|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | |  |
|  | | |
|  | | |
| ½Ô²穙k1 | | | |
|  | OceanStor 9000 Technical Proposal Template | | ¸½¼þ1-16K |
|  | |
| **Issue** | **V1.0** |
| **Date** | **201701** |
|  | |
| HUAWEI TECHNOLOGIES CO., LTD. | |
|  | | |

|  |
| --- |
| Copyright © Huawei Technologies Co., Ltd. 2017. All rights reserved.  No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.  Trademarks and Permissions  image001 and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.  All other trademarks and trade names mentioned in this document are the property of their respective holders.  **Notice**  The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.  The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied. |

|  |  |
| --- | --- |
| Huawei Technologies Co., Ltd. | |
| Address: | Huawei Industrial Base  Bantian, Longgang  Shenzhen 518129  People's Republic of China |
| Website: | http://enterprise.huawei.com |

Contents

[1 Overview 4](#_Toc425581509)

[1.1 Project Background 4](#_Toc425581510)

[1.2 Live Network Status 4](#_Toc425581511)

[1.3 Scale and Objectives 5](#_Toc425581512)

[2 Design Basis 6](#_Toc425581513)

[2.1 Solution Design Basis 6](#_Toc425581514)

[2.2 Solution Design Principles 6](#_Toc425581515)

[3 System Design Scheme 8](#_Toc425581516)

[3.1 Project Requirements 8](#_Toc425581517)

[3.2 System Design 9](#_Toc425581518)

[3.3 Topology 11](#_Toc425581519)

[3.4 Software and Hardware Configurations 12](#_Toc425581520)

[4 Features and Advantages 14](#_Toc425581521)

[4.1 Flexible Expansion 14](#_Toc425581522)

[4.2 Outstanding Performance 15](#_Toc425581523)

[4.3 Reliable Data Protection 17](#_Toc425581524)

[4.4 Simplified Management 20](#_Toc425581525)

[4.5 Fined-Grained Quota Management 20](#_Toc425581526)

[5 Product Introduction 22](#_Toc425581527)

[5.1 Product Overview 22](#_Toc425581528)

[5.2 Product Composition 22](#_Toc425581529)

[5.3 Product Specifications 24](#_Toc425581530)

[6 List of Abbreviations 27](#_Toc425581531)

# Overview

*This chapter describes the project construction background.*

## Project Background

*Introduce the project, ensuring to use language that describes the particulars of the project. Avoid using overly negative words like "fail". State the project needs without using language that is condescending towards the customer or their IT investments.*

*Example*

The pace of business development at XXX (fill in customer's name) has caused significant increase in service data, adding tremendous strain to its current IT system. Server disks struggle to deliver the required reliability, capacity, and performance. XXX is in need of a centralized storage system able to share data, dramatically improve service system performance, and ensure data security while enabling ease of capacity expansion on local disks.

## Live Network Status

*You are advised to provide a network topology in the description of live network structure. The topology indicates the quantity and types of existing devices.*

*Example*

In XXX's network, unstructured data is stored across various storage devices, including SAN and NAS devices that support clustered networking. There are 10-odd types of upper-layer services, including web disk, email, and office automation (OA). Operating systems include Windows, Linux, and UNIX. Mission-critical data is backed up while other data is sent to archives.

The following figure shows the live network topology.

Table 1-1 provides the device type and quantity deployed on the live network.

Devices deployed on the live network

|  |  |
| --- | --- |
| Storage Device | Quantity |
| Huawei S6800T | 2 |
| Huawei N8500 | 1 |
| EMC Centera | 2 |
| Switch | Quantity |
| Cisco 3900 | 10 |
| Huawei S5352 | 20 |
| Service |  |
| Web disk |  |
| Outlook |  |
| SMS |  |

## Scale and Objectives

*Describes the conceptualization (scale and objectives)* *of the new network in this section.*

*Example*

The new system XXX (customer name) intends to build will need to support 10-odd service systems, including CRM and ERP. The systems will run various operating systems, including Windows, Linux, and UNIX. Databases will include Oracle and SQL databases. Some of the service system data will be stored in local disks on servers.

XXX's detailed storage needs:

* Massive data of various service systems must be stored. Placing the OceanStor 9000 in the data center enables unstructured data of all service systems to be stored in a central repository.
* The estimated daily amount of generated is between 100 GB to 500 GB, which means the total volume will reach 500 TB within XX years. The build must provide ample initial storage capacity to satisfy these needs and the storage system mush have the scalability to meet capacity expansion needs over the long term.
* Growing access concurrency needs must also be met as services continue to grow. That is, concurrency capabilities must be enhanced in line with capacity expansions to improve overall performance.
* The storage system must support various types of services and store unstructured data.

