

Ao

**GOVERNO DO ESTADO DO ESPÍRITO SANTO**

**COMPANHIA ESPÍRITO SANTENSE DE SANEAMENTO - CESAN**

**A/C: Exma. Sra. THATIANA SANTOS DE MELLO**

Divisão de Compras e Suprimentos (A-DCS)

Pregoeira responsável pelo Pregão Eletrônico em referência

**REF: RECURSO ADMINISTRATIVO - PREGÃO ELETRÔNICO Nº 017/2025**

**CONTRA: Decisão do Pregoeiro que declarou vencedora a empresa COMPWIRE INFORMÁTICA LTDA. (CNPJ: 01.181.242/0001-91)**

### **RECURSO ADMINISTRATIVO**

A empresa **MICROWARE TECNOLOGIA DE INFORMAÇÃO LTDA**, empresa constituída e existente de acordo com as leis do Brasil, com sede a rua Av. Lagoa Encantada, nº 220 – Galpão A6 – Sala 01 – Vale Encantado, Vila Velha/ES, CEP.: 29.113-515 inscrita no CNPJ sob o Nº 01.724.795/0007-39, na qualidade de licitante do PREGÃO ELETRÔNICO em referência, vem respeitosamente, perante Vossa Senhoria, com fundamento no **art. 56 da Lei Federal nº 13.303/2016 (Lei das Estatais)**, interpor o presente **RECURSO ADMINISTRATIVO** em face da decisão proferida pela Sra. Pregoeira que declarou como vencedora a empresa **COMPWIRE INFORMÁTICA LTDA.**, pelas razões de fato e de direito a seguir expostas.

#### **I. DOS FATOS**

Em 22/08/2025, foi realizada a sessão pública do Pregão Eletrônico CESAN nº 017/2025, ocasião em que o Sr. Pregoeiro, após análise das propostas apresentadas, declarou como vencedora a empresa COMPWIRE INFORMÁTICA LTDA.

Ocorre que, conforme se demonstrará adiante, a proposta da referida empresa apresenta **graves inconsistências técnicas, contradições documentais e flagrante descumprimento aos requisitos editalícios**, o que impõe sua imediata desclassificação do certame.





## II. DO DIREITO

### 2.1. DOS PRINCÍPIOS FUNDAMENTAIS DAS LICITAÇÕES

A Lei nº 13.303/2016 estabelece, em seu artigo 31, os princípios que devem nortear os procedimentos licitatórios das sociedades de economia mista, dentre os quais destacamos:

- **Legalidade:** toda licitação deve observar rigorosamente as normas legais e editalícias;
- **Impessoalidade:** vedação a qualquer tratamento preferencial ou discriminatório;
- **Moralidade:** exigência de conduta ética e probidade;
- **Publicidade:** transparência nos procedimentos;
- **Eficiência:** busca pela melhor relação custo-benefício;
- **Interesse público:** finalidade precípua de atender às necessidades coletivas;
- **Probidade administrativa:** dever de atuação honesta e íntegra;
- **Igualdade:** tratamento isonômico entre os licitantes;
- **Vinculação ao edital:** estrita observância às regras previamente estabelecidas.

### 2.2. DA OBRIGATÓRIA DESCLASSIFICAÇÃO DE PROPOSTAS IRREGULARES

A Lei nº 13.303/2016, em seu **artigo 32, §2º**, é clara ao dispor que:

“Não se admitirá, no processo de licitação, a inclusão de exigências ou documentos não previstos no instrumento convocatório, **bem como deverão ser desclassificadas as propostas que não atenderem às exigências do edital.**”

Além disso, o **artigo 32, §1º** da mesma lei determina que:





“O julgamento das propostas será objetivo e será realizado em conformidade com os critérios previamente estabelecidos no edital.”

Portanto, fica evidente que qualquer proposta que apresente **documentação desnecessária, divergente, incompleta ou incompatível com o edital** deve ser **desclassificada**, sob pena de violação ao princípio do **julgamento objetivo**.

