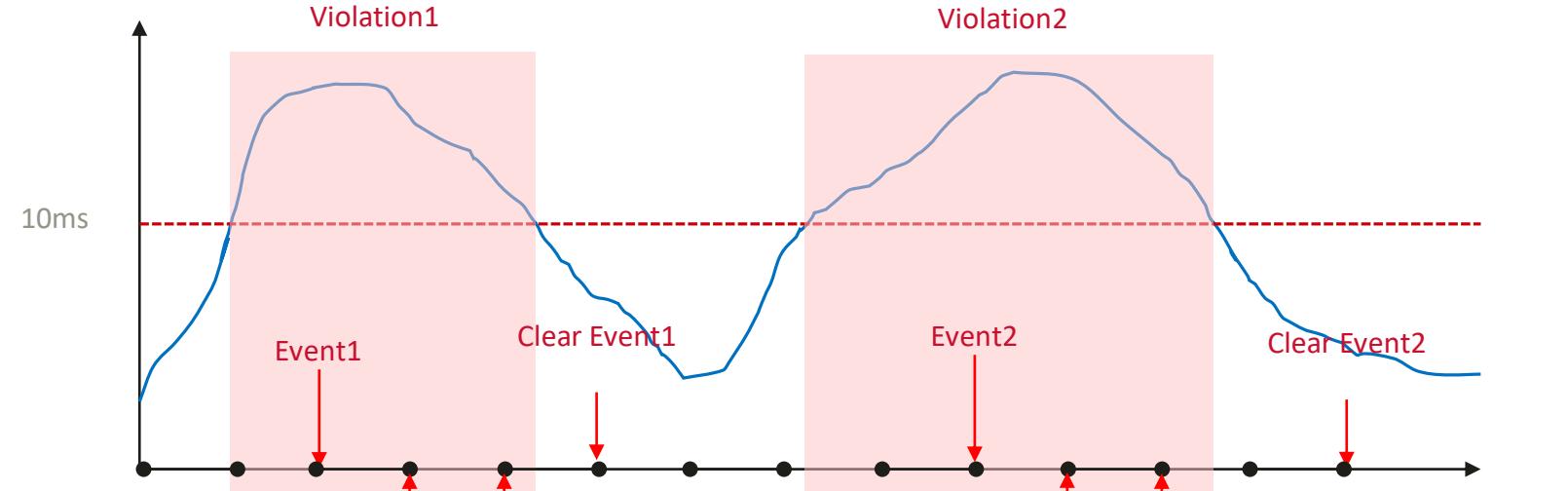
Planning

Deployment

O&M

Optimization

Read response time

**Violation1:**

First violation -> generate event1
Next violation -> update last occurrence time
No violation -> clear event1

Violation2:

First violation -> generate event2
merge event1 to event2
Next violation -> update last occurrence time
No violation -> clear event2

Proactive Problem Identification: Capacity Prediction

Planning

Deployment

O&M

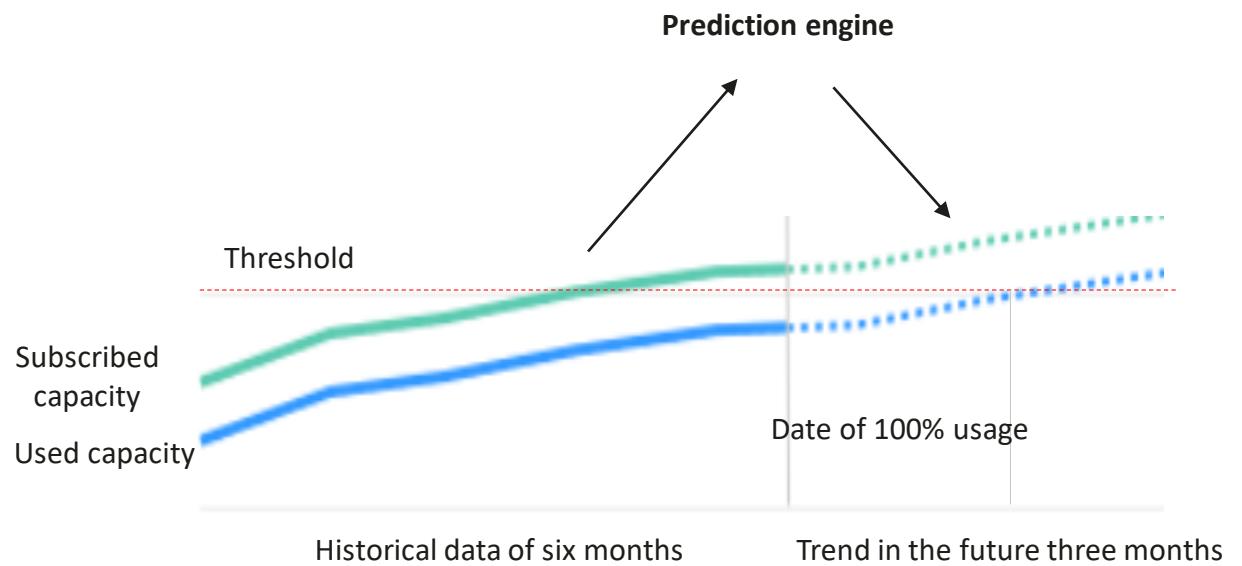
Optimization

Capacity data collection

- Collect capacity data of storage devices, pools, and file systems and aggregates the data based on service levels, storage devices, and storage pools.
- Data is collected every 30 minutes and stored for six months.
- Data aggregated once a day will be stored for two years. Data aggregated once a month will be stored for five years.

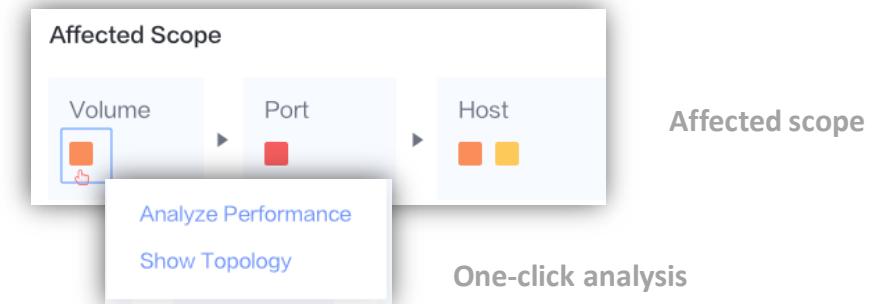
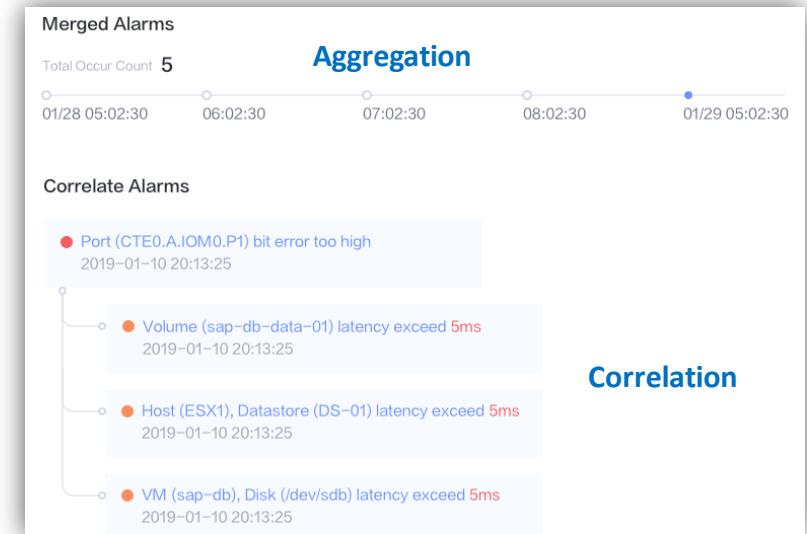
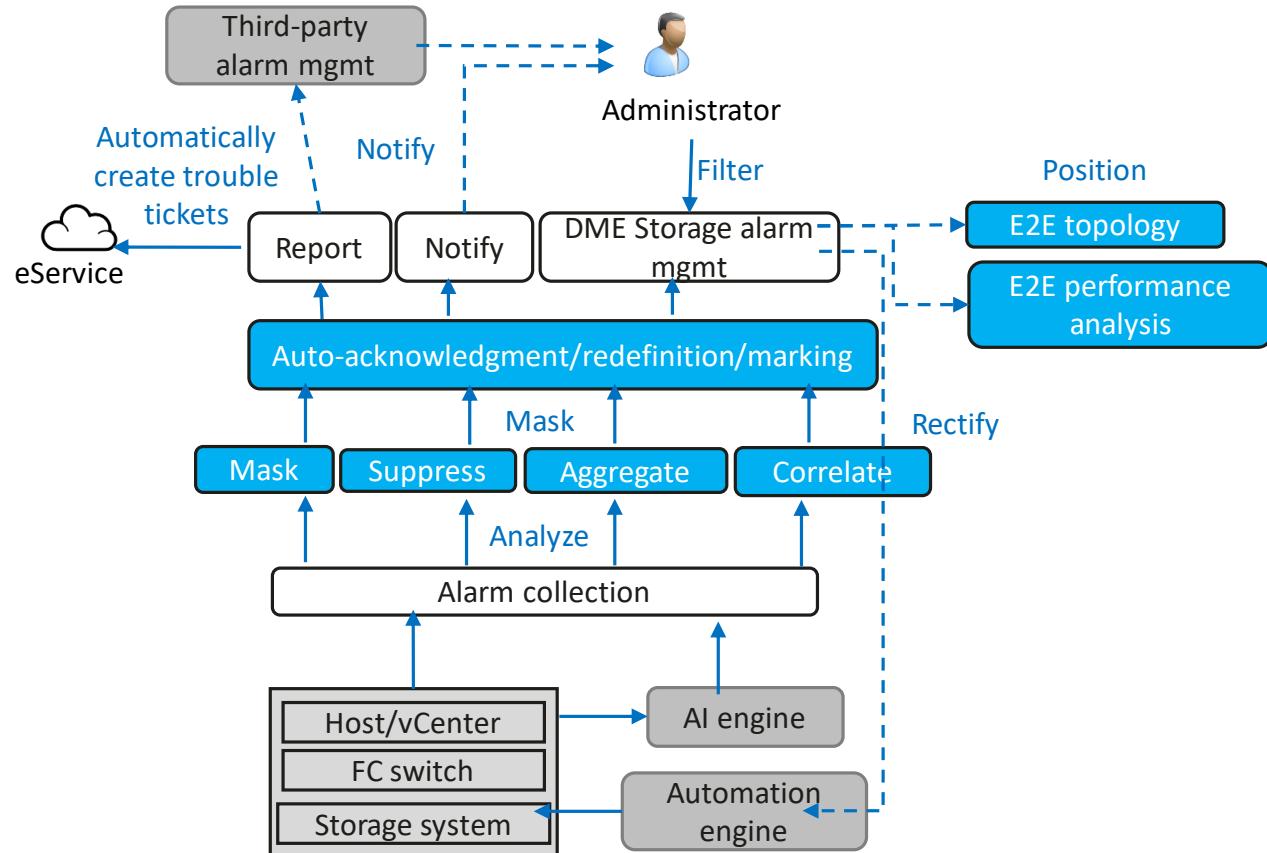
Capacity prediction

- Predicts the future capacity trends of pools, service levels, hosts, and host groups
- Predict the capacity trend once a day, read the data of the past six months, and generate the trend of the next three months.



Automatic Problem Analysis

Massive alarm notification --> Automatic association, aggregation, suppression, and masking
 Manual CLI-based fault locating --> One-click topology and performance analysis



Quick Fault Locating

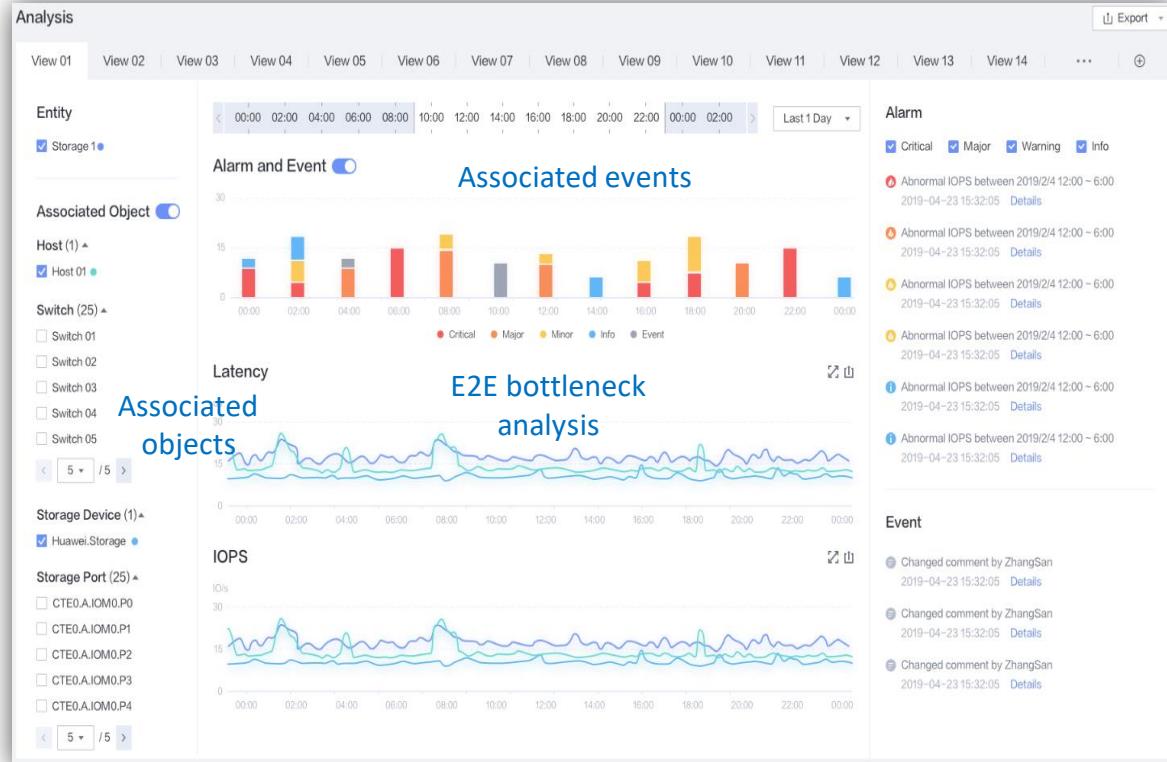
Planning

Deployment

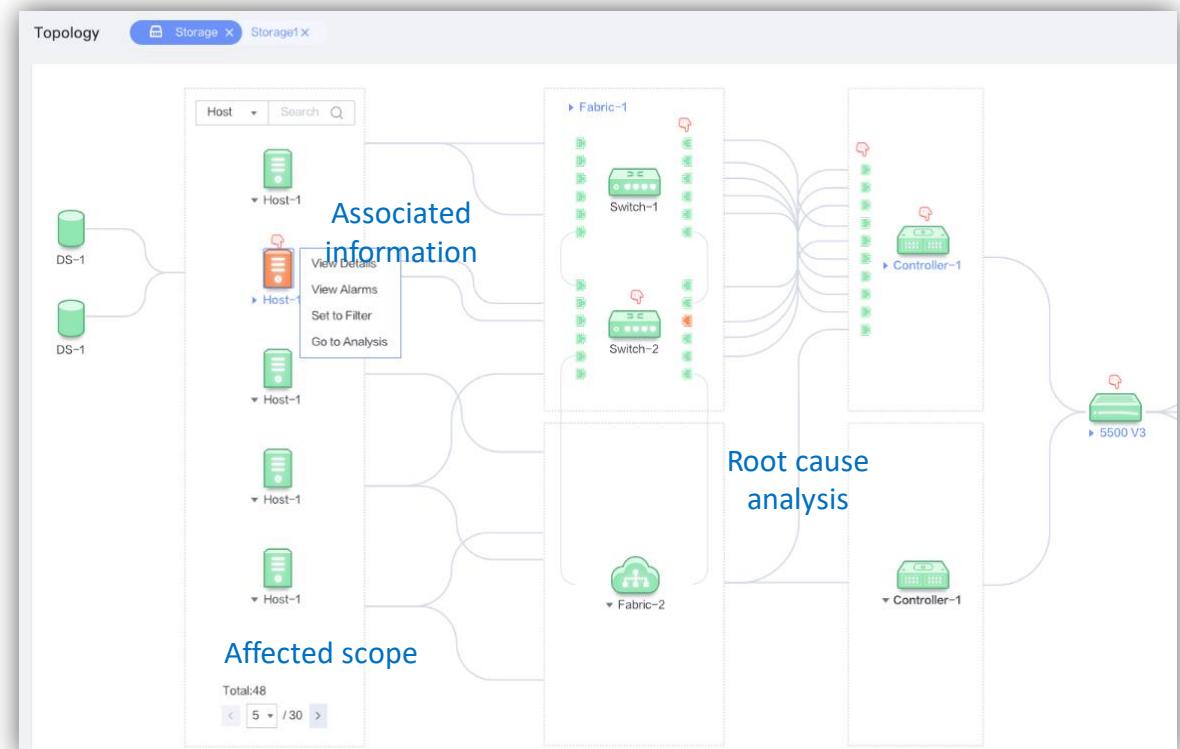
O&M

Optimization

E2E performance analysis



E2E topology



Quick Fault Locating: End-to-End Performance Analysis

Planning Deployment **O&M** Optimization

Create Chart

* Chart Name:

* Entity Type:

Total 100 | Selected 2

Name	Status	Capacity
Volume 001	Normal	Used 40.00 TB of 100.00TB
Volume 001	Normal	Used 40.00 TB of 100.00TB
Volume 001	Normal	Used 40.00 TB of 100.00TB
Volume 001	Normal	Used 40.00 TB of 100.00TB
Volume 001	Normal	Used 40.00 TB of 100.00TB
Volume 001	Normal	Used 40.00 TB of 100.00TB

Metrics: Latency IOPS Bandwidth

Associated Object: Host Switch Port Storage Device

Associated Object: Storage Port Storage Pool Disk

Analysis

View 01 View 02 View 03 View 04 View 05 View 06 View 07 View 08 View 09 View 10 View 11 View 12 View 13 View 14 ...