*In the next paragraph, explain how Huawei Storage can meet all project requirements.*

The Huawei OceanStor 9000 can meet all project requirements, able to centrally store and share all types of service data, provide more than ample initial storage space for current needs and the scalability to meet future needs, and ensure performance through expansions as concurrent access requests increase.

# Design Basis

*This chapter briefly explains the design basis and principles of the solution. It is also possible to combine the sections into one section or add other sections to suit the particular proposal.*

## Solution Design Basis

*Describe the basis of the solution design in this section.*

*Example*

This solution design described herein has fully considered all requirements specified in *XX Bidding Specifications*.

## Solution Design Principles

*This section describes the solution design principles.*

*Example*

To fully satisfy project requirements on the XX system, the Huawei solution provides the following:

* Security and reliability
* Distributed architecture helps ensure data remains protected and available.
* High availability in terms of hardware, data, and services.
* High reliability with support of service and storage clustering.
* Impressive data redundancy and error tolerance capabilities.
* Advanced technologies and mechanisms to ensure data reliability.
* Protection against data loss in the event a storage node becomes damaged (including disk and cabinet faults).
* Data self-healing utilities provide added assurances in recovering damaged data.
* Adaptable to a full range of operating environments.
* Secure login and access, protection against attacks.
* Protection against data loss in the event of a power failure; after the power supply recovers, the storage system automatically restarts and restores connections.
* System components can work continuously for 24/7 hours under heavy loads.
* Scalability
* Robust scalability, providing up to over 10 PB in total capacity.
* Performance in line with capacity expansion.
* Standardized components enable ease of replacement and scale out.
* Compliance and technical relevance
* Complies with international standards, national standards, and the in-country norms of the communications industry.
* Huawei regularly updates it technology and devices in line with development trends in the storage and IT industry. All product models selected for this project are able to ensure high-volume production.
* Cutting-edge technologies going into the solution ensure technical relevance for years to come.
* Industry-leading processing capabilities and the flexibility to meet future upgrade requirements.
* Openness
* Support for international standard network storage protocols and open application protocols.
* Compatibly with mainstream servers.
* Support for mainstream operating systems, volume management software, and applications.
* Able to integrate with third-party management platforms, enabling customized management and maintenance approaches.
* Sufficient scalability to suit future service requirements.
* Maintainability
* User-friendly multilingual user interface.
* Support for permission management, log management, fault management, and automatic alarming utilities.
* System devices are easy to install and use, freeing up higher-level maintenance staff to perform higher-end tasks.
* On-demand system capacity expansion without the need to suspend services.
* Hitless upgrade of system functions.
* Support for web-based management.
* Cost-effectiveness
* Optimal price/performance ratio when compared to similar offerings.
* Energy efficiency and environmental protection
* Low noise output.
* Lead-free components.
* Energy-saving technologies.
* Environmental management certified; packaging uses recyclable materials.

# System Design Scheme

*This chapter explains the design scheme in detail. You can combine or delete sections to suit the particular proposal.*

With developments in IT, the amount of unstructured data produced is soaring upwards as new business and service models continue to constantly spring up everywhere. Conventional storage systems often cannot meet the mushrooming requirements. Huawei released the OceanStor 9000, a Big Data storage system, to address storage needs now and in the future. OceanStor 9000 provides massive storage space with 100 PB maximum capacity for a single file system. The impressive appliance is capable of carrying various services, adapting to an extensive range of use cases with external NAS access interfaces coupled with the openness to interoperate with a wide-range of applications. The deployment scheme and configurations specific to this project are described in the following sections of this chapter.

## Project Requirements

As XXX's business continues to develop, a variety of new applications are being deployed in its IT environment. These deployments cause an exponential increase in data and place much higher requirements on storage performance and capacity. These changes bring the following challenges to XX:

* Rapid increase in unstructured data. The total amount of production data has reached xx TB, straining the existing storage system that does not support horizontal expansion; expanding this system would entail high costs.
* Average of XX TB in data generated each day. XXX must expand system capacity to handle the amount of data being generated each day as current resources become exhausted.
* Growing diversity in data sources poses higher requirements on data reliability. Disk utilization must be balanced while data reliability requirements are met. This balancing prevents sacrificing large amounts of storage space to meet reliability requirements (required storage space increases several fold compared to the original amounts of data).
* Variety of service systems must be able to share data. In the existing system, storage space provided by multiple storage systems cannot be shared, and management is complex. Multiple IT administrators are needed to maintain various storage devices and networks as maintenance costs continue to increase, leading to a soaring TCO.
* Need for improved power efficiency. Existing devices are power intensive and must operate for long periods given the lack of ample resources. In addition, requirements on the working temperature and heat dissipation in the equipment room must be met, leading to ever-increasing electricity costs at the data centers.