### III. DA ANÁLISE TÉCNICA DA PROPOSTA IRREGULAR

Submetida a rigorosa análise técnica, a proposta apresentada pela empresa **COMPWIRE INFORMÁTICA LTDA** revelou-se incompatível com os requisitos estabelecidos no edital. Foram apresentadas **mais de 8.000 (oito mil) páginas de documentação**, contendo inúmeros **catálogos de produtos não solicitados e diferentes versões de documentos**, numa clara tentativa de **confundir a análise técnica e dificultar a aferição do atendimento às exigências editalícias**.

Essa conduta **viola diretamente** o princípio do **julgamento objetivo**, previsto no **art. 32, §1º** da Lei nº 13.303/2016, e **compromete a economicidade e a transparência** que devem nortear os processos licitatórios. Além disso, ao se analisar detalhadamente o conteúdo apresentado, verificam-se **múltiplas e graves inconsistências** que comprometem irremediavelmente a idoneidade da proposta, conforme será demonstrado a seguir.

#### 3.1. MISTURA INADMISSÍVEL DE VERSÕES DE SOFTWARE

A proposta da empresa recorrida apresenta **confusão documental**, utilizando simultaneamente documentações das versões 1.5.0, 1.6.0 e 1.7.0 do produto OceanProtect, sem qualquer critério técnico ou justificativa plausível.

Esta **incoerência documental grosseira** gera incerteza jurídica inaceitável sobre qual versão será efetivamente entregue, configurando proposta **imprecisa e contraditória**, em frontal violação ao princípio da vinculação ao edital.

**ITEM DESCLASSIFICATÓRIO:** Funcionalidades essenciais como SourceDedupe/DataTurbo aparecem em documentos das versões 1.6.0 e 1.7.0, enquanto trechos da documentação fazem referência à versão 1.5.0, que





**comprovadamente não possui tais recursos.** Tal inconsistência revela **total desconhecimento técnico** e compromete a exequibilidade da proposta.

### **3.2. DESCUMPRIMENTO DOS REQUISITOS DE PROTEÇÃO CONTRA RANSOMWARE (Item 1.1.12 do edital)**

O item 1.1.12.3 do Termo de Referência **exige categoricamente** isolamento completo dos dados, sem exposição via protocolos CIFS/NFS, visando à proteção contra ataques de ransomware.

A documentação apresentada pela recorrida menciona o uso da tecnologia DataTurbo, porém, **contraditoriamente, parte significativa da documentação técnica ainda referencia protocolos tradicionais CIFS/NFS, gerando dúvida técnica inaceitável sobre o real atendimento ao requisito de segurança.**

**ITEM DESCLASSIFICATÓRIO: Se o equipamento ofertado permitir acesso via CIFS/NFS aos dados protegidos, haverá DESCUMPRIMENTO TOTAL do requisito editalício de invisibilidade da superfície de ataque, comprometendo a segurança dos dados da CESAN.**

### **3.3. AUSÊNCIA DE COMPROVAÇÃO DE LICENCIAMENTO INTEGRAL (Item 1.4 do edital)**

O edital **exige expressamente** que todas as funcionalidades estejam licenciadas para toda a capacidade útil solicitada. A proposta menciona licenças para 40TB + 230 unidades de 41–300TB, porém **omite completamente** o detalhamento sobre se todas as funcionalidades críticas (dedupe, replicação, criptografia, etc.) estão ativadas para os 200TB úteis.

Esta **omissão técnica** impossibilita a verificação do real atendimento aos requisitos editalícios, configurando proposta **incompleta e imprecisa.**

### **3.4. DESDUPLICAÇÃO GLOBAL NÃO COMPROVADA (Item 1.4 do edital)**





O edital demanda **expressamente** deduplicação global entre protocolos e clientes. A documentação apresentada menciona apenas deduplicação por sistema, **sem qualquer comprovação técnica** de que a deduplicação é efetivamente global e cruzada entre os protocolos CIFS/NFS/OST.