Last 1 Day

Entity

Storage 1

Associated Object

Host (1) Host 01

Switch (25) Switch 01 Switch 02 Switch 03 Switch 04 Switch 05

Storage Device (1) Huawei.Storage

Storage Port (25) CTE0.A.IOM0.P0 CTE0.A.IOM0.P1 CTE0.A.IOM0.P2 CTE0.A.IOM0.P3 CTE0.A.IOM0.P4

Alarm and Event

Associate d events

Latency

E2E bottleneck analysis

IOPS

Event

Critical Major Warning Info

Abnormal IOPS between 2019/24 12:00 ~ 6:00 2019-04-23 15:32:05 Details

Abnormal IOPS between 2019/24 12:00 ~ 6:00 2019-04-23 15:32:05 Details

Abnormal IOPS between 2019/24 12:00 ~ 6:00 2019-04-23 15:32:05 Details

Abnormal IOPS between 2019/24 12:00 ~ 6:00 2019-04-23 15:32:05 Details

Abnormal IOPS between 2019/24 12:00 ~ 6:00 2019-04-23 15:32:05 Details

Changed comment by ZhangSan 2019-04-23 15:32:05 Details

Changed comment by ZhangSan 2019-04-23 15:32:05 Details

Changed comment by ZhangSan 2019-04-23 15:32:05 Details

E2E Performance Analysis

1. Supports association analysis of the storage SAN.
2. Allows users to view different performance indicators of multiple objects at the same time.
3. Alarms of related resources can be viewed on the performance monitoring page.
4. Currently, VM performance monitoring is not supported.
5. The real-time performance monitoring period is 5 seconds.
6. The historical performance monitoring period is 1 minute.
7. Historical performance data can be stored for a maximum of two years.

Quick Fault Locating: E2E Topology

Details
Action ▾
Planning
Deployment
O&M
Optimization

Basic Information

Sequence:	1234568	ID:	0xF01080015
Level:	Major	Type:	Performance
Occurred Time:	2019-01-28 11:02:30	Trap Time:	2019-01-28 11:02:30
Name:	Latency threshold violation,Latency threshold violation		
Location:	ME=Huawei.Storage,LUN=volume1		
Matched Rules:	Auto-acknowledge rule1		
Details:	FC front-end port CTE0.A.IOM0.P1 has too many bit errors.The system performance may be affected.		

Violation

Check name: Latency threshold check Violated Conditions: read latency >= 5ms

Affected Scope

Volume Analyze Performance Show Topology Host DataStore VM

Redirecting to the topology

DME STORAGE displays the topology of the SAN network. In the topology, you can view the link relationship between LUNs and hosts and highlighted objects that have alarms. You can be directly redirected from the topology view to the performance analysis page.

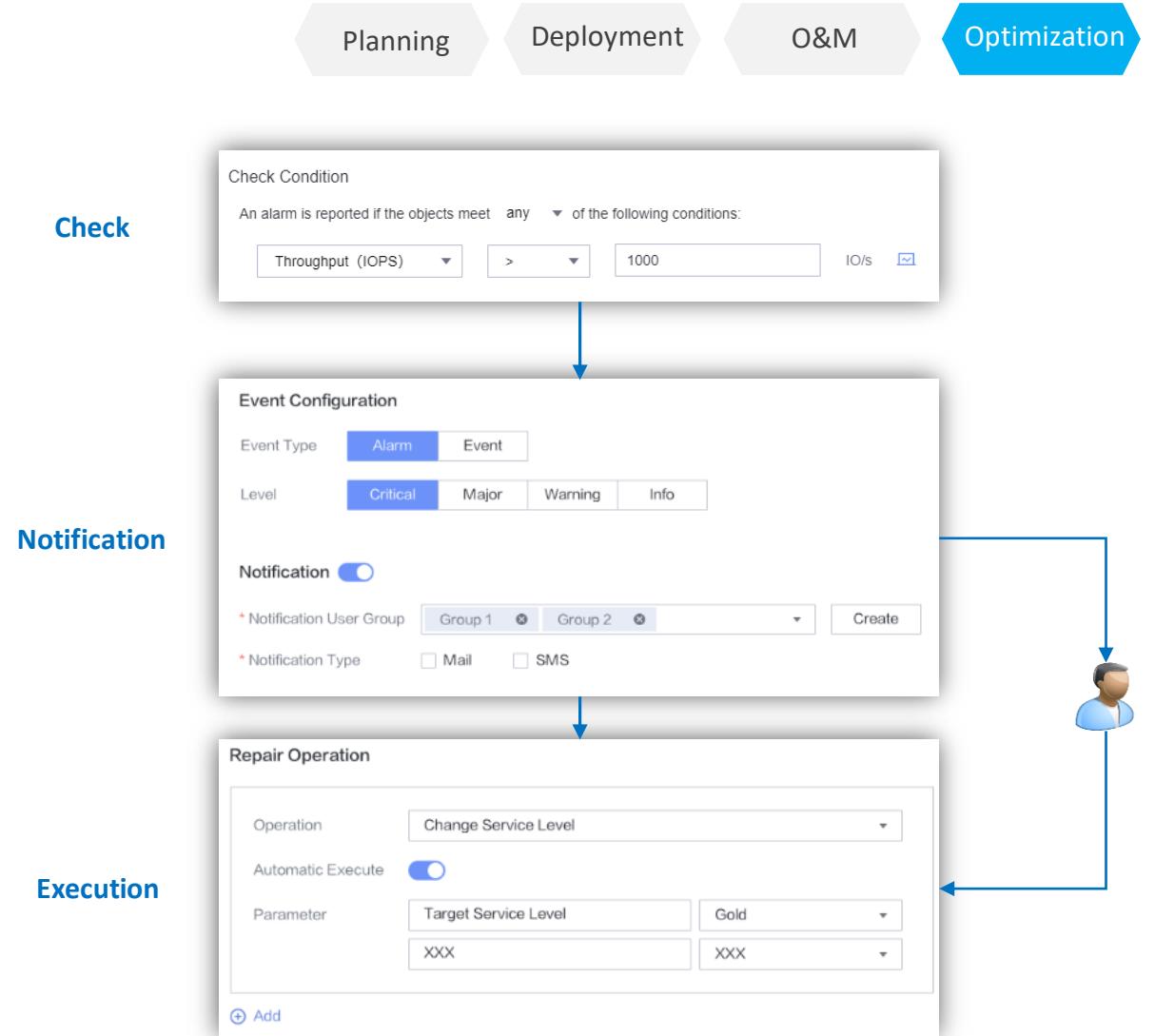
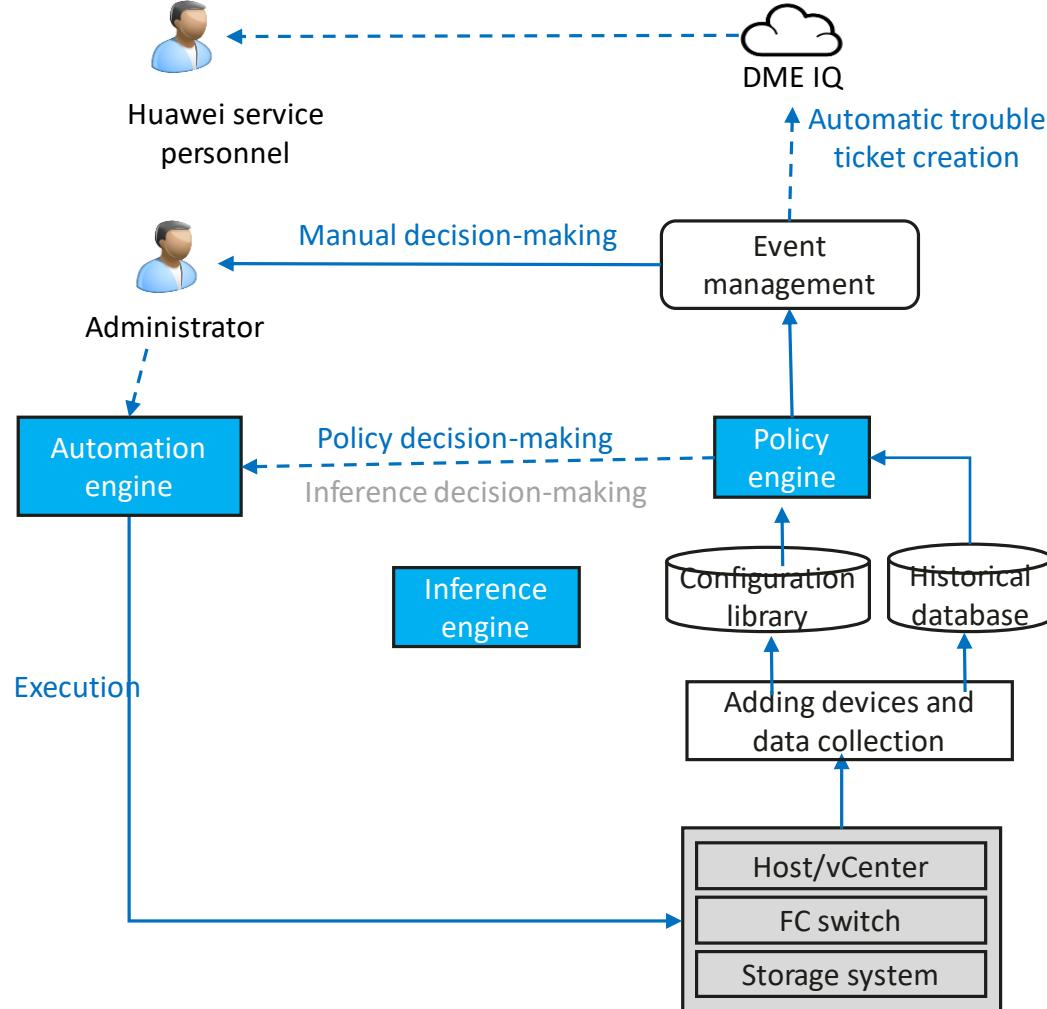
Topology

Associated information

E2E root cause analysis

Affected scope

Automatic Problem Resolution



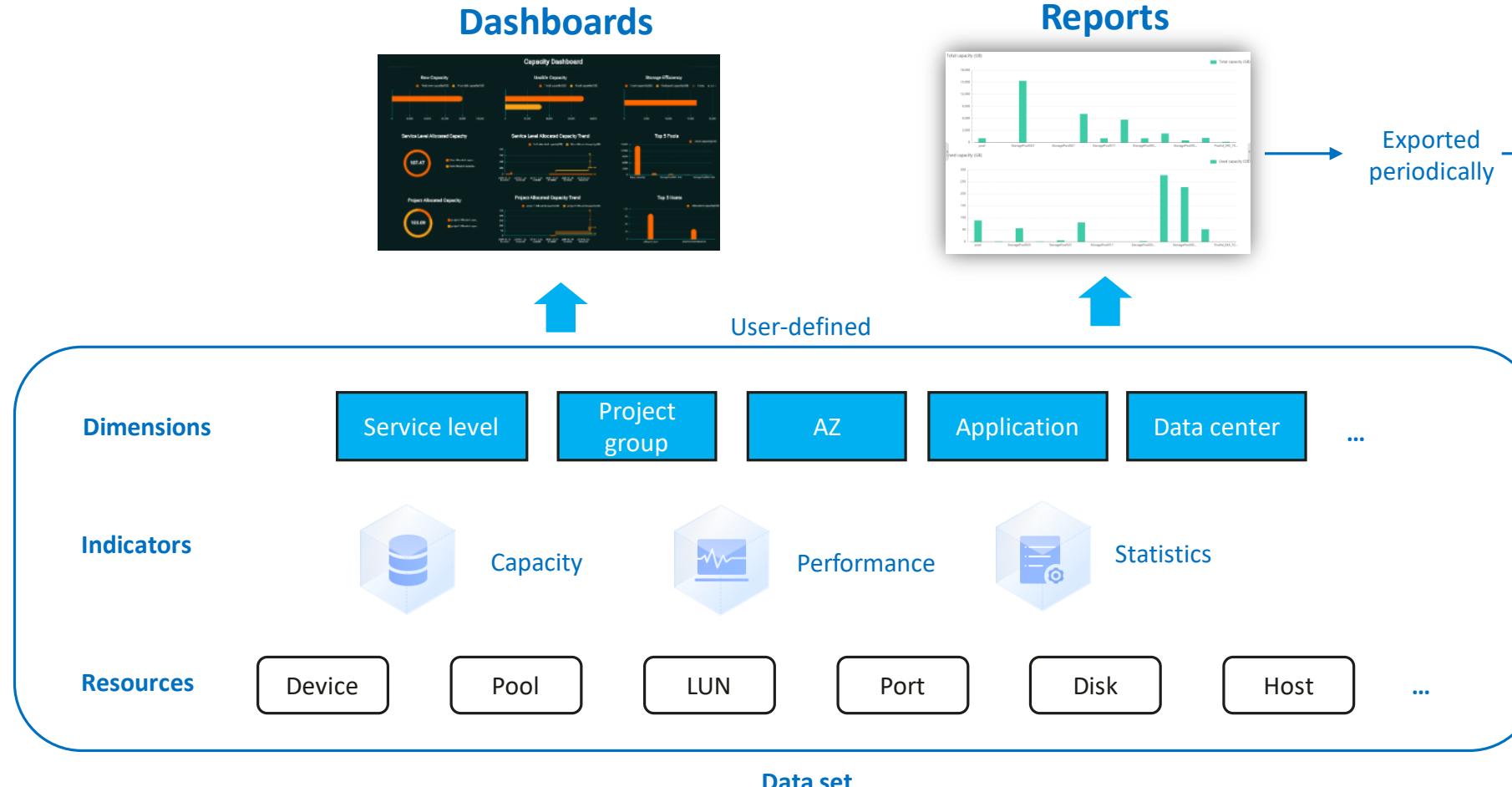
Rich Capabilities in Report Presetting and Customization

Planning

Deployment

O&M

Optimization



Competitiveness Comparison Between DME STORAGE and Competitors

Planning

Deployment

O&M

Optimization

DME Storage has the highest degree of convergence, and the convergence concept has led the industry planning. Currently, both Vendor N and Vendor H management software evolve to convergence

The DME storage modes of automatic provisioning are leading, and the basic O&M capability is the same as that of competitors

	Vendor E SRM	Vendor N AIQUM	Vendor H Ops Center	Huawei DME Storage 2.1.0
Provisioning	NO	★★★	★★★	★★★★
O&M	★★★	★★★★	★★★	★★★★
Copies	★★	★★★	★★★	★★★
Flow	NO	★★★	★★★	NO
Openness/ Ecosystem	★★★★	★★★★	★★	★★★