The above considerations culminate into the need for the new storage models and systems while allowing for integration, to the extent possible, with existing devices. The new models and systems are expected to feature the following:

* Support for massive amounts of data storage and horizontal expansion.
* The flexibility to expand storage capacity whenever needed without compromising performance.
* The capability to store and share data from a wide range of services and systems.
* Highly reliable in data storage.
* Support for thin provisioning.
* Simplified storage management.
* High efficiency and energy savings.

## System Design

The Huawei OceanStor 9000 is a distributed storage solution designed to handle a variety of upper-layer file storage services, allowing customers to store various types of service data. The solution requires that upper-layer services comply with their interface specifications. The solution features:

* Unified namespace capable of storing files, videos, images, and other data generated from a variety of services.
* Three–node initial configuration employs a symmetric distributed architecture for easy capacity expansion, meeting growing storage needs.
* Performance improves in line with capacity expansion, preventing performance bottlenecks caused by increases in data.
* Support for hitless online expansion. In addition, multiple nodes can be expanded in batches.
* Clustering deployment and erasure code data protection across the entire system.
* N+1 to N+4 data reliability.
* Automatic fault detection and repair technologies able to ensure system reliability.
* External NAS interfaces and support for NFS and CIFS sharing, applicable to such application scenarios as interworking with file and image services.
* Support for thin provisioning enabling flexible space expansion to suit the specific needs of each service, preventing waste of storage space due to use of fixed quotas.
* Wide range of value-added features such as quota management, tiered storage, snapshot, WORM, and replication for NAS storage and deduplication, Multi-tentant and data transmission encryption for object storage.
* The web UI management tool allows administrators to manage the storage device in a unified manner.

Huawei Storage also offers a variety of nodes for use with the OceanStor 9000 to suit different application scenarios. These nodes can coexist in the same storage system and provide a unified namespace. The appropriate number of each type of node to meet specific storage tiering needs can be tailored to each use case, thereby reducing total cost of ownership (TCO). Huawei offers P and C series nodes, each with distinct performance and capacity specifications.

* P series, applicable to high-performance use cases

P25E: 2 U, 25 disks, support for 600 GB SSD, 900 GB SSD, 600 GB SAS, 900 GB SAS, 1.2 TB SAS and 1.8 TB SAS

P36E: 4 U, 36 disks, support for 600 GB SSD, 900 GB SSD, 2 TB/4 TB/6 TB/10 TB SATA, 2 TB/4 TB/6 TB NL-SAS

P12E: 2 U, 12 disks, support for 600 GB SSD, 900 GB SSD, 2 TB/4 TB/6 TB/10 TB SATA, 2 TB/4 TB/6 TB NL-SAS

* C series, applicable to video surveillance and near-line archiving

C36E: 4 U, 36 disks, support for 2 TB/4 TB/6 TB/10 TB SATA, 2 TB/4 TB/6 TB NL-SAS

C72: 4 U, 72 disks, support for 2 TB /4 TB/6 TB/10 TB SATA

Different types of nodes can be intermixed to achieve optimal performance. However, the same node type must be added in groups of three or more.

The same node can be configured with the NAS interface or S3 interface or Swift interface, if storage need to support a variety of interfaces in a system, the configuration of each interface at least 3 nodes.

This solution is most widely used in XXX application scenarios requiring XXX TB of effective capacity and XXXX performance.

*Refer to the following examples to calculate the number of OceanStor 9000 nodes needed to achieve the required capacity and performance for the project. The larger of the two values is used to determine the number of required nodes.*

Calculating the required capacity for this project:

*The following describes the formulas used to calculate the required capacity for two typical industries. Apply this basic idea to the particular industry the customer is engaged in and supply the pertinent content to suit the case.*

*Example from media industry:*

Code stream (for example: 100 Mbit/s) x Number of edit layers (4) x Number of workstations (30)/8 = 1.5 GB/s

Effective capacity = Bandwidth x Time (second)

*Example from video surveillance industry:*

Effective capacity = Number of channels x Code stream x 1.1 (bit rate fluctuation coefficient) x (3600 x 24 x 30) /8/1024/1024

Based on the calculated capacity required, the number of required OceanStor 9000 nodes is as follows:

Effective capacity of SAS disks = Raw capacity x Utilization rate x 0.9 (0.9 is the file system loss rate. The utilization rate is related to the number of nodes and the N+M value, and it can be obtained from the redundancy ratio table) x 0.909 (optional, 1024 band conversion).

Number of SAS disks = Raw capacity/Capacity per node

Effective capacity of SATA disks = Raw capacity x Utilization rate x 0.9 (0.9 is the file system loss rate. The utilization rate is related to the number of nodes and the N+M value, and it can be obtained from the redundancy ratio table).