**ITEM DESCLASSIFICATÓRIO:** A ausência de comprovação técnica objetiva sobre este requisito essencial **torna a proposta tecnicamente inadequada e incompatível** com as necessidades da CESAN.

### **3.5. AUSÊNCIA COMPLETA DE SUPORTE AO PROTOCOLO OST (Item 1.8.13 do edital)**

**ITEM DESCLASSIFICATÓRIO:** O edital exige **CATEGORICAMENTE** que o equipamento suporte os protocolos CIFS, NFS e OST **SIMULTANEAMENTE**.

O datasheet oficial do Huawei OceanProtect X6000 lista como protocolos suportados: NFS, SMB/CIFS, S3, FC, iSCSI, FTP/SFTP e NDMP, **NÃO CONSTANDO O PROTOCOLO OST** (Veritas OpenStorage).

A empresa recorrida, em tentativa **tecnicamente inaceitável**, tenta equiparar sua tecnologia proprietária "DataTurbo" ao protocolo OST exigido. Trata-se de **tentativa de burlar** os requisitos editalícios, uma vez que DataTurbo é recurso próprio da Huawei e **NÃO SE CONFUNDE** com o protocolo OST especificado no edital.

**ITEM DESCLASSIFICATÓRIO:** Há **descumprimento objetivo e incontestável** ao requisito editalício, o que **impõe obrigatoriamente** a desclassificação da proposta.

### **3.6. DESCUMPRIMENTO DO REQUISITO DE CRIPTOGRAFIA SEM IMPACTO DE PERFORMANCE (Item 1.1.19 do edital)**

O Termo de Referência estabelece **inequivocamente** que a solução deve ofertar criptografia de dados **sem impacto de desempenho**.

**ITEM DESCLASSIFICATÓRIO:** O próprio manual técnico da Huawei **expressamente alerta** que a ativação da criptografia pode impactar o desempenho, recomendando inclusive análise específica no configurador ("**consultar eDesigner para impacto**").





A recorrida declarou throughput de 45 TB/h com criptografia ativa, porém **não comprovou** a inexistência de impacto em relação ao cenário sem criptografia, apresentando apenas **alegação sem substrato técnico**.

**ITEM DESCLASSIFICATÓRIO:** O requisito não foi atendido, restando apenas alegação genérica **desprovida de comprovação técnica objetiva**.

### 3.7. INTEGRAÇÃO DEFICIENTE COM SOFTWARES DE BACKUP (Item 1.9.3 e 2.18)

O edital **exige categoricamente** integração **PLENA** com os softwares de mercado adotados pelo órgão (NetBackup e Commvault).

A empresa recorrida apresentou **apenas declaração genérica** de fabricante, **desprovida** de:

- Matriz oficial de compatibilidade;
- Lista detalhada de versões suportadas;
- Comprovação técnica efetiva da interoperabilidade.



**Site oficial da Commvault demonstra a ausência de compatibilidade com Huawei para appliance de backup** conforme link a seguir: <https://www.commvault.com/supported-technologies#>

**ITEM DESCLASSIFICATÓRIO:** O fabricante Huawei não consta na matriz de compatibilidade conforme portal Commvault.

### 3.8. PROPOSTA TECNICAMENTE INCONSISTENTE E NÃO FIRME





A análise global da proposta revela **padrão sistemático de inconsistências técnicas**, com apresentação de documentações de versões diferentes de software, informações contraditórias e ausência de precisão técnica.

Tal configuração caracteriza **proposta não firme, em clara violação aos princípios da segurança jurídica e vinculação ao edital**, impossibilitando a adequada avaliação técnica e comprometendo a execução contratual.