Quiz

1. (Multiple-choice) Which of the following capabilities are the northbound ecosystem capabilities of DME Storage?
 - A. Provide northbound RESTful interfaces to connect to platforms such as CMP
 - B. Provide SNMP interfaces to connect to the centralized alarm management platform
 - C. Provide vCenter and CSI plug-ins to integrate into the ecosystem
 - D. Provide SMI-S interfaces to connect to third-party NMSs.
2. (Single-choice) What is the biggest difference between DME IQ and DME Storage in terms of the key features they provide?
 - A. DME IQ can provide capacity prediction, but DME Storage cannot
 - B. DME IQ can provide performance prediction, but DME Storage does not
 - C. DME IQ supports E2E topology, but DME storage does not
 - D. DME IQ does not provide the storage configuration function, but DME Storage can provide storage manual configuration and SLA-based automatic provisioning

Contents

1. Brief Introduction to DME Storage
2. DME IQ and DME STORAGE Features
- 3. Customer Cases**

With DME, BDO Bank of the Philippines Optimizes its Storage System to Make Services more Agile and Efficient



30%
Resource Utilization Improvement

- ✓ Single device → Resource pool management

70%
Shortened Service Rollout Time

- ✓ Quickly allocate transaction, card, and analysis services based on SLAs

3X
Management Efficiency Improvement

- ✓ Large-screen display, 12-month capacity and 60-day performance forecast in advance to support procurement

Summary

- ✓ Service capabilities of Huawei storage O&M data management system (DMS) and corresponding three-layer management software
- ✓ DME IQ usually is used for remote lightweight management, providing business workload simulation and risk prediction
- ✓ DME Storage northbound ecosystem constructed using the REST interface, SNMP interface, and ecosystem plug-ins
- ✓ DME Storage's infrastructure management protocols, such as HTTP-based Rest, SSHv2, and SNMPv3, are security protocols recognized by the industry.
- ✓ Different automation capabilities of DME Storage and applicable service scenarios
- ✓ Automatic allocation process of DME Storage, involving storage resource allocation capability and FC switch allocation capability
- ✓ DME Storage's basic strategies in data protection capability
- ✓ Capability building of proactive O&M (alarm management) and supported remote notification functions
- ✓ Performance association analysis and topology visualization of DME Storage
- ✓ Report customization on DME Storage and corresponding periodic reports and report export capabilities

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 Data Storage POC Test Design



Foreword

- The POC(Proof of Concept) test is designed based on the specific service requirements of specific customers. The purpose of POC is to determine the service requirements of proper system or software product versions and solutions for customers. POC is widely used in projects such as admission projects, PK projects, and centralized procurement.
- This chapter describes precautions and recommend test cases of flash storage POC test.

Objectives

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

- Describe the functions and values of POC in a project
- Describe POC test solutions recommended for data storage
- Precautions for a POC test
- How to conduct a POC test

Contents

- 1. POC Overview**
2. Recommended POC Contents for Flash Storage
3. Precautions for flash POC Performance Tests
4. Precautions for flash POC Reliability Tests
5. OceanProtect POC Tests
6. Precautions for DME storage POC
7. Precautions for Scale out Storage POC
8. How to Conduct a Satisfactory POC

Overview and Objectives

- This section describes POC test classification and process.
- On completion of this section, you will be able to:
 - Understand the purpose/classification and process of POC test.
 - Know how to conduct a POC test.

Ice-breaking

- How to **change customers' perceptions** if they have preconceived notions?
- How to **change the perception of partners?**
- How to make breakthroughs in **technical-level customer relationships?**



Truth-seeking

- How to reply to the **customers' specific requirements?**
 - Specific I/O model
 - Specific applications
 - Specific application scenarios

**Do we have the required functions? Are the functions working well?
Is the user experience good?**

Benefits of POC

Driving business success:

Makes product technical capabilities visible.

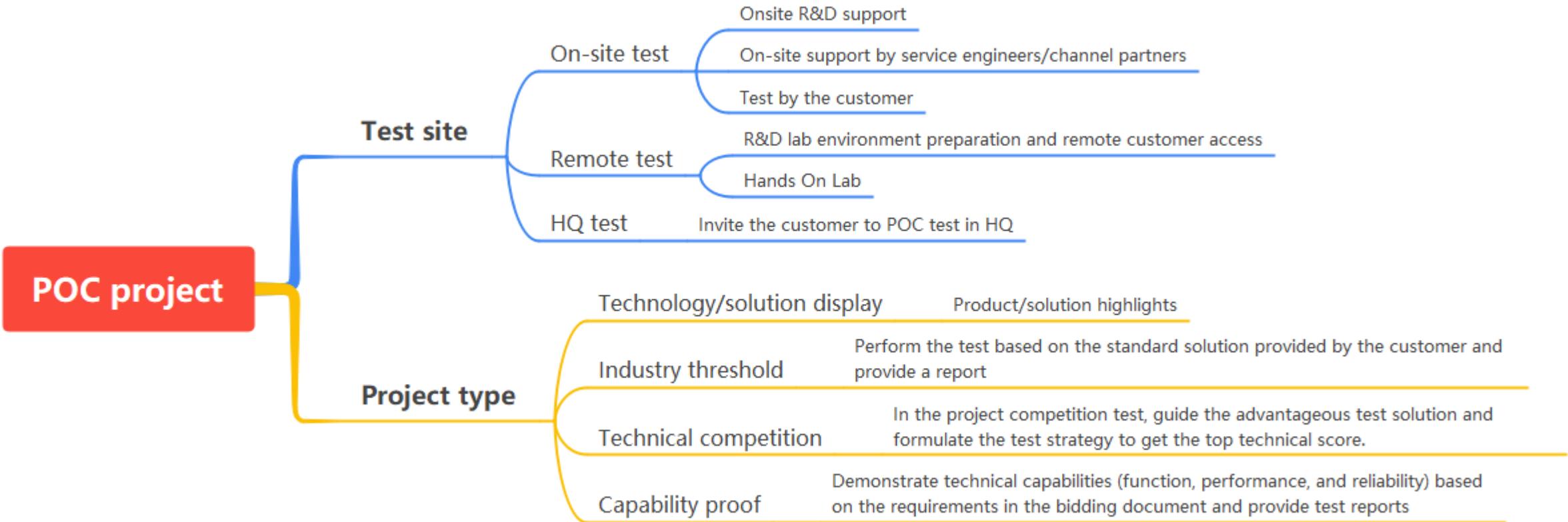
Resolves technical doubts and obtains customer recognition.

Demonstrates technical capabilities.

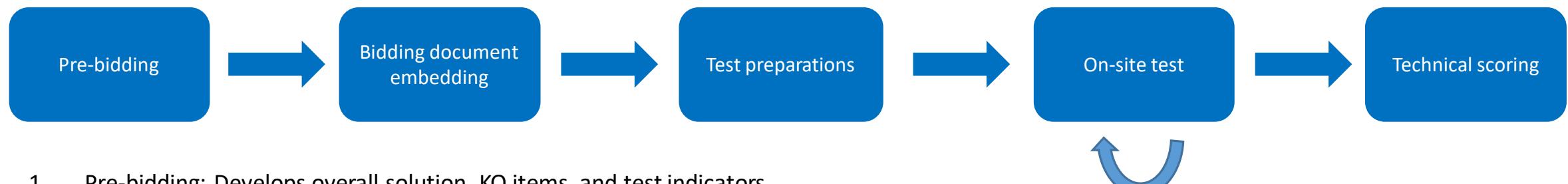
Providing reassurance:

Identifies project risks, verifies configuration compliance, and verifies solution feasibility.

POC Classification



General On-site POC Process



1. Pre-bidding: Develops overall solution, KO items, and test indicators.
2. Bidding document embedding: Scoring rules, functions and features, KO items, bonus items, and test methods.
3. Test preparation: Materials, personnel, site, test referee, and detailed test solution.
4. On-site test: Test sequence and test scheme execution.
5. Result application: Ensure that the test result is used for technical scoring.

Quiz

1. (True or False) Do all projects require POC testing?

2. (Multiple-choice) Which of the following types of POC tests are included?
 - A. On-site test
 - B. Remote test
 - C. HOL test
 - D. HQ test

Contents

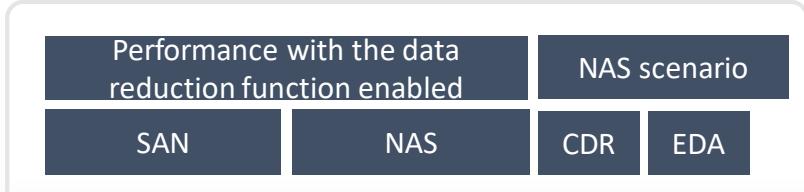
1. POC Overview
- 2. Recommended POC Contents for Flash Storage**
3. Precautions for flash POC Performance Tests
4. Precautions for flash POC Reliability Tests
5. OceanProtect POC Tests
6. Precautions for DME storage POC
7. Precautions for Scale out Storage POC
8. How to Conduct a Satisfactory POC

Overview and Objectives

- This section describes the recommend test case of flash storage POC
- On completion of this section, you will be able to:
 - Understand the recommended test solution and OceanStor Dorado & OceanStor

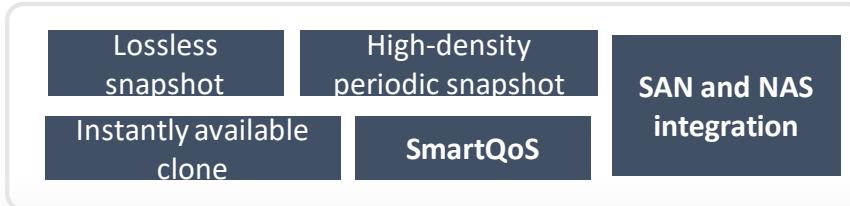
Recommended Content in flash storage POC - OceanStor Dorado

Ever fast



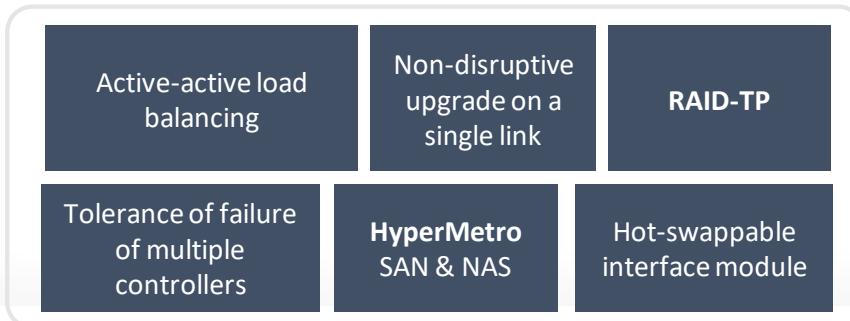
- Demonstrates the advantages of high IOPS and low latency.
- Suits I/O model to service scenarios.

AI-powered



- Snapshot/Clone application scenarios: Data extraction and query in the operation analysis scenario; Development and test.
- In data backup and service cutover scenarios, snapshots are required to ensure quick rollback. Therefore, the impact of snapshots on the production system is critical.
- High-density snapshots with shorter intervals greatly reduce data loss.
- SAN and NAS share the same resource pool, and no independent resource pool needs to be planned, improving resource utilization.

Always online



- High-end storage **tolerates failures of a single engine, seven out of eight controllers, three out of four controllers on a single link, and both controllers on a single link**.
- Active-active LUNs can dynamically balance workloads among multiple controllers, eliminating the need for complex planning during initial deployment and manual adjustment during subsequent maintenance.
- The upgrade is not interrupted, and the host is unaware of the upgrade. The upgrade does not require attendance of maintenance personnel.
- The controller failover time is short, ensuring stable running of upper-layer applications of the core database in high concurrency scenarios. Otherwise, I/Os may be overstocked, affecting service continuity.

Recommended Content in flash storage POC - OceanStor

Comprehensive Functions, Full Specifications, and Convergence and Interworking

Excellent performance

SAN	Max. performance of multiple LUNs when the data reduction function is disabled
NAS	Large files and large I/Os Small files and small I/Os

SmartQos/HyperLock/
SmartVirtualization/
SmartMigration/

Deep convergence

Active-active solution for SAN and NAS		FC/iSCSI/NFS/SMB/NoF
Instantly available clone	Secure Snapshot	QoS

HyperMetro
HyperReplication/HyperCDP
HyperSnap/HyperClone /3DC

High reliability

Tolerance of failure of three out of four controllers.	Fast reconstruction	Hot-swappable interface module
--	---------------------	--------------------------------

CloudBackup/
Unified-Namespace/
SmartMulti-Tenant/
SmartQuota/
Anti-virus



Quiz

1. (True or False) The NAS performance of Dorado is comparable to NetApp and at the same level product.
2. (Multiple-choice) Which of the following items are recommended for testing by OceanStor Dorado?
 - A. Tolerance of failure of multiple controllers
 - B. SAN and NAS integration
 - C. SmartQoS
 - D. RAID-TP

Contents

1. POC Overview
2. Recommended POC Contents for Flash Storage
- 3. Precautions for flash POC Performance Tests**
4. Precautions for flash POC Reliability Tests
5. OceanProtect POC Tests
6. Precautions for DME storage POC
7. Precautions for Scale out Storage POC
8. How to Conduct a Satisfactory POC

Overview and Objectives

- This section describes the recommend test case of flash storage POC
- On completion of this section, you will be able to:
 - Understand the recommended test contents of flash storage in performance.
 - Know the networking configuration of flash storage.



- Do you know the I/O model and performance expectations for the customer's production IT systems? The goal is clear for the POC test?
- Will you guide the I/O model for performance testing?
- Is large I/O performance worse than traditional storage for all-flash storage?

Recommended Solutions for flash storage SAN Performance Tests

Testing the mixed read/write performance of 8 KB random data on multiple LUNs and a single LUN

- **8 KB is the main I/O size of database OLTP services**

In OLTP services, the size of I/Os delivered by the data LUN of the service system is the same as the block size of the database. Generally, the size is 4 KB or 8 KB, and the ratio of random read/write is 6:4 or 7:3.

- **Fill the capacity for one time**

After the capacity is filled once, read I/Os are actually delivered to SSDs. Otherwise, read I/Os are directly returned after the metadata is queried. After the LUN capacity has been fully filled, the overall performance of the storage system can be accurately tested.

- **Capability of a single LUN**

LUNs do not belong to any controller, greatly simplifying management and O&M. You do not need to plan a large number of LUNs and do not need to worry about network planning.

The capability of a single LUN is an important means to prove the active-active capability. If the entire LUN belongs to a controller, a single LUN cannot fully utilize the overall system performance. It is good practice to use the single-LUN capability to beat vendors who have owning controllers, especially for high-end storage. This is also an advantageous item in mid-range scenarios.

In the single-LUN test, multiple servers are required due to the bottleneck of a single host. Therefore, pay attention to this point during configuration.

- **Value-added features**

SmartDedupe and SmartCompression

HyperSnap, HyperReplication, and HyperMetro

Recommended Solutions for flash storage NAS Performance Tests

- **Reserve sufficient IP resources**

Limited by the number of TCP connections, an IP mount point of a host can provide only about 20,000 to 50,000 OPS. A host can provide a maximum of 100,000 to 150,000 OPS.

You need to reserve a port for testing the ultimate OPS performance. You can also estimate the performance that can be reached based on the number of ports.

- **Performance Test Model**

Generally, NAS is used in specific scenarios based on customer requirements.

If there is no specific requirement, the small I/O model for small files and the large I/O model for large files are recommended.

Small files and small I/Os are used for OA and carrier CDR services.

Large files and large I/Os are for HPC scenarios.

- **Storage configurations**

In NFS scenarios, one or more file systems and **multiple sub directories** are recommended;

In CIFS scenarios, multiple file systems and multiple sub directories are recommended.



- List the factors that you think will affect storage performance testing.

Number of controllers

Number of Disks

RAID level

Network

Test Server

...

Configuration/Networking Suggestions

Performance is a systematic project and needs to be viewed from top to bottom and from end to end.

- **Storage disk configuration suggestions**

For mid-range storage, it is recommended that at least 25 SSDs be configured to maximize the controller capability. Disk enclosures can be configured based on site requirements.

For high-end storage, you are advised to configure at least 50 SSDs to maximize the controller's capability.