Number of SATA disks = Raw capacity/Capacity per node

Calculating the required performance for this project:

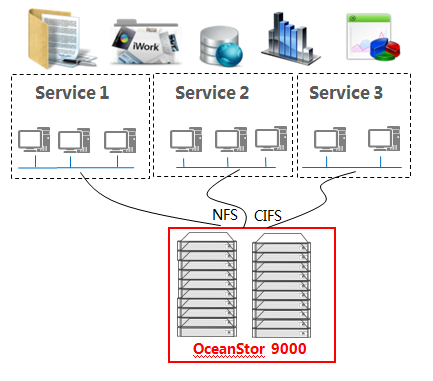
*The following describes the calculation for determining the needed level of performance for this project. Use this calculation or similar and adapt the content to case.*

The system requires XXXX bandwidth, so the required level of performance is as follows:

Number of nodes = System bandwidth/Bandwidth per node/0.8 (0.8 is the average linear coefficient. It is determined by the number of nodes and redundancy ratio, and can be obtained from the specifications list.)

Based on the capacity and performance requirements, configure X P25E nodes and Y C36E nodes to achieve a raw capacity of XX. The redundancy ratio is N+M, so the effective capacity is XX and performance is XX, meeting the needs of the XX service system of XXX customer.

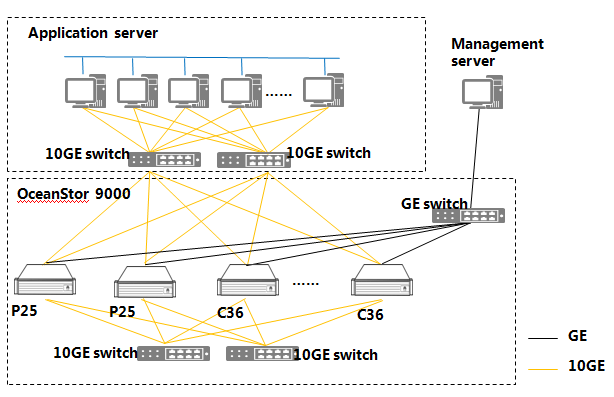
OceanStor 9000 solution architecture



## Topology

The OceanStor 9000 supports a variety of networking modes. Both internal and external networks can employ 10GE, 56Gbps InfiniBand, and GE networking modes. Specific networking modes can be flexibly selected based on project and service requirements. System capacity can be smoothly expanded to more than 10 PB as services grow. Figure 3-1 shows the storage network topology.

Storage network topology



The solution features:

* Centralized storage of all mission-critical services and support for multi-protocol sharing.
* 10GE networking for both internal and external networks.
* Improved access efficiency with use of proprietary TOE network adapters and remote direct memory access (RDMA) technologies.
* Support for client connection load balancing and quota management.
* Dynamic storage tiering (DST) that stores hotspot data on the faster P25E nodes while cold data is stored on C36E or C72 nodes.
* Snapshot and other utilities for rapid data recovery and backup.

## Network Deployment

Configuration x and y units P25E node node station C36E. Cabinet networking is deployed as follows:



Network Assistant can refer to the following link:

<http://app.huawei.com/unistar/edesigner/configuration!showParas.action?projectId=10323974&siteId=473365>

## Software and Hardware Configurations

Software and hardware configurations

| Site | Device Name | Specifications or Model | Quantity | Remarks |
| --- | --- | --- | --- | --- |
| XX equipment room | OceanStor 9000 P25E | Dual-channel eight -core CPU, 48 GB memory, 600 GB SSD + 21.6 TB SAS, 2 x 10GE network adapters | X |  |
| XX equipment room | OceanStor 9000 C36E | Dual-channel eight -core CPU, 48 GB memory, 600 GB SSD + 140 TB SATA, 2 x 10GE network adapter | X |  |
| XX equipment room | 10GE switch | CE6800 | 2 |  |
| XX equipment room | GE switch | S5700 | 1 |  |
| XX equipment room | Cabinet |  | 1 |  |
| XX equipment room | KVM |  | 1 |  |
| XX equipment room | SMS modem |  | 3 | Each system has three management nodes, each requiring an SMS modem. |
| XX equipment room | Wushan distributed file system software |  | X |  |
| XX equipment room | Load balancing software |  | X |  |
| XX equipment room | Quota management software |  | X |  |
| XX equipment room | DST software |  | X |  |

The configuration consists of x P25E, y C36E, 2 Gigabit switches and 1 Gigabit Switch, maximum system power consumption is xxxx, weight is xxxx.



product capability assessment can refer to the following link:

<http://app.huawei.com/unistar/edesigner/solutionAction!showHome.action>

# Features and Advantages

Designed for Big Data, the Huawei OceanStor 9000 features impressive scale-out capabilities. Service dynamics constantly change and data increases rapidly in Big Data use scenarios, placing ever-increasing requirements on data security for storage systems. The OceanStor 9000 delivers flexible space, outstanding performance, simplified management, and solid reliability, addressing the pressing challenges in storing and accessing vast amounts of data while reducing capital expenditure (CAPEX) and TCO for enterprises.