#### IV. DOS PEDIDOS

Diante do exposto e considerando:

- a) As **graves e múltiplas inconsistências técnicas** identificadas na proposta da empresa COMPWIRE INFORMÁTICA LTDA.;
- b) O **descumprimento** aos requisitos editalícios fundamentais;
- c) A **violação aos princípios fundamentais** que regem os procedimentos licitatórios;
- d) A **impossibilidade técnica** de execução adequada do objeto contratual;
- e) **O risco iminente ao interesse público e à segurança dos dados da CESAN;**

**REQUER-SE** que seja **ACOLHIDO** o presente recurso administrativo para:

1. **REFORMAR** a decisão do Sr. Pregoeiro que declarou vencedora a empresa COMPWIRE INFORMÁTICA LTDA.;
2. **DECLASSIFICAR** definitivamente a referida empresa do certame, em razão das irregularidades técnicas e descumprimentos editalícios demonstrados;
3. **PROSEGUIR** com o certame na ordem de classificação, convocando-se a próxima empresa classificada para apresentação da documentação de habilitação;
4. Subsidiariamente, caso entenda necessário, **ANULAR** o procedimento licitatório para correção das irregularidades identificadas.

#### V. DAS CONSIDERAÇÕES FINAIS

O presente recurso fundamenta-se em **análise técnica rigorosa e objetiva**, demonstrando **inequivocamente** que a proposta da empresa recorrida não atende






aos requisitos editalícios fundamentais, apresentando **graves inconsistências** que comprometem irremediavelmente sua adequação técnica e exequibilidade.

A manutenção da decisão recorrida representaria **grave violação** aos princípios da legalidade, moralidade, eficiência e interesse público, além de comprometer a segurança e a qualidade dos serviços prestados pela CESAN.

**Confia-se** no acerto da decisão de Vossa Senhoria, certo de que prevalecerá o **interesse público** e a **estrita observância aos princípios legais** que devem nortear todo procedimento licitatório.

Termos em que,  
Pede e espera deferimento.

Vila Velha/ES, 28 de agosto de 2025.

Documento assinado digitalmente  
 **EDUARDO PORTO RANGEL**  
Data: 28/08/2025 16:06:46-0300  
Verifique em <https://validar.iti.gov.br>

Eduardo Porto Rangel  
Supervisor Comercial



# Huawei OceanProtect 1.5.0

## Technical White Paper

Issue	04
Date	2024-01-19



HUAWEI TECHNOLOGIES CO., LTD.



**Copyright © Huawei Technologies Co., Ltd. 2024. 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



HUAWEI 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: <https://e.huawei.com>

# Security Declaration

## Product Lifecycle

Huawei's regulations on product lifecycle are subject to the *Product End of Life Policy*. For details about this policy, visit the following web page:

<https://support.huawei.com/ecolumnsweb/en/warranty-policy>

## Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process*. For details about this process, visit the following web page:

<https://www.huawei.com/en/psirt/vul-response-process>

For vulnerability information, enterprise customers can visit the following web page:

<https://securitybulletin.huawei.com/enterprise/en/security-advisory>

## Preconfigured Digital Certificate

The digital certificates preconfigured on Huawei devices are subject to the *Rights and Responsibilities of Preconfigured Digital Certificates on Huawei Devices*. For details about this document, visit the following web page:

<https://support.huawei.com/enterprise/en/bulletins-service/ENEWS2000015789>

## Huawei Enterprise End User License Agreement

This agreement is the end user license agreement between you (an individual, company, or any other entity) and Huawei for the use of the Huawei Software. Your use of the Huawei Software will be deemed as your acceptance of the terms mentioned in this agreement. For details about this agreement, visit the following web page:

<https://e.huawei.com/en/about/eula>

## Lifecycle of Product Documentation

Huawei after-sales user documentation is subject to the *Product Documentation Lifecycle Policy*. For details about this policy, visit the following web page:

<https://support.huawei.com/enterprise/en/bulletins-website/ENEWS2000017761>

# Contents

**1 Abstract..... 1**

**2 Product Overview..... 2**

2.1 Data Protection Appliance Mode..... 3

2.1.1 Supported Models and Specifications..... 3

2.1.2 Feature Description..... 4

2.1.3 Customer Benefits..... 7

2.1.4 Network Planning..... 8

2.1.4.1 Network Plane Division..... 8

2.1.4.2 Configuration Principles for Network Interface Card (NIC)..... 10

2.1.4.3 Typical Network Connection and IP Address Configuration..... 11

2.1.4.3.1 Backup Networking..... 11

2.1.4.3.2 Backup + Archiving Networking..... 13

2.1.4.3.3 Backup + Replication Networking..... 14

**3 Hardware Architecture..... 17**

3.1 Hardware..... 17

3.1.1 OceanProtect X8000..... 18

3.1.2 OceanProtect X6000..... 19

3.1.3 SAS Disk Enclosures..... 21

3.1.3.1 2 U 2.5-Inch SAS Disk Enclosures..... 21

3.1.3.2 4 U 3.5-Inch SAS Disk Enclosures..... 22

3.1.4 Power Consumption and Heat Dissipation..... 23

3.2 Fully Interconnected Controllers Across Controller Enclosures..... 24

3.3 High Availability of Management Nodes in Multiple Clusters..... 25

**4 Storage System Software Design..... 26**

4.1 Software Architecture..... 26

4.1.1 SAN+NAS Parallel Architecture..... 26

4.1.1.1 Active-Active Logical Architecture for SAN..... 27

4.1.1.1.1 Global Load Balancing..... 28

4.1.1.2 Distributed Logical Architecture for NAS..... 28

4.1.1.2.1 NAS Protocols..... 28

4.1.1.3 RAID 2.0+..... 32

**5 Data Backup Software Design..... 33**

Huawei OceanProtect Technical White Paper	Contents
5.1 Software Architecture.....	33
5.1.1 Backup Software Architecture.....	33
5.2 Highlights.....	34
5.2.1 Forever Incremental Backup (Synthetic Full Backup).....	34
5.2.2 Backup in Native Format.....	36
5.2.3 Instant Availability of Copies.....	38
5.2.4 E2E Ransomware Protection.....	40
5.2.4.1 Encrypted Transmission.....	41
5.2.4.2 Encrypted Storage.....	42
5.2.4.3 WORM.....	44
5.2.4.4 Ransomware Detection.....	44
5.2.5 Global Search.....	49
5.3 Protection Ecosystem for Production Systems.....	51
5.3.1 Host/NAS Backup.....	51
5.3.2 Database Ecosystem Backup.....	52
5.3.3 Virtualization, Cloud, and Container Ecosystem Backup.....	53
5.3.4 Big Data Ecosystem Backup.....	54
5.3.5 Data Warehouse Backup.....	56
5.4 Copy Lifecycle Management.....	57
5.4.1 Copy Archiving.....	57
5.4.1.1 Cloud Archiving.....	57
5.4.1.2 Tape Archiving.....	59
5.4.2 Copy Replication.....	60
5.5 Copy Data Anonymization.....	61
5.5.1 Data Anonymization.....	61
<b>6 System Performance Design.....</b>	<b>64</b>
6.1 Front-end Network Optimization.....	66
6.2 Intra-Controller Optimization.....	67
6.2.1 Intelligent Multi-Core Technology.....	67
6.2.1.1 vNode Processing Domain.....	68
6.2.1.2 Lock-free Design Between Cores.....	68
6.2.1.3 Intelligent Dynamic Load Balancing of CPUs.....	69
6.2.2 Huawei-developed Efficient Fingerprint Algorithm.....	71
6.3 Back-end Network Optimization.....	71
6.3.1 Multistreaming.....	71
6.3.2 Large-Block Sequential Write.....	74
6.4 Backup Software Performance Optimization.....	77
6.4.1 Data Passthrough to Storage.....	77
6.4.2 Distributed Concurrent Stream Backup.....	77
6.5 Backup Media Performance Optimization.....	78
6.5.1 Optimization for Backup and Recovery Jobs.....	78
6.5.2 Recovery Performance Improvement.....	81
Issue 04 (2024-01-19)	iv