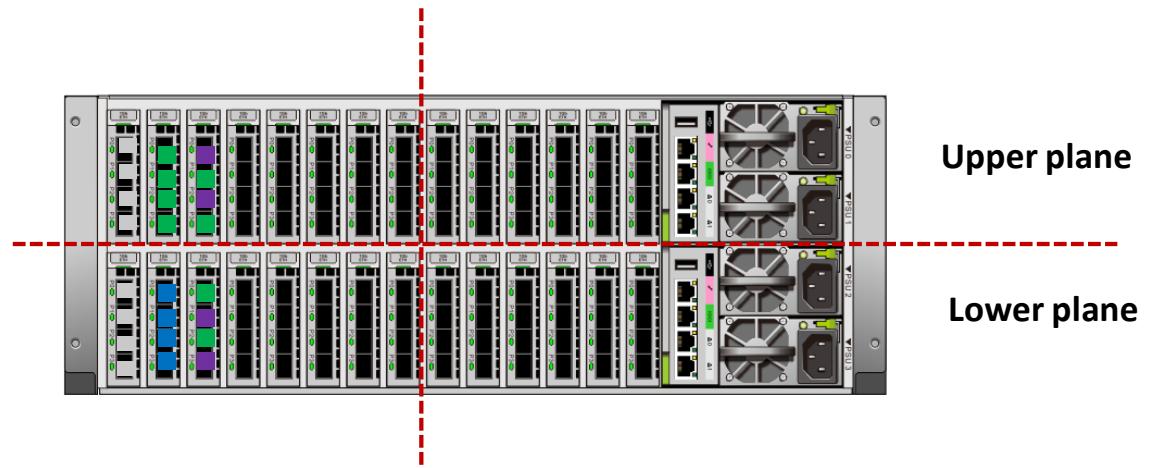
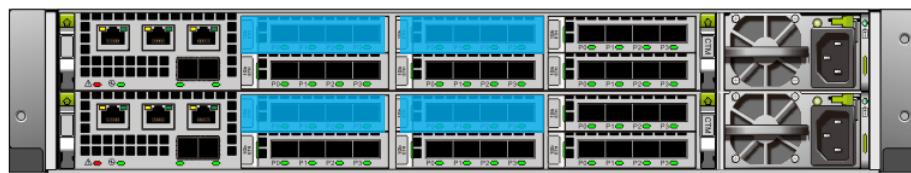
- **Suggestions on configuring front-end storage ports**

Ensures that links and interface modules are redundant and that the performance of key services can be ensured in fault scenarios.

For mid-range storage, it is recommended that two interface modules be configured for each controller to meet redundancy requirements and fully utilize CPU resources of controllers.

For high-end storage, ensure redundancy in upper and lower planes.

Ultrapath is recommended for easy and efficient configuration.



Configuration/Networking Suggestions

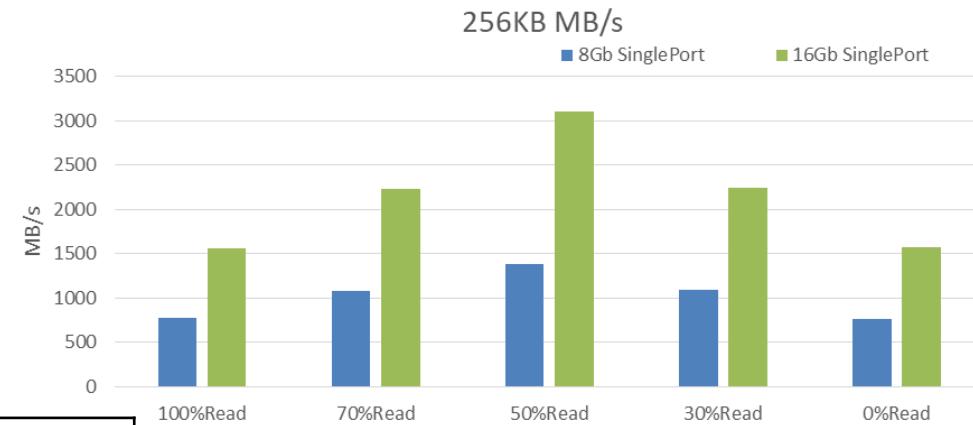
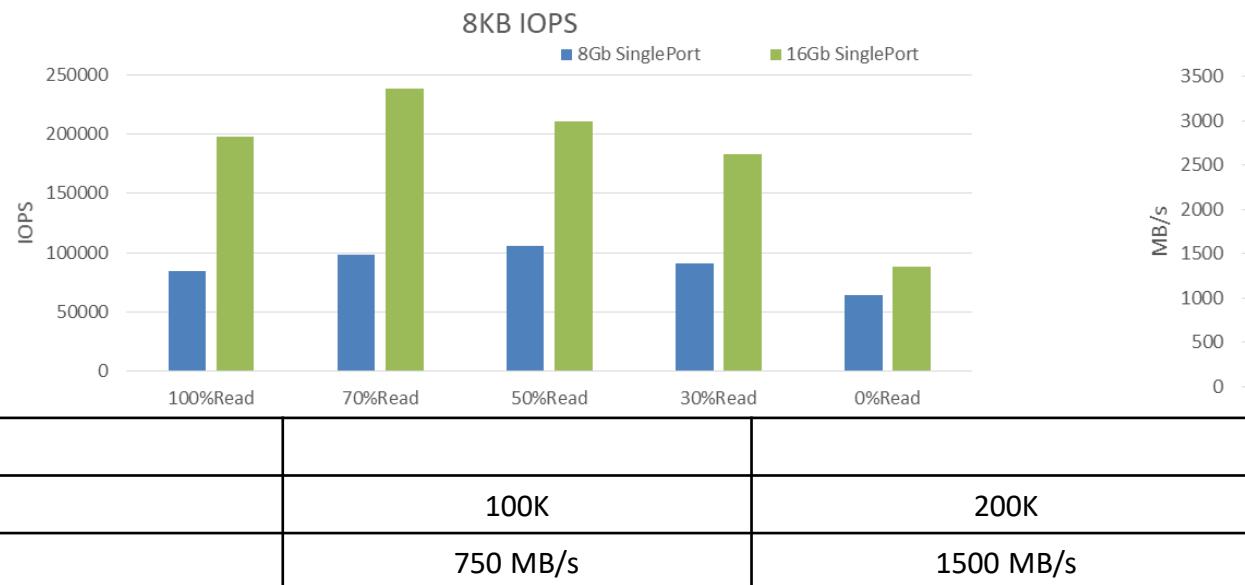
Performance is a systematic project and needs to be viewed from top to bottom and from end to end

- **Server model, CPU, and I/O channel rate and quantity**

Compared with mid-range computers such as AIX and Solaris, 2-socket x86 servers are easier to test for higher performance using I/O tools.

I/O-intensive services consume a large amount of CPU resources. The newer the CPU generation and the higher the frequency, the higher the performance.

- **I/O channel rate and quantity of servers**

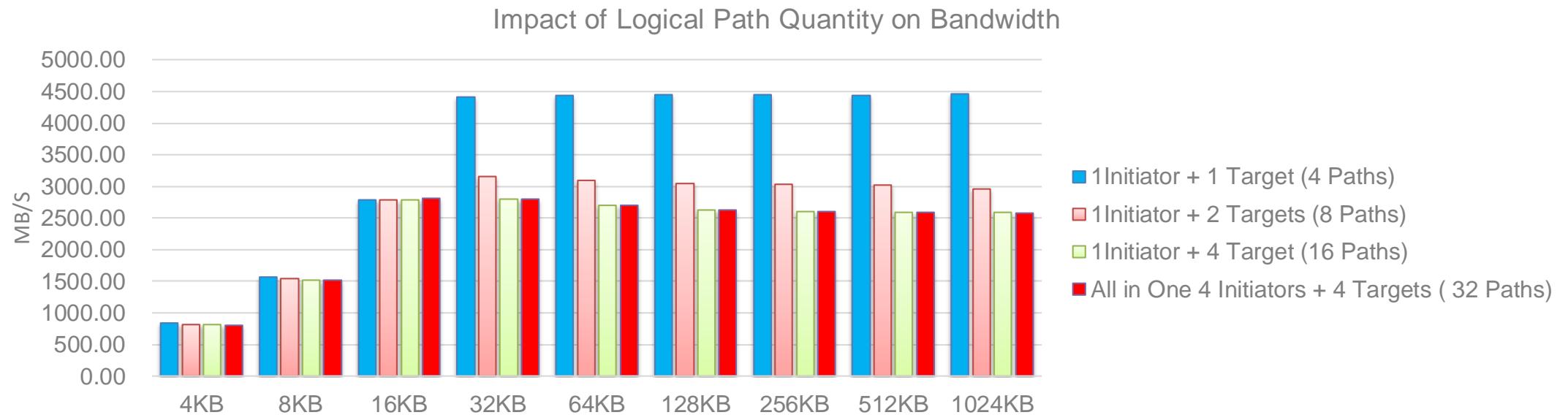


Configure sufficient I/O channel resources (host HBAs and front-end interface modules of storage devices) based on the actual reachability of products. The following is used as a reference.

Impact of Multipathing on Performance

Recommended zone configuration for FC switches:

Each zone contains only one initiator and one target. Ports are not shared. The number of logical paths is the same as that of initiators.



Increasing the logical paths may have 30%+ of impact on the performance of large I/Os. Performance of small I/Os is not affected.

Verification environment: Four FC ports on the host are connected to ports 0, 1, 2, and 3 on the switch, and four FC ports on the storage device are connected to ports 4, 5, 6, and 7 on the switch.

- 4 paths: One-to-one zone planning, no port sharing. (0,4); (1,5); (2,6); (3,7).
- 8 paths: One-to-two zone planning. Storage ports are shared twice. (0,4,5); (1,6,7); (2,4,5); (3,6,7).
- 16 paths: One-to-four zone planning. Storage ports are shared four times. (0,4,5,6,7); (1,4,5,6,7); (2,4,5,6,7); (3,4,5,6,7).
- 32 paths: A large zone or no zone is created. (0,1,2,3,4,5,6,7).

Quiz

1. (True or False) Huawei UltraPath is recommended for projects that require high performance. UltraPath has higher transmission efficiency than the native multipath.
2. (Multiple-choice) Which of the following statement about flash storage NAS performance tests are correct?
 - A. Small files and small I/Os are used for OA and carrier CDR services
 - B. Large files and large I/Os are for HPC scenarios
 - C. In CIFS scenarios, one or more file systems and multiple sub directories are recommended
 - D. In CIFS scenarios, multiple file systems and multiple sub directories are recommended.

Contents

1. POC Overview
2. Recommended POC Contents for Flash Storage
3. Precautions for flash POC Performance Tests
- 4. Precautions for flash POC Reliability Tests**
5. OceanProtect POC Tests
6. Precautions for DME storage POC
7. Precautions for Scale out Storage POC
8. How to Conduct a Satisfactory POC

Overview and Objectives

- This section describes the recommend test case of flash storage POC
- On completion of this section, you will be able to:
 - Understand the recommended test contents of flash storage in reliability.
 - What are the common test tools in the flash storage POC test?

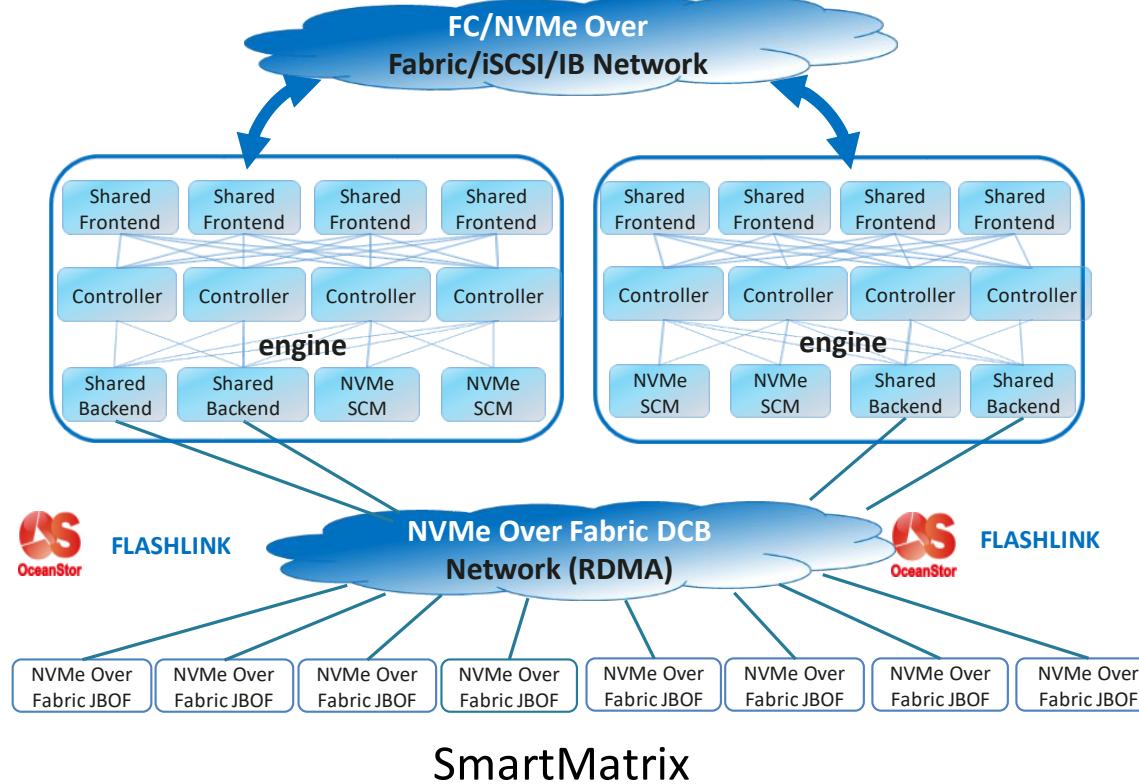


- Do you know the symmetric Active-Active architecture ?
- Do you know which storage model are using the symmetric AA architecture?
- Do you know what the advantage of symmetric AA is? And how to verify symmetric AA ?
- If the LUN does not show you the own controller ship, and multiple controllers just can read and write a single LUN. This can be considered as an active-active architecture ?

High Reliability — Active-Active Architecture

What Is Active-Active Controller	No LUN ownership Multiple controllers can read and write a single LUN. (Non-forwarding)	System load balancing For a single LUN, the system can still implement load balancing. (CPU, not only front-end IOPS)	Simplified O&M You do not need to manually plan the controller ownership and controller usage.
Customer Concerns	<ol style="list-style-type: none">1. How long is the controller failover time? (Theoretically, the controller switchover time is short.)2. What are the impacts of I/O switchover on services during the system software upgrade?3. How to achieve system utilization load balancing? Can I balance the utilization of a single LUN?4. Can active-active architecture improve reliability?		

Advantages of OceanStor Dorado



FIM

- The front-end shared I/O module intelligently identifies host I/Os and distributes them based on specific rules. In this way, host I/Os are directly sent to the optimal controller without being forwarded by the controller, implementing full interconnection between the front-end module and controllers in the engine.

Symmetric interconnection of four controllers

- Controllers are interconnected through 100 Gbit/s RDMA high-speed channels on the passive backplane in the chassis. The cache mirror data is mirrored to other controllers without being forwarded, implementing full interconnection between the four controllers in an engine.
- Cross-engine three-copy + Persistent cache, supporting data sharing and redundancy protection.

Cross-engine disk enclosure sharing

- Back-end interconnect I/O modules (BIMs) are also inserted into the chassis. These modules can be simultaneously accessed by controllers. A disk enclosure can be accessed by controllers in the engine at the same time, implementing full interconnection between the disk enclosure and controllers in the engine;
- Smart disk enclosures can be connected to two engines at the same time. In this way, the same disk enclosure can be accessed across engines at the same time, and back-end disk enclosures are fully interconnected across engines.