## Flexible Expansion

Conventional storage systems, whether employing the early monolithic architecture or the more current modular architecture, are all based on scale-up design. Such types of design require adding computing resources, such as processors, to meet performance requirements as applications are added — a rather cost-intensive approach to improving performance. In addition, achieving improved performance through the scale-up design is restricted by many aspects, such as the number of disks, the number of connections, the cache size, and the number of controllers. Therefore, most current storage systems employ a scale-out storage architecture. Enterprises can add different servers and applications to suit their storage needs. Multiple servers and storage systems compute and execute tasks in parallel. In addition, functions such as load balancing and error tolerance are implemented. In this manner, computing capabilities and system reliability are improved. The scale-out architecture no longer adopts the centralized memory and large-scale controller model. Instead, independent modules are used, and each module has its own processing units, caches, host interfaces, and systems. When performance or capacity requirements change, modules can be adjusted to meet the need. Adding to the attraction, cache size, processing capabilities, number of disks, and bandwidth can be increased or reduced in proportion to the modules being added or deleted, thereby achieving optimal use of resources and making the entire storage system more agile, stable, and scalable.

The OceanStor 9000, which employs a scale-out storage architecture, distributes the loads from the massive data to multiple storage nodes that execute the tasks in concurrency. Data and metadata are evenly distributed on all nodes, preventing competition for resources among different applications. System performance (throughput) is expanded in proportion to scale-outs and loads are balanced among storage nodes, preventing performance bottlenecks from occurring on a single node. The OceanStor 9000 architecture delivers improved scalability while retaining the high reliability and performance advantages of object-based storage systems. Based on the scale-out architecture, OceanStor 9000 allows easy capacity expansion from 3 to 288 nodes.

The OceanStor 9000 allows smooth capacity expansion in line with service growthand is able to automatically identifys each newly added node. Single nodes can be added in less than 60 seconds, making added storage space available across the entire system almost immediately. Based on the load balancing principle, capacity expansion management is simplified, and operation costs are reduced As capacity and performance can be expanded linearly. Capacity of a single file system reaches up to 100 PB, truly achieving "scalability on demand" and ensuring a superb price/performance ratio for the long-term.

## Outstanding Performance

With up to five million operations per second (OPS), less than 40 ms of latency, and over 400 GByte/s of bandwidth, OceanStor 9000 delivers industry-leading NAS read and write performance to address the requirements of such fields as media editing and high-performance computing.

The OceanStor 9000 employs the key technologies outlined in the following sections to achieve its impressive performance.

### SSD Acceleration and DST

Solid-state drives (SSDs) provide low latency and high throughput to eliminate I/O performance bottlenecks. The OceanStor 9000 uses SSDs to store metadata, improving metadata access efficiency and processing capabilities.

In addition, the InfoTier utility (DST software provided by the OceanStor 9000), automatically identifies small-sized files and hotspot data. Based on the service value of the information and collection of intelligent I/O access frequency statistics, the value of data to the phase or application can be accurately determined. Based on the access frequency, data is migrated between different storage media to achieve optimum distribution of data. High value data is automatically distributed to high-level storage media while low-level data is automatically migrated to cheaper storage media, improving resource utilization. The migration process is transparent to services.

InfoTier implements fine-grained data statistics collection and automatic migration. In addition, InfoTier allows users to configure policies based on specific service characteristics, such as file attributes (path, file name, creation time, access time, modification time, file size, and UID/GID) and access frequency. InfoTier improves access to hotspot data, making better use of high-value disks. The statistical period has a major impact of the analysis and handling of cold and hot data. Therefore, ensuring the proper period is set helps the system obtain the actual frequency at which certain types of data are accessed.

### Global Cache

Reading or writing data on the cache is much faster than doing so on a disk. However, caches have little room for expansion. In addition, if the cache of one node is independent from that of another node, it is difficult to increase the cache hit ratio, hindering increased assurances in data consistency.

The global cache mode on the OceanStor 9000 consolidates all node caches, forming a virtual cache pool of up to 73 TB. Only one copy of data of the same file is stored in the cache pool and is accessible to any node. This increases the data hit ratio, reduces the number of disk reads and writes, and shortens access latency, thereby improving overall system performance.