6.6 Backup Performance Monitoring..... 82

**7 System Reliability Design..... 83**

7.1 Data Reliability Design.....83

7.1.1 Cache Data Reliability.....84

7.1.1.1 Written Data Mirroring..... 84

7.1.1.2 Power Failure Protection..... 85

7.1.2 Persistent Data Reliability..... 85

7.1.2.1 Intra-disk RAID..... 85

7.1.2.2 RAID 2.0+..... 86

7.1.2.2.1 RAID for Disk Redundancy..... 86

7.1.2.3 Dynamic Reconstruction..... 87

7.1.2.4 Background Data Consistency Scanning..... 87

7.1.3 Data Reliability on I/O Paths..... 87

7.1.3.1 End-to-end PI..... 88

7.1.3.2 Matrix Verification..... 88

7.1.4 Automatic Cross-site Data Repair..... 89

7.2 Service Availability..... 89

7.2.1 Interface Module and Link Redundancy Protection..... 90

7.2.2 Controller Redundancy..... 90

7.2.3 Storage Media Redundancy..... 91

7.2.3.1 Fast Isolation of Disk Faults..... 91

7.2.3.2 Disk Redundancy..... 91

**8 Data Reduction (SmartDedupe and SmartCompression).....93**

8.1 Data Preprocessing..... 94

8.2 Deduplication..... 94

8.2.1 Source Deduplication..... 95

8.2.1.1 SourceDedupe Client — DataTurbo..... 95

8.2.1.2 Application-Aware SourceDedupe..... 96

8.2.2 Variable-Length Deduplication..... 96

8.2.2.1 Intelligent Multi-Layer Variable-Length Chunking..... 97

8.2.2.2 Deduplication Principles..... 98

8.3 Deduplicated Replication..... 100

8.4 Compression..... 101

8.4.1 Compression After Combination..... 101

8.4.2 Compression (SmartCompression) Process..... 102

8.4.3 Data Compaction..... 103

8.4.4 Data Rearrangement..... 104

**9 System Security Design..... 105**

9.1 Overall Security Architecture..... 105

9.2 Security Capabilities..... 106

**10 System Serviceability Design..... 107**

10.1 Storage System Management.....	107
10.1.1 DeviceManager.....	107
10.1.1.1 Storage Space Management.....	108
10.1.1.1.1 Flexible Storage Pool Management.....	108
10.1.1.2 Configuration Task.....	108
10.1.1.3 Fault Management.....	109
10.1.1.3.1 Monitoring Status of Hardware Devices.....	109
10.1.1.3.2 Alarm and Event Monitoring.....	110
10.1.1.4 Performance and Capacity Management.....	110
10.1.1.4.1 Built-In Performance Data Collection and Analysis Capabilities.....	111
10.1.1.4.2 Independent Data Storage Space.....	111
10.1.1.4.3 Capacity Prediction.....	112
10.1.1.4.4 Performance Threshold Alarm.....	112
10.1.1.4.5 Scheduled Report.....	112
10.1.2 CLI.....	113
10.1.3 RESTful APIs.....	113
10.1.4 SNMP.....	113
10.1.5 SMI-S.....	113
10.1.6 Tools.....	114
10.2 Backup System Management.....	114
10.2.1 OceanProtect GUI.....	114
10.2.2 RESTful APIs.....	115
10.2.3 Centralized Management of Multiple Devices.....	115
10.2.4 Client push installation in batches.....	117
10.3 Intelligent Cloud Management.....	118
10.3.1 Scope of Information to Be Collected.....	119
10.3.2 Intelligent Fault Reporting.....	119
10.3.3 Capacity Prediction.....	120
10.3.4 Disk Health Prediction.....	121
10.3.5 Device Health Evaluation.....	123
10.3.6 Performance Fluctuation Analysis.....	124
10.3.7 Performance Exception Detection.....	125
10.3.8 Performance Bottleneck Analysis.....	126
10.4 OceanProtect Appliance Upgrade.....	127
10.4.1 Upgrading Storage System.....	127
10.4.2 Upgrading Backup Software.....	128
10.4.3 Upgrading Backup Clients in Batches.....	129
<b>11 Acronyms and Abbreviations.....</b>	<b>130</b>

# 1 Abstract

As the next-generation intelligent all-flash storage benchmark, the Huawei OceanProtect backup system (OceanProtect for short) is designed to back up data of mission-critical services in data centers of large enterprises in especially financial and manufacturing industries.