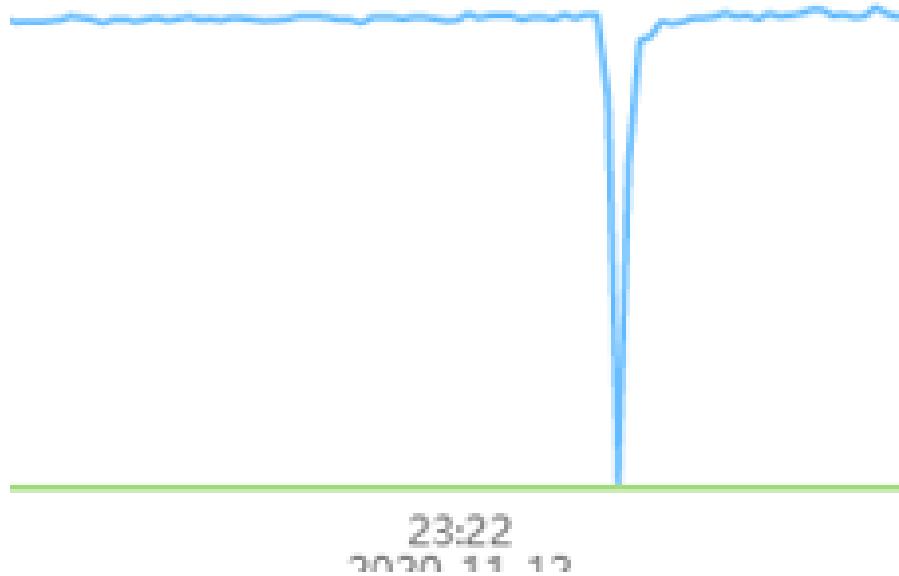
NDU

- The upgrade is component-based. Services are not affected during the upgrade, links are not interrupted, and performance does not deteriorate.
- Short upgrade duration: I/O components are upgraded within 1 second.

Controller Failover Time

What is I/O suspension? Is the service interrupted when the I/O returns to zero?

Under the same test conditions, using the same controller failure simulation method, observation dimension, and I/O model, you are advised to observe test results on the host side.



Observation on the storage system

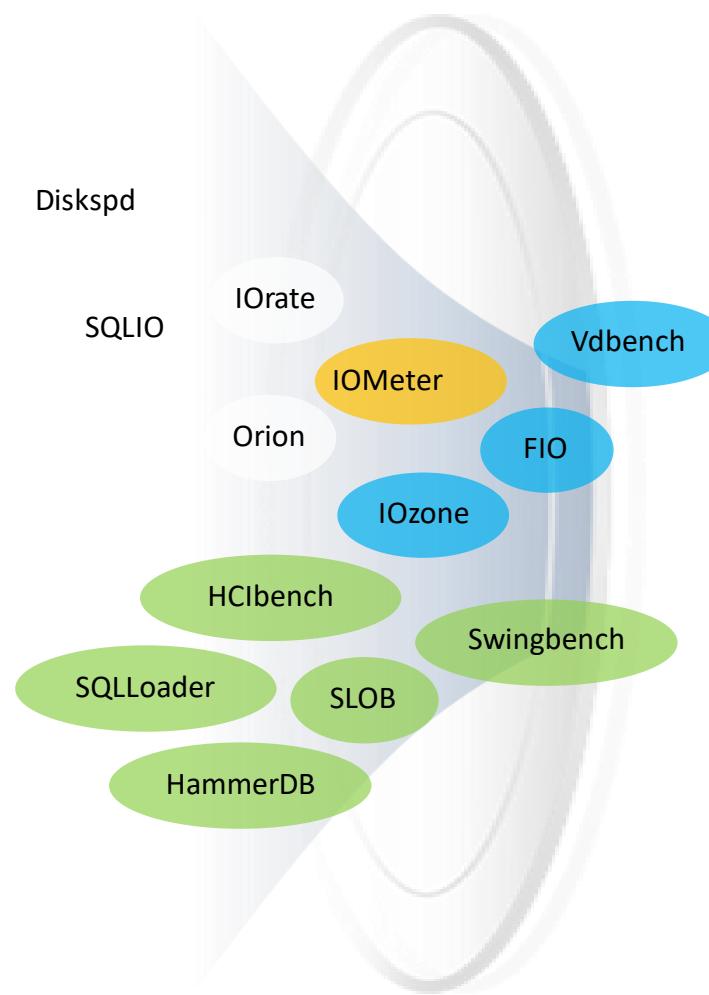
Time	interval	i/o	MB/sec	bytes	read	resp	read	write	resp	resp	queue	cpu%	cpu%		
		rate	1024**2	1/o	pct	time	resp	resp	resp	max	stddev	depth	sysu	sys	
1920	20:27:57.041	2620	55381.00	432.66	8192	70.10	1.495	1.699	1.016	207.823	6.354	255.9	2.0	1.5	
1921	20:27:58.041	2621	0.00	0.00	0	0.00	0.000	0.000	0.000	0.000	0.000	0.000	255.8	0.1	0.0
1922	20:27:59.041	2622	0.00	0.00	0	0.00	0.000	0.000	0.000	0.000	0.000	0.000	256.1	0.1	0.0
1923	20:28:00.041	2623	37.00	0.29	8192	72.97	1659.724	2274.378	0.159	3419.423	1727.805	256.1	0.0	0.0	
1924	20:28:01.041	2624	5461.00	42.66	8192	70.66	150.363	212.624	0.383	4735.543	775.767	255.8	0.2	0.1	
1925	20:28:02.041	2625	77235.00	603.40	8192	69.93	2.499	3.352	0.517	953.524	9.074	255.8	2.8	2.1	
1926	20:28:03.042	2626	139520.00	1090.00	8192	69.97	1.395	1.821	0.403	96.559	4.672	255.6	5.2	3.7	
1927	20:28:04.042	2627	213278.00	1666.23	8192	69.85	0.896	1.161	0.282	70.323	2.824	255.0	7.0	5.5	
1928	20:28:05.045	2628	237849.00	1858.20	8192	70.05	0.797	0.963	0.409	231.784	3.861	255.3	7.7	6.2	
1929	20:28:06.042	2629	160109.00	1250.85	8192	70.44	1.105	0.884	1.632	235.469	9.495	255.3	5.3	4.2	
1930	20:28:07.043	2630	154211.00	1204.77	8192	69.80	1.345	1.031	2.069	231.125	10.873	255.6	4.9	3.8	
1931	20:28:08.042	2631	126577.00	988.88	8192	70.13	1.433	1.189	2.005	228.654	10.706	255.4	4.5	3.6	
1932	20:28:09.041	2632	123642.00	965.95	8192	70.12	1.595	1.196	2.532	235.858	12.181	255.7	4.3	3.4	
1933	20:28:10.042	2633	109851.00	858.21	8192	69.90	1.760	1.291	2.849	239.200	13.056	255.7	3.8	2.9	
1934	20:28:11.042	2634	120770.00	945.52	8192	69.88	1.597	1.681	1.401	241.820	9.439	255.5	4.3	3.4	
1935	20:28:12.042	2635	111831.00	873.68	8192	70.05	1.675	1.682	1.660	236.258	10.319	255.7	3.7	2.9	
1936	20:28:13.042	2636	111227.00	868.96	8192	70.29	1.675	1.612	1.825	249.238	10.634	255.5	4.0	3.0	
1937	20:28:14.042	2637	113659.00	887.96	8192	69.74	1.771	1.780	1.752	247.237	10.600	255.7	3.9	3.0	
1938	20:28:15.042	2638	125819.00	982.96	8192	70.19	1.493	1.439	1.619	315.900	9.926	255.7	4.4	3.4	
1939	20:28:16.042	2639	103349.00	807.41	8192	69.70	1.893	1.508	2.778	238.501	13.406	255.5	3.6	2.8	
1940	20:28:17.045	2640	103618.00	809.52	8192	69.79	1.862	1.786	2.038	244.301	11.620	255.7	3.8	2.7	
1941															
1942	Feb 12, 2019	interval	i/o	MB/sec	bytes	read	resp	read	write	resp	resp	queue	cpu%	cpu%	
1943			rate	1024**2	1/o	pct	time	resp	resp	resp	max	stddev	depth	sysu	sys
1944	20:28:18.042	2641	103893.00	807.76	8192	70.16	1.783	1.689	2.006	247.284	11.214	255.5	3.6	2.7	
1945	20:28:19.041	2642	138865.00	1084.88	8192	69.84	1.419	1.465	1.312	246.209	8.933	255.6	4.8	3.7	
1946	20:28:20.041	2643	123688.00	966.31	8192	70.02	1.525	1.386	1.850	264.123	10.592	255.7	4.3	3.4	
1947	20:28:21.042	2644	107132.00	836.97	8192	69.91	1.791	1.308	2.913	330.193	13.299	255.5	3.8	2.9	
1948	20:28:22.042	2645	111702.00	872.67	8192	70.04	1.746	1.377	2.611	248.369	12.661	255.8	3.7	2.8	
1949	20:28:23.042	2646	140747.00	1099.59	8192	69.87	1.351	1.238	1.611	236.665	9.802	255.4	4.8	3.8	
1950	20:28:24.041	2647	103695.00	810.12	8192	70.05	1.818	1.188	3.290	248.879	14.158	255.6	4.1	2.9	
1951	20:28:25.041	2648	103668.00	809.91	8192	69.93	1.853	1.650	2.324	249.585	12.165	255.5	3.8	3.0	
1952	20:28:26.042	2649	116405.00	909.41	8192	70.17	1.603	1.544	1.741	242.847	10.360	255.7	4.0	3.2	
1953	20:28:27.041	2650	108281.00	845.95	8192	70.06	1.153	1.250	0.927	234.883	7.344	255.4	4.0	3.3	
1954	20:28:28.041	2651	6204.00	48.47	8192	70.65	0.637	0.811	0.219	74.258	2.930	256.0	0.3	0.2	
1955	20:28:29.041	2652	0.00	0.00	0	0.00	0.000	0.000	0.000	0.000	0.000	0.000	256.1	0.0	0.0
1956	20:28:30.041	2653	0.00	0.00	0	0.00	0.000	0.000	0.000	0.000	0.000	0.000	255.8	0.0	0.0
1957	20:28:31.041	2654	0.00	0.00	0	0.00	0.000	0.000	0.000	0.000	0.000	0.000	256.1	0.1	0.0
1958	20:28:32.042	2655	0.00	0.00	0	0.00	0.000	0.000	0.000	0.000	0.000	0.000	265.8	0.1	0.0
1959	20:28:33.042	2656	478.00	3.73	8192	65.06	1000.434	995.395	1009.819	36838.494	4926.747	246.1	0.4	0.1	
1960	20:28:34.041	2657	81637.00	637.79	8192	69.70	18.189	20.520	12.828	37580.871	676.146	255.9	3.9	2.0	
1961	20:28:35.042	2658	139450.00	1089.45	8192	70.05	11.867	14.155	6.515	38719.239	531.611	255.6	4.8	3.7	

Observation on the host



- What I/O tools have you used to verify storage performance/failure behavior?
- Do you know of any I/O tools that can generate duplicate and compressed data ?

Recommended Test Tools



Tool Type	Tool	Recommendation Index	Recommendation Reason
Common tools	Vdbench	★★★★★	SAN and NAS are universal and adapt to multiple platforms. The test model can be flexibly configured. The test result statistics are detailed and easy to extract and analyze.
	FIO	★★★	Mainly used for SAN, not applicable to NAS, and mainly applicable to the Linux platform. The test model configuration has no fixed syntax, and the test result statistics are difficult to extract and analyze. If the test is performed by the customer, obtain the test script from the customer in advance.
	IOzone	★★★	Mainly used to test the file system and applicable to the Linux platform. The test model is limited by the bandwidth model. The test scenarios are limited.
	IOmeter/IOrate/Diskspd/Orion/SQLIO	Not recommended	It is a simple tool. The delivered I/Os are basically all zeros or all duplicate data. Therefore, it is not applicable to performance tests with deduplication and compression enabled. Some tools have bottlenecks and do not use the comparison test.
Test tools	Swingbench	★★★★★	Mainstream Oracle performance test tool, which is used for OLTP load tests and general platforms. The test result is displayed on the GUI. This tool is not applicable to extreme IOPS performance test.
	SLOB	★★★★★	A simple tool to test Oracle performance, with easy data construction. A large amount of data can be quickly constructed to test the ultimate storage performance.
	HammerDB	★★★★★	Database performance test tool, which adapts to multiple types of databases and provides GUIs. The amount of constructed data is small, meeting the requirements of simple verification.
Others	There are other methods or tools for testing real service scenarios, such as Oracle RMAN, SQL Loader, and containers. Focus on customer requirements and perform tests based on site conditions.		

Quiz

1. (True or False) A LUN does not have an owning controller, and multiple controllers can read and write a single LUN. This can be considered as an active-active architecture.
2. (Multiple-choice) Which of the following tools are recommended for testing?
 - A. Vdbench
 - B. IOmeter
 - C. IOzone
 - D. HammerDB

Contents

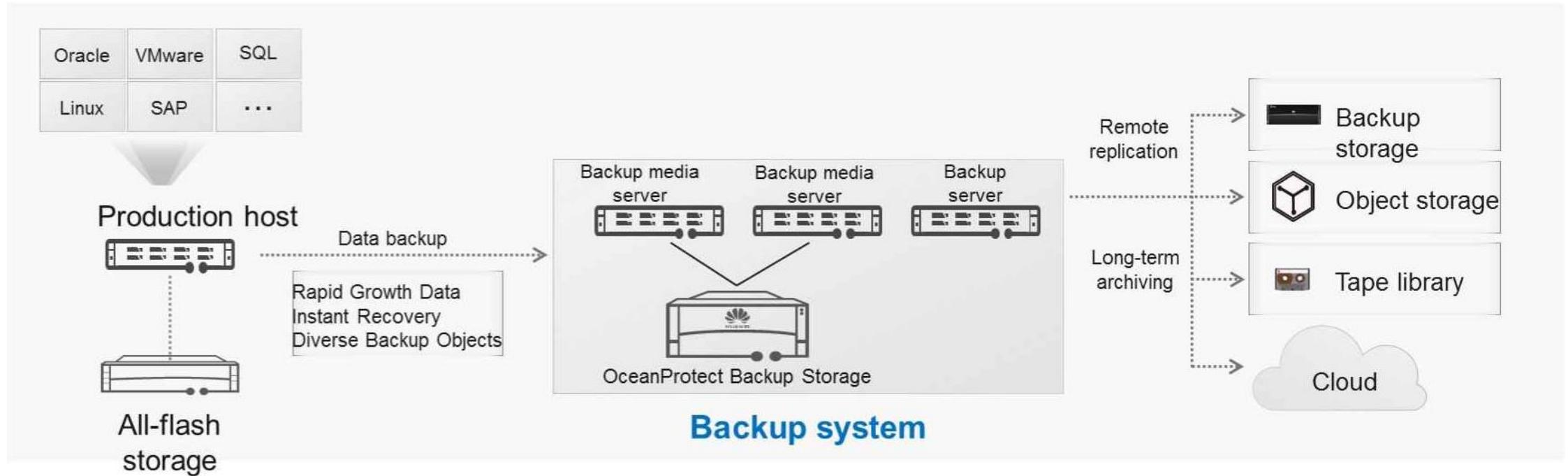
1. POC Overview
2. Recommended POC Contents for Flash Storage
3. Precautions for flash POC Performance Tests
4. Precautions for flash POC Reliability Tests
- 5. OceanProtect POC Tests**
6. Precautions for DME storage POC
7. Precautions for Scale out Storage POC
8. How to Conduct a Satisfactory POC

Overview and Objectives

- This section describes the recommend test case of OceanProtect.
- On completion of this section, you will be able to:
 - Understand OceanProtect back-up solution.
 - Describes the highlights and recommend test scenarios of OceanProtect.