### High-Speed Interconnection

The OceanStor 9000 supports 10GE, and each node provides four 10GE ports. Additionally, the OceanStor 9000 supports InfiniBand (IB) networks that deliver up to 56 Gbps of bandwidth, meeting demanding performance requirements.

By interconnecting nodes at high speed, the OceanStor 9000 reduces internal latency and provides outstanding performance, adapting to different networking requirements.

### Concurrent Access

Traditional NAS systems consist of engines and storage units. All concurrent data access events must be processed by NAS engines first, making these engines a performance bottlenecks. The OceanStor 9000 employs a symmetric logical architecture that enables all nodes to provide services simultaneously. The load balancing design enables loads from data access requests to be distributed evenly in a cluster, greatly improving access concurrency.

The OceanStor 9000 provides a unified domain name for the external to access. The domain name query request service is integrated with the load balancing service. When a client initiates a domain name query request, the load balancing service calculates the load based on the configured load balancing policy and then returns the IP address of an appropriate node to the client to access the OceanStor 9000. The OceanStor load balancing utility intelligently manages connections to clients and evenly balances loads and switches over services for the paths. This function greatly improves system availability and effectively eliminates system performance bottlenecks.

### DST

InfoTier is the dynamic storage tiering (DST) software provided by the OceanStor 9000, which automatically identifies hotspot data in the system. Based on the service value of the information and collection of intelligent I/O access frequency statistics, the value of data to the phase or application can be accurately determined. Based on the access frequency, data is migrated between different storage media to achieve optimum distribution of data. High value data is automatically distributed to high-level storage media while low-level data is automatically migrated to cheaper storage media, improving resource utilization. The migration process is transparent to services.

InfoTier implements fine-grained data statistics collection and automatic migration. In addition, InfoTier allows users to configure policies based on specific service characteristics, such as file attributes (path, file name, creation time, access time, modification time, file size, and UID/GID) and access frequency. InfoTier improves access to hotspot data, making better use of high-value disks. The statistical period has a major impact of the analysis and handling of cold and hot data. Therefore, ensuring the proper period is set helps the system obtain the actual frequency at which certain types of data are accessed.

## Reliable Data Protection

As IT develops, enterprises have come to realize that data is one, if not the most, important asset to just about every operation. Data loss can have a crippling effect on a company. Therefore, increasingly more importance is being attached to data protection. The OceanStor 9000 adopts the best data protection mechanism in the industry to ensure high data reliability.

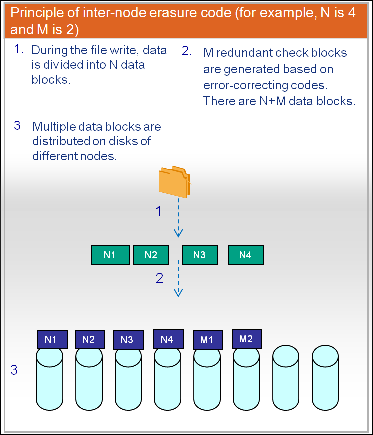
### Erasure Code

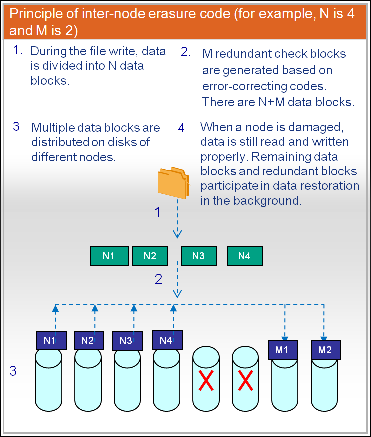
The OceanStor 9000 employs erasure code technology to ensure data security and reliability in scenarios requiring massive amounts of storage. Erasure code is a RAID superset that supports redundancy policies that are more flexible and provides higher reliability than traditional RAID algorithms.

The erasure code mechanism divides files into N original slices, and M redundant check block files are generated based on error-correcting codes. When writing a file, the client performs slicing (N) and transcoding (M), and N+M pieces of data are generated. The storage system automatically selects a disk from each storage node to from a disk group that consists of N+M disks, and writes the N+M pieces of data to the N+M disks, respectively. In the event that any one piece of data is damaged, the data can be restored using the other pieces of data. Damage to a maximum of M pieces of data is tolerated.

The OceanStor 9000, which employs the erasure code algorithm to ensure data reliability, stores N+M pieces of data on N+M virtual nodes, where N ranges from 2 to 16, and M ranges from 1 to 4(Distributed File M ranges from 1/2/3/4, distributed Object M ranges from 1/2/3). If four nodes become damaged at the same time, the OceanStor 9000 is capable of restoring the original data using other data slices without affecting service continuity, providing better data protection than traditional RAID 5 (which tolerates damage to one piece of data) and RAID 6 (which tolerates damage to two pieces of data). Figure 4-1 shows the working principle of the erasure code mechanism.