In today's digital era, IT construction is facing huge challenges. The challenges can be analyzed from two aspects: disaster recovery (DR) center and big data center.

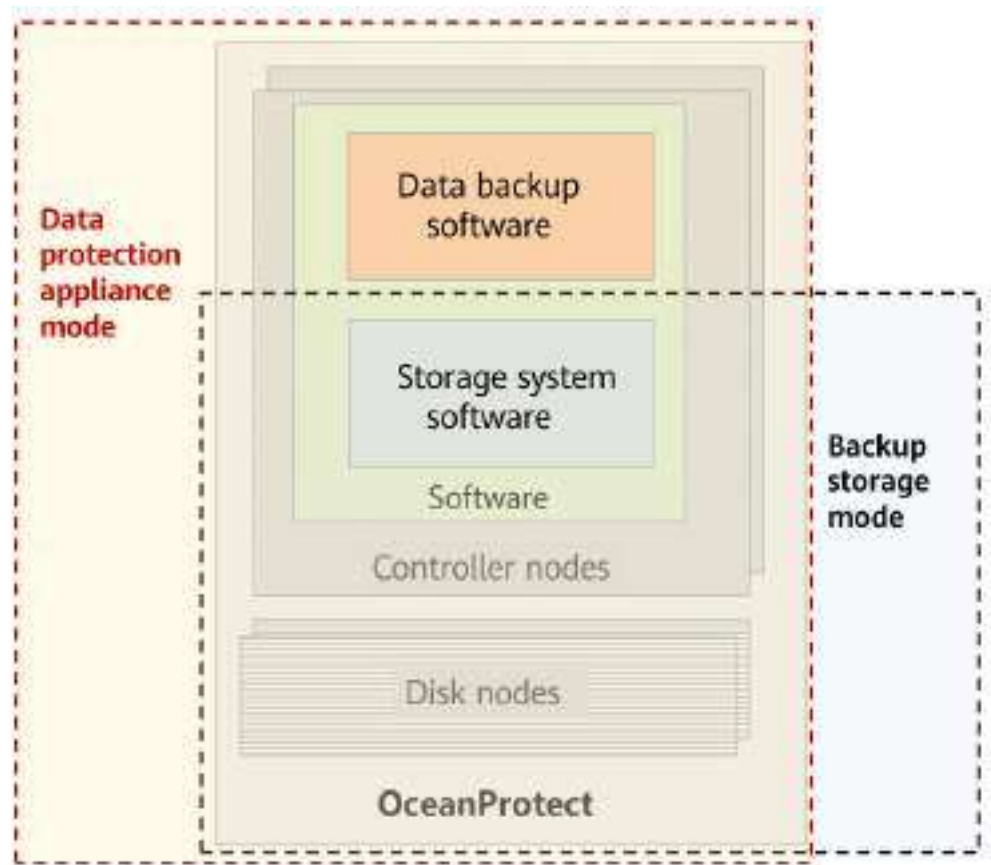
- DR center
  - Backup data is used only for data recovery, resulting in a low return on investment (ROI) of DR construction.
  - Data protection efficiency is low.
  - Verification of data recovery is difficult.
- Big data center
  - It takes a long time to collect and back up massive data, increasing the impact of backup failures.
  - The nearline data storage duration is prolonged, the amount of backup data increases sharply, and management becomes more difficult.
  - Data recovery takes a long time, and service recovery from interruption takes more time.

Based on end-to-end acceleration and the active-active high-reliability architecture, Huawei OceanProtect features rapid backup, rapid recovery, efficient reduction, and high reliability. With the fastest recovery speed, it can help users achieve efficient backup and recovery and greatly reduce the TCO. It is widely used in industries such as government, finance, carrier, healthcare, and manufacturing.