OceanProtect POC Tests

OceanProtect Back-up Solution

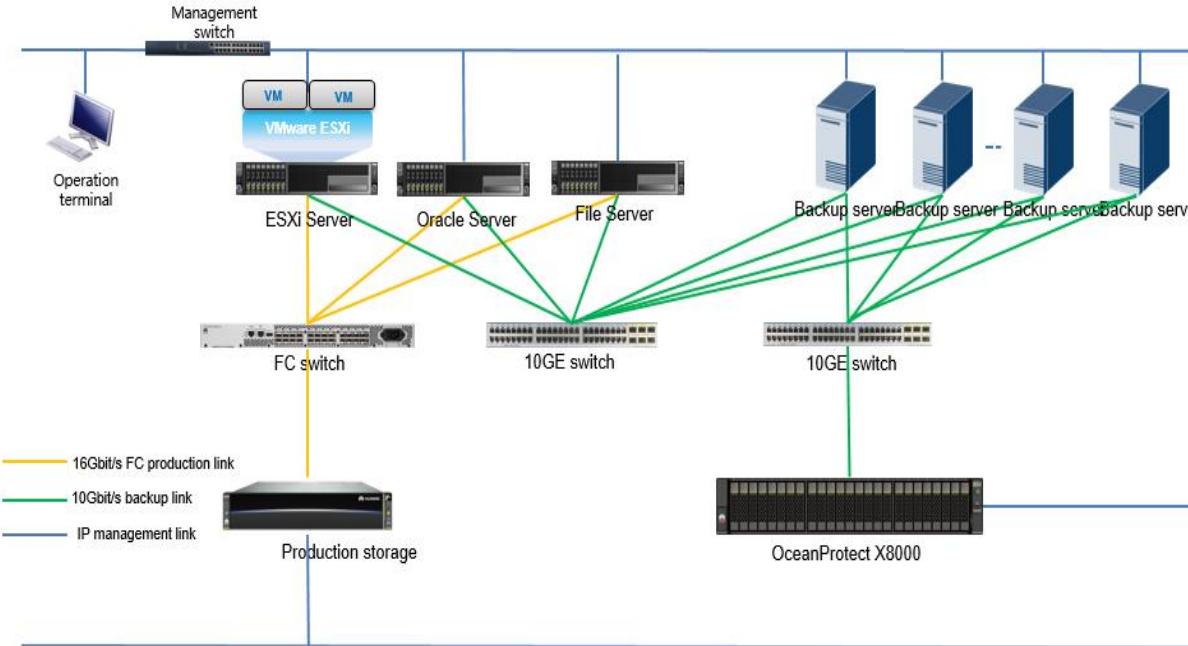


OceanProtect backup storage:

- Including OceanProtect X6000, OceanProtect X8000 & OceanProtect X9000 , Each product is available in all-flash form and HDD hybrid flash form;
- Compatible with the mainstream backup software (including Veritas NetBackup,CommVault,Veeam, etc.) , The NAS (NFS/SMB) protocol is recommended for interconnect with backup software.
- **Customers application compatibility, performance, reduction ratio, and reliability. The performance and reduction ratio depend on factors such as environment configuration, application configuration, data model, and backup policy. Before the evaluation, obtain detailed customer information.**

Recommended Solutions of OceanProtect Backup Storage

The OceanProtect dedicated backup storage works with backup software to back up applications in an end-to-end manner, and protect application data throughout the lifecycle. It can be connected to mainstream backup software.



Recommended Test :

- **High performance**

All-flash architecture, DTOE accelerates end-to-end backup, improves backup and recovery bandwidth of the backup storage system.

- **High data reduction ratio**

Innovative deduplication and compression technologies for ultimate data reduction.

- **High reliability**

Active-Active high-reliability architecture, supports RAID-TP, three-disk failure, and hot swap of front-end interface cards, ensuring end-to-end network and storage device hardware reliability, ensuring high reliability of backup services.

POC Test Precautions

Test type	Key impact factor	POC Test Precautions
Performance Test	<ol style="list-style-type: none"> 1. Hardware configuration and quantity (including application servers, backup servers, and backup storage) 2. Application type (file, virtualization, and database) 3. Application service pressure (data volume model) 4. Backup network 5. Backup software configuration (whether to enable deduplication, compression, and encryption) 6. Back up storage configurations (such as storage pools and file systems). 	<p>The OceanProtect backup storage NAS (NFS/SMB) protocol is recommended for performance tests.</p> <ol style="list-style-type: none"> 1. The performance test depends on the service pressure, backup network, and backup storage configuration. 2. In the OceanProtect maximum performance test, ensure that the front-end service pressure, backup server and network bandwidth are sufficient. 3. It is recommended that all-flash storage be used as the OceanProtect backup storage, and created multiple file systems to improve the overall system bandwidth. 4. It is recommended that deduplication and compression be disabled for the backup software during the performance test.
Reduction Ratio Test	<ol style="list-style-type: none"> 1. Application type (file, virtualization, and database) 2. Full application data model (data volume, data model, deduplication ratio, and compression ratio) 3. Incremental application data model (data volume, data model, deduplication ratio, and compression ratio) 4. Backup software configuration (whether to enable deduplication, compression, and encryption) 5. Backup policy and copy retention policy 6. Back up storage configurations (such as storage pools and file systems). 	<p>The OceanProtect backup storage NAS (NFS/SMB) protocol is recommended for the reduction ratio test.</p> <ol style="list-style-type: none"> 1. The reduction ratio test depends on many factors, different data models (database, virtual machine, VDI, file (regular/irregular) and backup policies. The reduction ratio varies with the backup policy. 2. It is recommended that VMware applications be backed up for the reduction ratio test. The full backup policy is used for the test. In addition, it is recommended that the backup software deduplication and compression be disabled. 3. It is recommended that data of the same application be backed up and written to the same file system.
Reliability Test	<ol style="list-style-type: none"> 1. Hardware configuration 2. Networking 	<p>You are advised to perform the following operations when backup services are running to observe the running status of backup services:</p> <ol style="list-style-type: none"> 1. A single controller failure. 2. RAID-TP 3 disks failure. 3. The interface card failure. 4. A single network port failure.

Quiz

1. (True or False) The entire backup system includes multiple components, such as front-end production applications, networks, backup servers, and backup storage.
2. (Multiple-choice) Which of the following are recommended for OceanProtect testing?
 - A. Performance test for backup to tape libraries
 - B. Performance test for archiving to tape libraries
 - C. Reduction ratio test
 - D. High reliability:

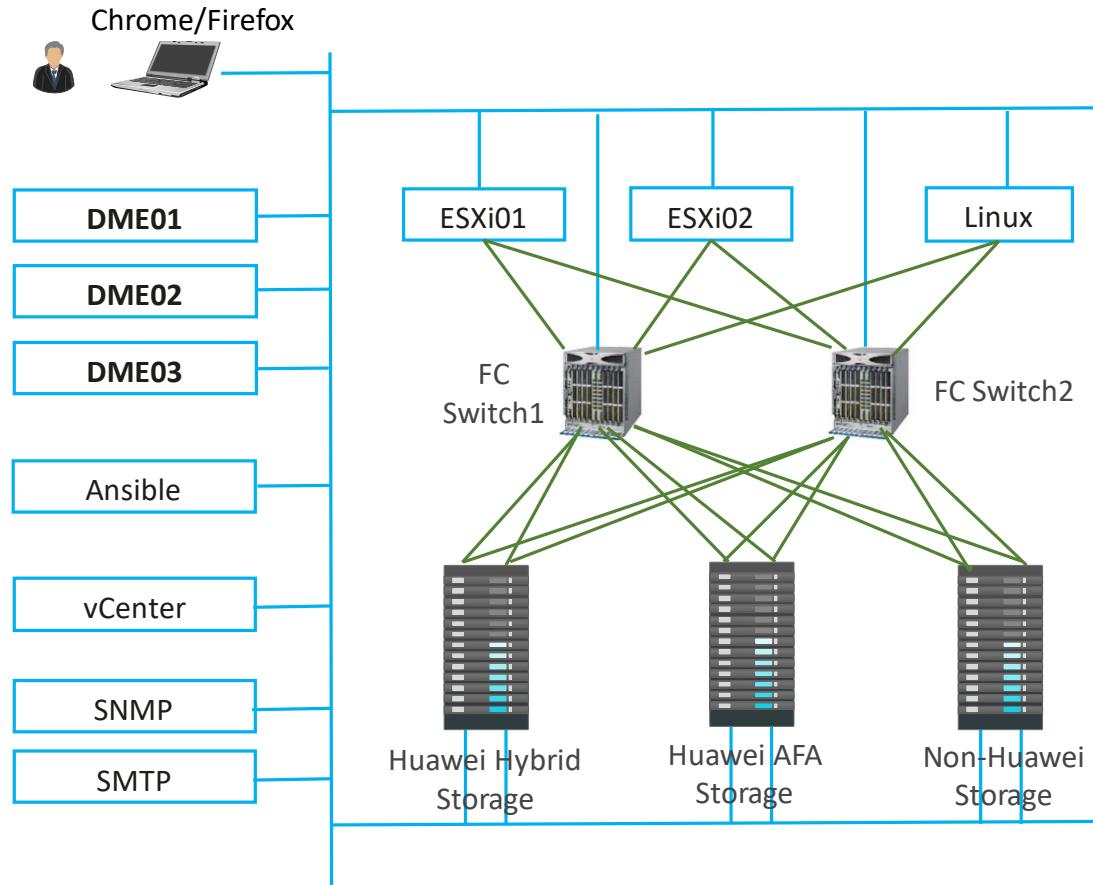
Contents

1. POC Overview
2. Recommended POC Contents for Flash Storage
3. Precautions for flash POC Performance Tests
4. Precautions for flash POC Reliability Tests
5. OceanProtect POC Tests
- 6. Precautions for DME storage POC**
7. Precautions for Scale out Storage POC
8. How to Conduct a Satisfactory POC

Overview and Objectives

- This section describes DME storage POC test.
- On completion of this section, you will be able to:
 - Understand the networking of DME storage.
 - Understand Precautions for DME Storage in the demo.

Precautions for DME Storage



- ✓ Network between DME internal nodes: The recommended network bandwidth is **1Gbit/s**, the network latency **<10ms**, and the packet loss rate **<0.1%**.
- ✓ Network between DME and southbound devices: Network bandwidth **>100Mbit/s**, the network latency **<50ms**, and the packet loss rate **<0.1%**.

Test Scenario	Description
Basic functions introduction	OceanStor DME, homepage and basic functions
Service catalog	Infrastructure Asset Management, Add new device, delete device
	Catalog definition, SLA, pools, multi tenancy, Quota
Storage as a service	Automatic Resource Provisioning, verify, resource pool reporting
Intelligent O&M	Intelligent O&M: Discover Risk → Trouble-shooting → Problem Solved
	Multi-dimensional Dashboard Customization

DME demo:

<https://info.support.huawei.com/storage/dme-demo/#/home>

Quiz

1. (True or False) DME Storage is defined as a full-lifecycle automated management platform

Contents

1. POC Overview
2. Recommended POC Contents for Flash Storage
3. Precautions for flash POC Performance Tests
4. Precautions for flash POC Reliability Tests
5. OceanProtect POC Tests
6. Precautions for DME storage POC
- 7. Precautions for Scale out Storage POC**
8. How to Conduct a Satisfactory POC

Overview and Objectives

- This section describes the recommend test case of scale out storage POC
- On completion of this section, you will be able to:
 - Understand the recommended test solution and networking of scale out storage POC .
 - Understand Precautions for Scale out Storage POC functions/performance/reliability Tests

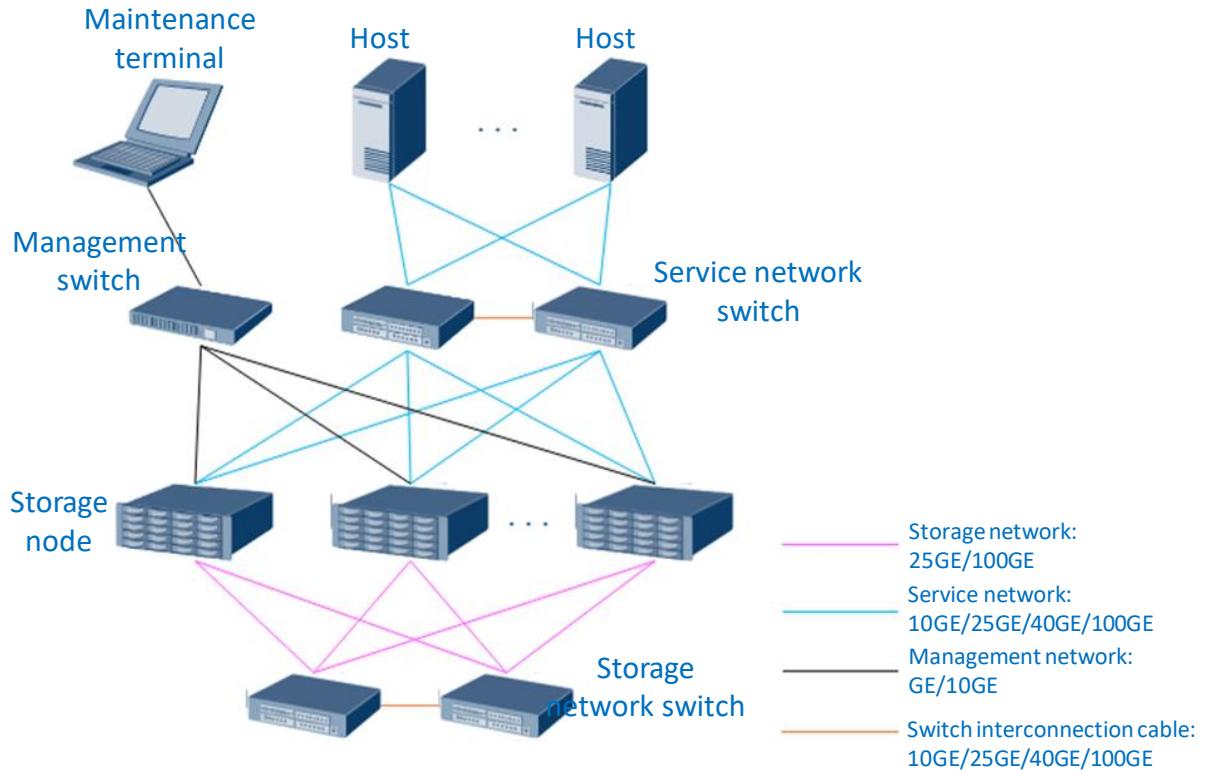


- What products do you know about Scale out storage? What services are supported?
- What are the differences in functions of different products? What are the main application scenarios?

Recommended Test Solution for Scale out Storage POC

File service	<table border="1"><tr><td data-bbox="327 357 481 423">OceanStor Pacific</td><td data-bbox="532 357 942 423">Parallel file system, optimal performance</td><td data-bbox="993 357 1198 423">Automated data tiering</td></tr><tr><td data-bbox="327 437 481 501">OceanStor 9000</td><td data-bbox="532 437 942 501">Mature and rich enterprise features</td><td data-bbox="993 437 1198 501">High reliability</td></tr><tr><td data-bbox="993 437 1198 501"></td><td data-bbox="993 437 1198 501"></td><td data-bbox="1198 437 1280 501">Media</td></tr><tr><td data-bbox="1198 437 1280 501"></td><td data-bbox="1198 437 1280 501"></td><td data-bbox="1280 437 1361 501">Video</td></tr></table>	OceanStor Pacific	Parallel file system, optimal performance	Automated data tiering	OceanStor 9000	Mature and rich enterprise features	High reliability			Media			Video	<ul style="list-style-type: none">• OceanStor Pacific: Supports parallel file systems and provides excellent performance in large and small I/O scenarios.• OceanStor 9000: Supports various enterprise features (snapshot, WORM, and remote replication) and provides stable reliability assurance mechanisms. Delivers excellent performance in media and intelligent video storage scenarios.
OceanStor Pacific	Parallel file system, optimal performance	Automated data tiering												
OceanStor 9000	Mature and rich enterprise features	High reliability												
		Media												
		Video												
Object services	<table border="1"><tr><td data-bbox="327 596 481 648">OceanStor Pacific</td><td data-bbox="532 596 942 648">Hundreds of billions of objects per bucket</td><td data-bbox="993 596 1198 648">Performance with a large number of objects</td></tr><tr><td data-bbox="327 682 481 734"></td><td data-bbox="532 682 942 734">Object-level deduplication</td><td data-bbox="993 682 1198 734">Synchronous or asynchronous remote replication</td></tr><tr><td data-bbox="1198 682 1280 734"></td><td data-bbox="1198 682 1280 734"></td><td data-bbox="1280 682 1361 734">Service QoS</td></tr></table>	OceanStor Pacific	Hundreds of billions of objects per bucket	Performance with a large number of objects		Object-level deduplication	Synchronous or asynchronous remote replication			Service QoS	<ul style="list-style-type: none">• A single bucket supports hundreds of billions of objects, meeting users' requirements for mass objects.• With hundreds of millions of objects in the bucket, the performance is continuously excellent and does not drop sharply.• Supports enterprise value-added features such as object deduplication and service QoS.• Synchronous/Asynchronous remote replication, meeting various remote DR requirements			
OceanStor Pacific	Hundreds of billions of objects per bucket	Performance with a large number of objects												
	Object-level deduplication	Synchronous or asynchronous remote replication												
		Service QoS												
HDFS service	<table border="1"><tr><td data-bbox="327 852 481 904">OceanStor Pacific</td><td data-bbox="532 852 942 904">Gateway-free native HDFS interface</td><td data-bbox="993 852 1198 904">Kerberos</td><td data-bbox="1198 852 1280 904">EC utilization</td></tr><tr><td data-bbox="327 904 481 956"></td><td data-bbox="532 904 942 956">Coexistence of old and new clusters</td><td data-bbox="993 904 1198 956"></td><td data-bbox="1280 852 1361 904">Flexible expansion of compute/storage</td></tr></table>	OceanStor Pacific	Gateway-free native HDFS interface	Kerberos	EC utilization		Coexistence of old and new clusters		Flexible expansion of compute/storage	<ul style="list-style-type: none">• No additional gateway or plug-in is required. The storage system directly provides native HDFS interfaces for external systems and supports unified authentication with Hadoop.• The EC redundancy algorithm is used to provide up to 90% or higher capacity utilization.• Flexible deployment: Supports independent deployment and coexistence of old and new devices. Independent and flexible expansion of computing and storage resources.				
OceanStor Pacific	Gateway-free native HDFS interface	Kerberos	EC utilization											
	Coexistence of old and new clusters		Flexible expansion of compute/storage											
Block service	<table border="1"><tr><td data-bbox="327 1080 481 1131">OceanStor Pacific</td><td data-bbox="532 1080 942 1131">Lossless snapshots</td><td data-bbox="993 1080 1198 1131">QoS policy</td><td data-bbox="1198 1080 1280 1131">EC performance</td></tr><tr><td data-bbox="327 1131 481 1183"></td><td data-bbox="532 1131 942 1183">Deduplication and compression</td><td data-bbox="993 1131 1198 1183"></td><td data-bbox="1280 1080 1361 1131">Active-active + Asynchronous replication</td></tr></table>	OceanStor Pacific	Lossless snapshots	QoS policy	EC performance		Deduplication and compression		Active-active + Asynchronous replication	<ul style="list-style-type: none">• Lossless snapshot: Minimizes the impact of snapshot creation and deletion on performance.• EC + deduplication and compression: Ensures optimal performance while maximizing available user space.• Active-Active/Asynchronous Replication: Multiple remote DR protection solutions.				
OceanStor Pacific	Lossless snapshots	QoS policy	EC performance											
	Deduplication and compression		Active-active + Asynchronous replication											