Erasure code principle





The OceanStor 9000 distributed storage system provides optimized storage of small-sized files. If the size of a file is not larger than that of one slice, the OceanStor 9000 working in N+M protection mode automatically divides the file into 1+M pieces and stores the data. If the size of a file is larger than that of one slice, the OceanStor 9000 working in N+M protection mode checks the data in the file.

The optimization improves performance in reading and writing small-sized files and enhances storage utilization.

### High Metadata Reliability

The OceanStor 9000 distributed storage system stores multiple copies of metadata. In addition, the number of metadata copies stored for a folder under N+M protection is automatically set to M+2. This functionality enhances the reliability of metadata stored on the system.

### Data Self-Healing

The OceanStor 9000 distributed storage system employs the InfoProtector, which is one of Huawei's patented technologies. InfoProtector provides data protection against simultaneous failure of up to four nodes. Also, InfoProtector helps ensure minimal reconstruction time of a single faulty disk with a restoration speed of up to 1 TB per hour.

Data to be stored is sliced, and slices are distributed to multiple storage nodes. If a data error occurs, the internal self-check mechanism automatically discovers the data fault. Then, the OceanStor 9000 automatically starts the data repair mechanism in the background with multiple nodes working in parallel. As each node needs to repair a small proportion of data only, performance bottlenecks are avoided and the impact on upper-layer services is minimized.

### Remote DR and Backup

The OceanStor 9000 provides a directory-level snapshot function which protects all the files and sub-directories under a specific path; the content of other files remains undisturbed. The InfoReplicator function is implemented based on the directory-level snapshot and asynchronously replicates data between the primary and secondary directories. Before data is synchronized, it creates a snapshot for the data to be backed up and then starts the synchronization based on the snapshot. This mechanism causes near-zero impact on customer services and ensures the consistency and integrity of the data. Enabling InfoReplicator allows critical data to be quickly and efficiently replicated to the backup site. This function can also be used for remote data archiving. It can replicate data to a remote site and to a level-2 storage system if needed. Then, the original local data is deleted and local storage space is released. The data archived remotely can be retained for a long period of time.

## Simplified Management

The OceanStor 9000 supports visualized and automatic deployment. The OceanStor 9000 provides a dedicated network management tool that supports a web UI. This wen UI displays the network topology containing all devices; each device's information such as capacity, CPU, and memory utilization; and the running status of each process and service. In addition, the tool displays system performance in real time and allows users to save and query operation logs. The tool is easy to use, providing centralized and efficient management and maintenance. An administrator can access the network management page using any node in a cluster, completly unifying management of hardware, software, clusters, and services. Single node failures do not affect user ability to log in to the network management page or perform operations on the network management page.

### Thin Provisioning

In the event that many applications are running on the client side, data storage space assigned to some applications may become used up, whereas that assigned to other applications may remain unused. The OceanStor 9000, with support for thin provisioning, dynamically allocates storage space based on customer needs. In addition, flexible capacity expansion is supported. When the amounts of data that need to be stored increases, more storage space is allocated. Customers are no longer burdened with trying to accurately plan the needed storage space for different applications at the initial stage of the project, and wastes in overplanning can be avoided.

## Fined-Grained Quota Management

The OceanStor 9000 implements fine-grained management on the storage space. The OceanStor 9000 supports quota management and configuration at three levels: user, user group, and directory. Also, nested quota management is supported for enhanced resource management. The InfoAllocator feature of the OceanStor 9000 provides the following quotas:

* Directory-specific quota: The administrator can set a quota for an empty shared directory to limit the maximum storage space available to the shared directory.
* User/User group-specific quota: The administrator can set a quota for any user or user group to limit the maximum storage space available to the user or user group.

# Product Introduction

## Product Overview

Designed for Big Data, the Huawei OceanStor 9000 features a fully symmetrical distributed architecture to deliver cutting-edge performance, large-scale horizontal expansion capabilities, and a super-large single file system – providing shared storage for structured and unstructured data. Adding to the appeal, OceanStor 9000 employs the Huawei proprietary erasure code data protection technology, delivers industry-leading read and write performance, and provides unified interfaces externally to carry various services, meeting the needs of the Big Data era.

## Product Composition

The OceanStor 9000 is able to achieve a superb balance in capacity and performance with flexibility to combine node types to suit the particulars of each use case. These nodes can coexist in the same storage system and provide a unified namespace to enhance system capabilities.

Huawei offers P and C series nodes, each with distinct performance and capacity specifications.