PoC Test Networking and Configuration Principles - Scale out Storage



Name	Hardware Configuration Principles
Storage node	<ol style="list-style-type: none"> 2 U or 4 U storage nodes (subsequent 5 U high-density nodes) can be configured. The number of CPUs and memory modules are configured based on the eDesigner. At least three nodes (single service) are required for in a test environment for basic function performance tests. If the capacity expansion function needs to be tested, at least four nodes are required (3 + 1 expansion) To test remote DR functions such as active-active and remote replication, at least six nodes (3 + 3 for two clusters) are required. If the project has specific requirements on performance, evaluate the number of required nodes based on the performance requirements.
Storage network switch	<p>To meet expectations, use recommended switches: Huawei 10GE switches (such as CE6881) or 25GE switches (such as CE6865 and CE6863) 100 Gbit/s IB switch, for example, Infiniband SB7800 EDR 100G switch</p>
Server	<p>2 U x86 server (RH2288 series recommended) To test the ultimate storage performance, it is recommended that the number of hosts be twice the number of storage nodes. Physical machines are recommended.</p>

compatibility query: <https://support-open.huawei.com/en/>

Precautions for Scale out Storage POC Function Tests

Service Type	Advantageous Test case	Description
OceanStor 9000 file service	Quota/snapshot/tiering/cross-protocol permission interworking/single-node multi-VLAN/file filtering	<p>Quota: supports default quotas, user quotas, and user group quotas.</p> <p>Snapshot: supports periodic snapshots.</p> <p>Hierarchy: Cold and hot levels are supported.</p> <p>Cross-protocol permission Interworking : The same user has same permission on the Linux&Windows platforms.</p> <p>Single-node multi-VLAN: network isolation</p> <p>File filtering: Blocks files with certain extensions to improve file system security.</p>
OceanStor Pacific converged service	Multi-protocol interworking/QoS/quota/metadata retrieval/recycle bin/soft encryption/Worm/cross-protocol user mapping/bucket policy control/object multi-version/object multi-DC	<p>Multi-protocol Interworking: supports NFS, CIFS, HDFS, and S3 protocols to access the same file.</p> <p>QoS: The bandwidth and OPS QoS can be configured on clients, accounts, and namespaces.</p> <p>Quota: supports user and user group quotas.</p> <p>Metadata search: supports the search of the home account, namespace, file, path, and last modification time.</p> <p>Recycle bin: Set the recycle bin for the command space. Deleted data is not deleted immediately. You can set the retention period of the recycle bin.</p> <p>Data soft encryption: Data encryption can be configured for namespaces. The encrypted data is stored on disks.</p> <p>Worm: Allows users to set Worm for namespaces, which can be written once and read multiple times.</p> <p>Cross-protocol user mapping: supports Linux to Windows user mapping and Linux to Unix user mapping.</p> <p>Bucket policy control: Allows users of the current account or other accounts to perform operations on buckets and objects in buckets.</p> <p>Object versioning: Object versioning can be enabled on buckets.</p> <p>Multi-DC object: DR control in two-site and three-center scenarios is supported.</p>
OceanStor Pacific block service	Volume capacity expansion/volume snapshot and clone/QoS/deduplication and compression	<p>Volume capacity expansion: supports volume capacity expansion.</p> <p>Volume snapshot and clone: Allows users to set snapshots for volumes and clone snapshots.</p> <p>QoS: QoS and QoS burst can be set for bandwidth and IOPS.</p> <p>Deduplication and compression: supports deduplication and compression settings.</p>

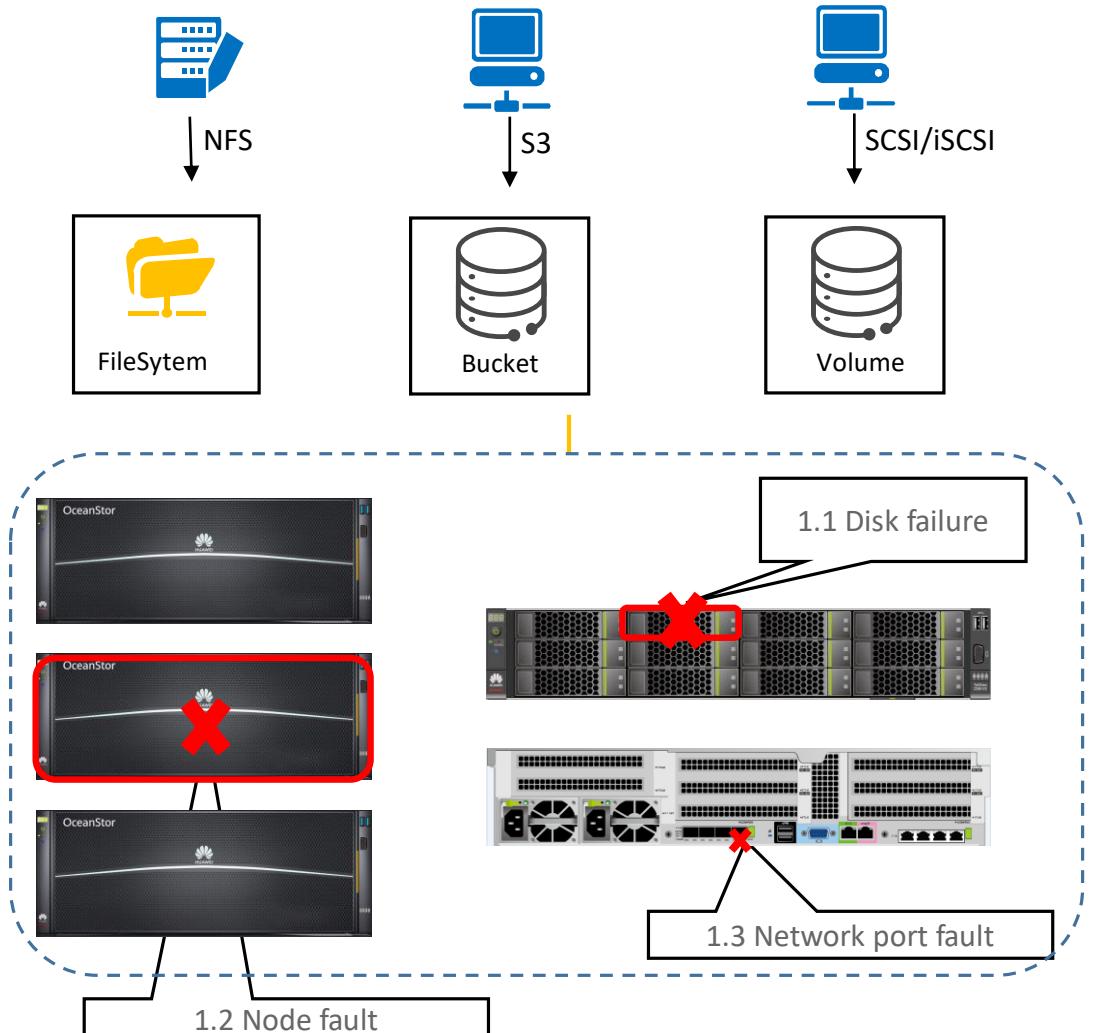


- What are the main causes that affect the performance test?
- Which of the following products are used for testing in video storage and small-file scenarios?

Precautions for Scale out Storage POC Performance Tests

Service Type	I/O Types for Advantageous Service Matching	Typical Performance Test Model and Key Test Points
OceanStor 9000 file service	Large-file and large-I/O scenarios such as media NLE and intelligent video storage	<p>Adapts to large block read/write tests OceanStor 9000 delivers excellent performance in large-block read/write scenarios, meeting project requirements in high-bandwidth scenarios.</p>
OceanStor Pacific file service	High-performance computing (HPC)	<p>Select an I/O model based on project requirements Large I/Os of large files, random small I/Os of large files, and read and write of small files.</p>
OceanStor Pacific object service	Upload and download performance of small- and large-sized objects in a bucket	<p>Test in the background of mass data Recommend the single-bucket test scenario. Pre-embed hundreds of millions of objects before the formal test to verify the performance when there are a large number of objects.</p>
OceanStor Pacific block service	Verify that the random write performance can reach 8 KB.	<p>Performance in EC mode You are advised to perform the test using the same data protection policy. EC of Pacific is more mature.</p> <p>8 KB is the main I/O size of database OLTP services In small-block (8 KB) random write scenarios, the performance is good, and the advantage in read scenarios is small.</p> <p>Test the real end-to-end I/O delivery capability in hybrid configuration. Avoid the full cache hit test of small-capacity LUNs. Otherwise, all data is stored on SSDs and the actual HDD storage performance cannot be tested. The capacity of the LUN to be tested must be larger than the storage cache. After the test volume is fully written and data is continuously read and written for a period of time, perform a formal test.</p> <p>All-flash test In all-SSD configuration, the network and CPU must be synchronized with the configuration (for example, 25GE RoCE network and 6-core CPUs) to prevent the network and CPU from becoming performance bottlenecks</p>

Precautions for Scale out Storage POC Reliability Tests



No	Common Test Item	Description & Observation Point
1.1	Disk fault	<p>In online service scenarios, remove data disks and observe the system and service performance.</p> <p>Observation Point:</p> <ol style="list-style-type: none"> I/Os return to zero when a disk is removed. Performance fluctuation after a disk is removed. After a disk is removed, check whether the system reports an alarm and whether the alarm points to the fault point. Whether the system automatically starts data reconstruction.
1.2	Storage node fault	<p>In online service scenarios, randomly power off nodes and observe the system and service performance.</p> <p>Observation Point:</p> <ol style="list-style-type: none"> Maximum number of nodes that can be powered off in a specified node configuration. I/Os suspension when a node is powered off.
1.3	Network port fault	<p>In the online service scenario, remove the network cable from the node network port and observe the system and service performance</p> <p>Observation Point:</p> <ol style="list-style-type: none"> I/Os return to zero when the network cable removed. Check whether the network port has subhealth detection capabilities such as delay and packet loss.

Quiz

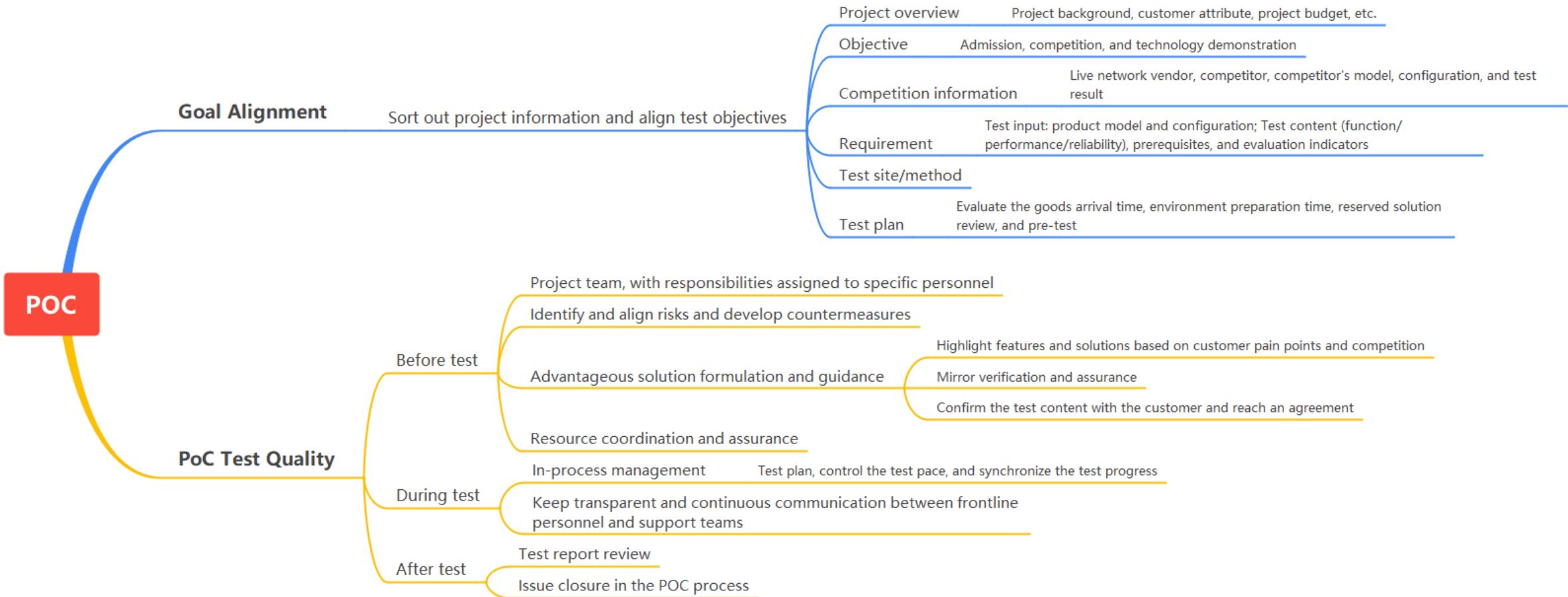
1. (True or False) It is recommended that OceanStor Pacific file service be tested in the single-bucket test scenario, Pre-embed hundreds of millions of objects before the formal test to verify the performance when there are a large number of objects.
2. (Multiple-choice) Which of the following are the advantages of converged service?
 - A. Worm
 - B. Data soft encryption
 - C. Snapshot
 - D. Object multi-DC

Contents

1. POC Overview
2. Recommended POC Contents for Flash Storage
3. Precautions for flash POC Performance Tests
4. Precautions for flash POC Reliability Tests
5. OceanProtect POC Tests
6. Precautions for DME storage POC
7. Precautions for Scale out Storage POC
- 8. How to Conduct a Satisfactory POC**

Overview and Objectives

- This section describes POC test tips.
- On completion of this section, you will be able to:
 - How to Conduct a Satisfactory POC.



Precautions for POC

- ❑ Test strategy: It is important to obtain and provide feedback on the project background, project nature, live network conditions, RFP/RFI or function and performance scoring table, and compatibility requirements in advance. The test strategy is an important input for formulating POC test strategies and evaluating risks.
- ❑ Device selection: When selecting test devices, ensure that the specifications and performance meet the requirements in the bidding document and test requirements, and consider the comparison of bidding models.
- ❑ **Test version:**
 - Select the corresponding version based on the actual test requirements and install the latest patch.
- ❑ Pre-test: If possible, perform more pre-tests to get familiar with products and solutions and train technical personnel.
- ❑ Technical Q&A and key messages: Key technical issues, roadmaps, and specifications must be unified.
- ❑ **Formal test:**
 - Do not perform POC tests in the production environment that has been delivered to the customer.
 - Do not use test devices to access the production environment. If the customer and frontline personnel need to forcibly access the production network, including accessing the production network through eService, they must obtain a written disclaimer from the customer before accessing the production network.
 - Onsite test risks and problems must be synchronized promptly. Onsite engineers can temporarily prevent the risks and problems, and seek help from back end immediately.

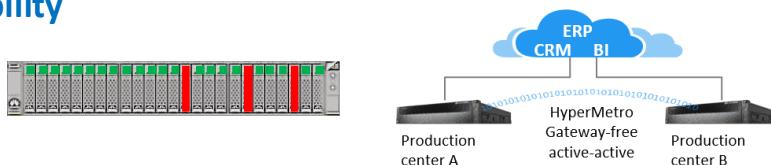
POC Suite + MOOC + HOL

What to Test? Recommended Test Scenarios

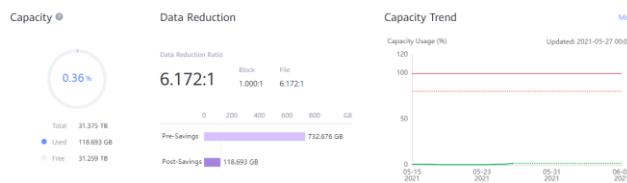
Excellent performance



High reliability



Rock solid



Know yourself and others, and select the most suitable solution for the customer

A set of test case standards, which can be used for special purposes based on project requirements

How to Test? Step by Step Operation Guide

2.2 Local Data Protection and Efficiency

2.2.1 High Density Zero-performance Impact Snapshot

Objective: To validate that the OceanStor Dorado supports high-density snapshots and the impact of high-density snapshots on performance.

Prerequisite:

1. The OceanStor Dorado storage system is running properly.
2. The host multipathing software (UltraPath) is installed on the host.
3. In the OceanStor Dorado storage system, one or more RAID 6 storage pools have been created.
4. Created sixteen 500GB LUNs with deduplication and compression.

Procedure:

1. Map 16 LUNs to the test servers.
2. Use the Vdbench tool to configure 256KB I/O without deduplication compression, and overwrite the LUNs once in 100% sequential mode.

Vdbench script sample (Linux):

Quick Simple Test

✓ Check the vdbench tool is available

--Linux: Enter the vdbench directory

Execute the command:

root@linux vdbench\$./vdbench -t

--Windows: Start Command Prompt

Enter the vdbench directory

Execute the command:

C:\vdbench\50406>vdbench -t

Note:

✓ When running ./vdbench -t Vdbench will run a hard-coded sample run. A small temporary file is

created and a 50/50 read/write test is executed for just five seconds.

✓ This is a great way to test that Vdbench has been correctly installed and works for the current OS

platform without the need to first create a parameter file.

Test case verification guide

All courses > MOOC > OceanStor Dorado 6.0 All-Flash Storage PoC Test Suite



Storage Online Class – Dorado

Dorado POC 2.0 suite

Chinese version: <http://3ms.huawei.com/documents/docinfo/396747674952970240?bookstackId=35621>

English version: <http://3ms.huawei.com/documents/docinfo/396747363111895040>

[Storage Online Class-PoC] Dorado V6 All-Flash Storage POC Test Suite

Chinese version: <https://ilearningx.huawei.com/portal/courses/HuaweiX+EBGTC00000465/about>

English version: <https://ilearningx.huawei.com/portal/courses/HuaweiX+EBGTC00000403/about>

Storage HandsOn online test platform

HandsOnLab

Homepage Cloud Computing Intelligence Data Intelligence Storage Server Help

Tags: All HandsOn Video OnlinePoC

Products: OceanStor Dorado OceanStor V3/V5 Solution Germany IT PoC LAB

FusionAccess Online PoC Config

Warm reminder: online POC course requires the use of Chrome browser

OceanStor Dorado 18000 Configuration 627

Huawei OceanStor Dorado 6000 V6 Storage Active-Active Solution 971

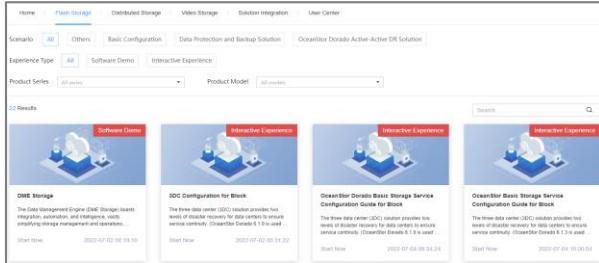
OceanStor Dorado 5000 RoCE vs FC Performance Test 6

OceanStor Dorado5000 V3(SAS) Configuration 9

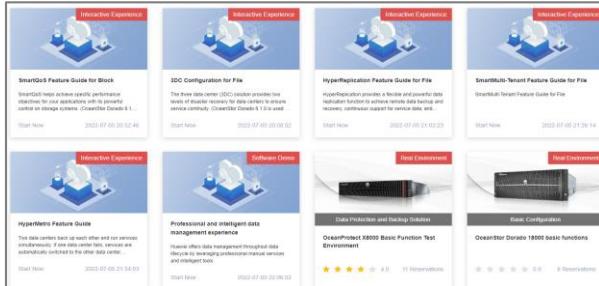
HOL Guide

What's HOL? Key Capabilities

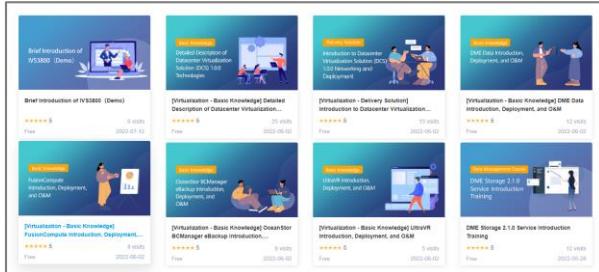
Massive Resource, Continuous Update



Global and Multi-environment Access



supporting course teaching

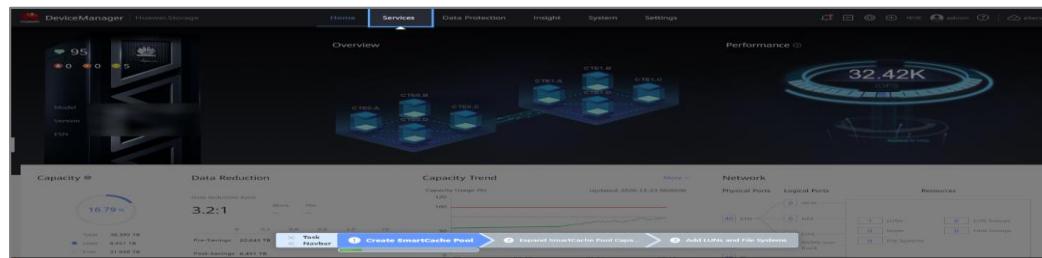


How to get HOL Resource? [link](#)

Interactive Experience & demo

Includes basic configuration guides for file and block storage. No approval is required.

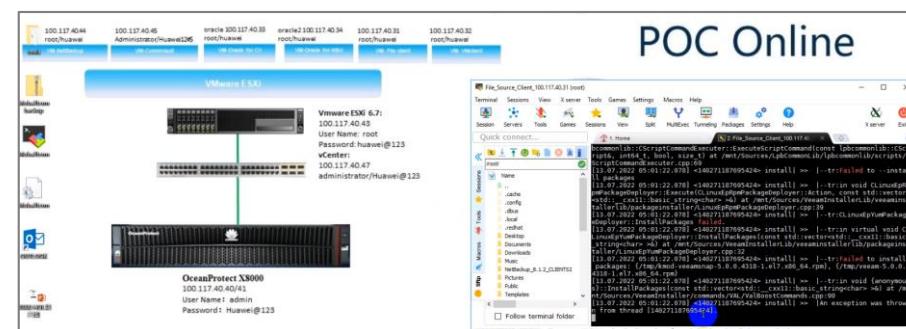
Steps: 【Scenario】 -> 【Software Demo/Interactive Experience】



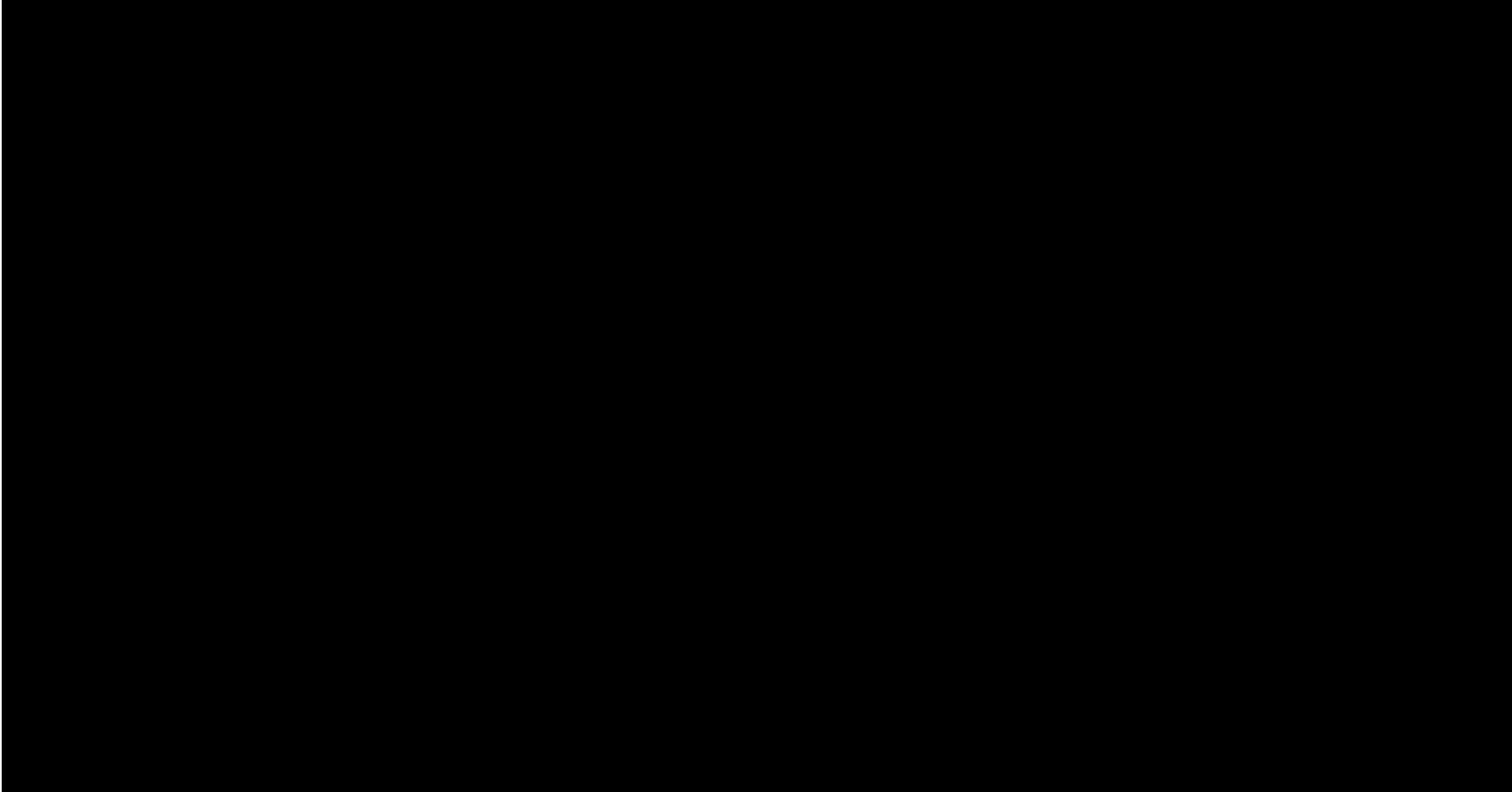
Real Environment for POC

Provide real test resources for the POC project, fill in the project information, and complete the demonstration/workshop after the project is approved (some resources do not need to be approved).

Steps: 【Scenario】 -> 【Real environment】 -> 【Reserve Test】 -> 【Approved】 -> 【User Center】 -> 【Start Test】

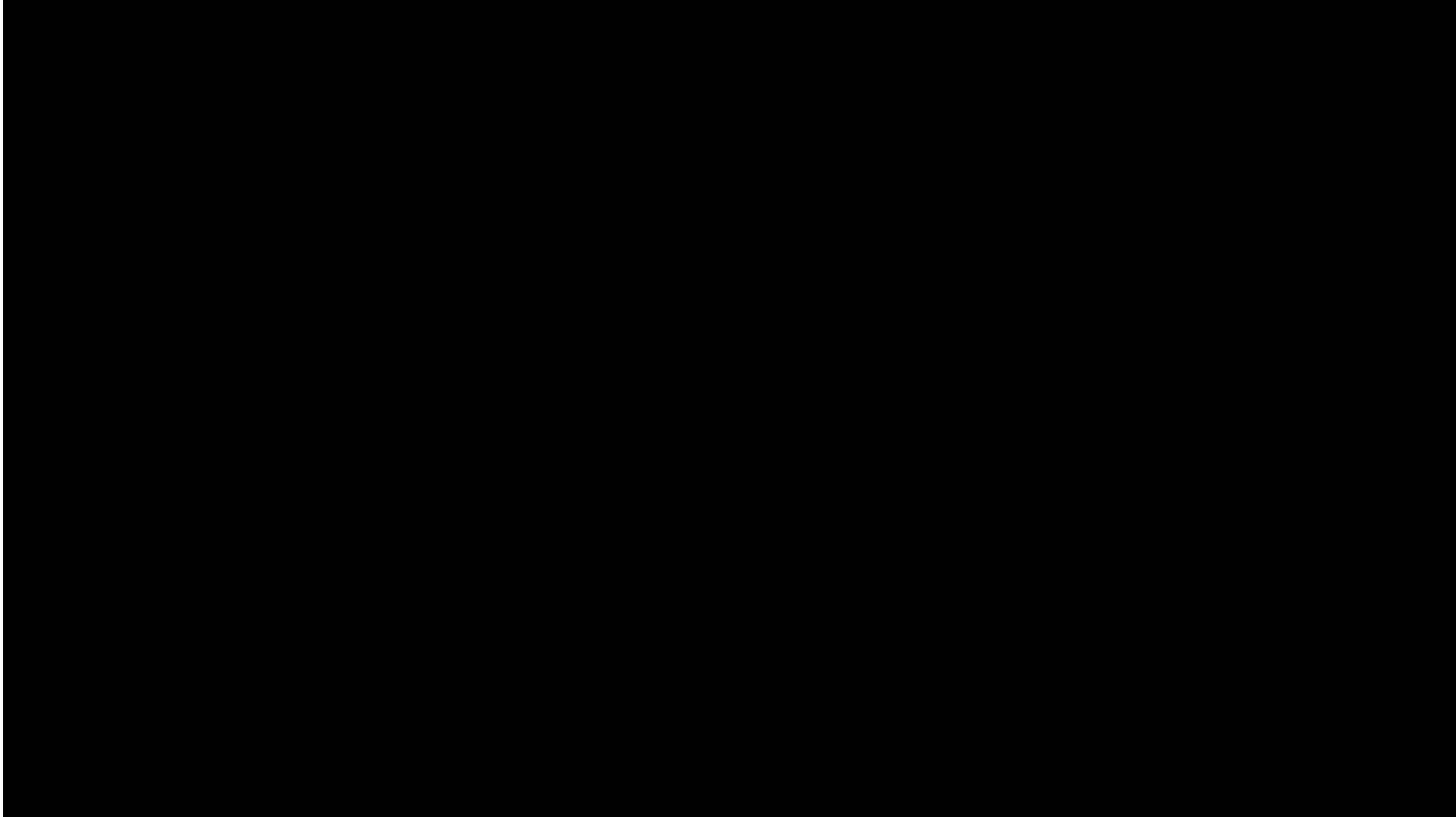


Scale Out storage POC Reliability Test - Disk Fault



1. Apply for Pacific test resources on [HOL](#).
2. Login HOL resource, get the environment information.
3. Follow the video and finish test in HOL.

Scale Out storage POC Reliability Test - Node Fault



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.