* P series, applicable to high-performance use cases

P25E: 2 U, 25 disks, 2.5 inch hard disk

P36E: 4 U, 36 disks, 3.5 inch hard disk

P12E: 2 U, 12 disks, 3.5 inch hard disk

* C series, applicable to video surveillance and near-line archiving

C36E: 4 U, 36 disks, 3.5 inch hard disk

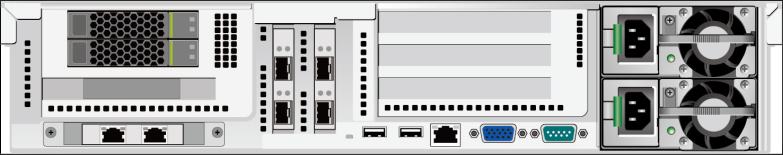
C72: 4 U, 72 disks, 3.5 inch hard disk

Different types of nodes can be intermixed to achieve optimal performance. However, the same node type must be added in groups of three or more.

The following figures show the front and rear views of each node.

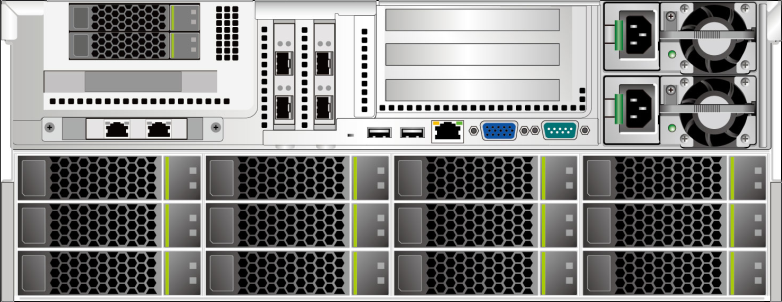
P25E front and rear views





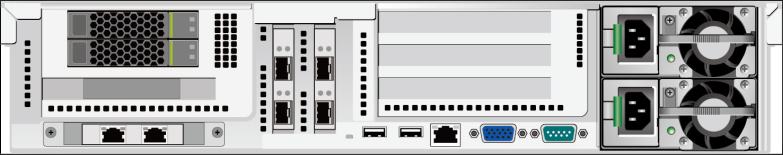
P36E/C36E front and rear views



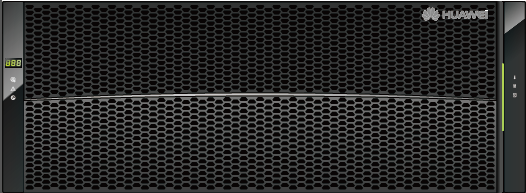


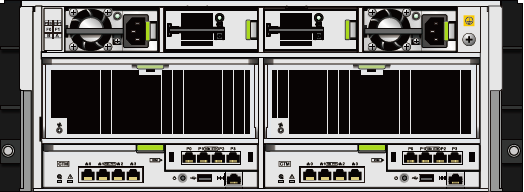
P12E front and rear views

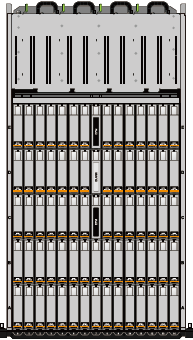




C72 front, rear and top views







OceanStor 9000 supports 10GE, InfiniBand, and GE internal and external networking. The front-end and back-end networks are separated. Table 5-1 lists the networking types supported by OceanStor 9000.

Networking types supported by OceanStor 9000

|  |  |  |  |
| --- | --- | --- | --- |
| Front-End Service Network | Back-End Storage Network | Switch Used in Internal Network | Is This the Recommended Networking? |
| 10GE access | 10GE access | CE6800 | Yes; highly recommended |
| 10GEaccess | IB access | Mellanox 6018F | - |
| IB access | IB access | Mellanox 6018F | - |
| GE access | 10GE access | CE6800 | - |

Figure 5-4 shows the rear view of the 10GE switch CE6800, which is used in 10GE internal networking.

CE6800 switch unit



# List of Abbreviations

This chapter lists the expansions for the acronyms found in this document.

*Example*

|  |  |
| --- | --- |
| Abbreviation | Full name |
| NAS | Network Attached Storage |
| UI | User Interface |
| SAN | storage area network |
| SAS | serial attached SCSI |
| SATA | Serial Advanced Technology Attachment |
| SSD | solid state disk |
| TCP | Transmission Control Protocol |
| UID | user ID |
| NFS | Network File System |
| CIFS | Common Internet File System |
| WORM | Write Once Read Many |
| S3 | Amazon S3 Interface |
| Swift | OpenStack Swift Interface |
| TOE | TCP Offload Engine |
| RDMA | Remote Direct Memory Access |