This document describes and highlights the unique advantages of OceanProtect in terms of the product positioning, hardware and software architectures, and features.

## 2 Product Overview

In terms of application scenarios, OceanProtect provides backup storage and data protection appliance modes. The two modes cannot be used together.



**Backup storage mode:** Only the storage system is included (without backup software installed). This mode applies to the scenario where a backup software system is available and only a high-performance and large-capacity storage device is required.

**Data protection appliance mode:** In addition to the storage system, the data backup software system is deployed in container mode. Storage system software and data backup software share controller hardware resources. The backup software system protects and manages connected applications. This mode applies

to the scenario where a complete data protection solution is required for the production system.

2.1 Data Protection Appliance Mode

2.1 Data Protection Appliance Mode

Compared with the backup storage mode, the data protection appliance mode applies to two hardware series: X6000 and X8000.

2.1.1 Supported Models and Specifications

Table 2-1 Supported models

Item	OceanProtect X6000 (All-Flash)	OceanProtect X6000 (HDD)	OceanProtect X8000 (All-Flash)	OceanProtect X8000 (HDD)
System architecture	2 U controller enclosure with integrated disks (25 x 2.5-inch disks)	2 U controller enclosure with integrated disks (25 x 2.5-inch disks)	2 U controller enclosure with integrated disks (25 x 2.5-inch disks)	2 U controller enclosure with integrated disks (25 x 2.5-inch disks)
CPU model	Kunpeng 920, 64-core 2.6 GHz	Kunpeng 920, 64-core 2.6 GHz	Kunpeng 920, 64-core 2.6 GHz	Kunpeng 920, 64-core 2.6 GHz
Memory per controller	256 GB (16 x 16 GB)	256 GB (16 x 16 GB)	512 GB (16 x 32 GB)	512 GB (16 x 32 GB)
Mirror channel per controller enclosure	2 x 100 Gbit/s RDMA	2 x 100 Gbit/s RDMA	2 x 100 Gbit/s RDMA	2 x 100 Gbit/s RDMA
Maximum number of controllers per controller enclosure	2	2	2	2
Data disk type	SSD	NL-SAS HDD	SSD	NL-SAS HDD
Capacity per data disk	3.84 TB/7.68 TB	4 TB/6 TB/8 TB/10 TB/14 TB	3.84 TB/7.68 TB	4 TB/6 TB/8 TB/10 TB/14 TB

Item	OceanProtect X6000 (All-Flash)	OceanProtect X6000 (HDD)	OceanProtect X8000 (All-Flash)	OceanProtect X8000 (HDD)
Minimum number of disks	8	HDD: 8 SSD: 6 (SSDs are mandatory for metadata cache.)	25	HDD: 24 SSD: 6 (SSDs are mandatory for metadata cache.)
Maximum number of disks	50	HDD: 96	175	168 HDDs + 50 SSDs
Available capacity per controller enclosure	16 TB to 300 TB	16 TB to 300 TB	150 TB to 1 PB	150 TB to 1 PB
Maximum logical capacity	Verification specifications: 6 PB/controller enclosure Maximum specifications: 21.6 PB/controller enclosure	Verification specifications: 6 PB/controller enclosure Maximum specifications: 21.6 PB/controller enclosure	Verification specifications: 16 PB/controller enclosure Maximum specifications: 72 PB/controller enclosure	Verification specifications: 16 PB/controller enclosure Maximum specifications: 72 PB/controller enclosure
Maximum backup bandwidth per controller enclosure	10.8 TB/hour	10.8 TB/hour	20 TB/hour	20 TB/hour
Data backup features supported	All features	All features	All features	All features

### 2.1.2 Feature Description

As the IT industry develops, data assets have become fundamental to the production, operation, and strategy making of enterprises. For the data protection of IT systems, the repurposing of backup data is crucial to maximizing the data value. Enterprise backup is facing the following challenges:

- Complex services, diversified applications, and large amount of data to be backed up in the production environment
- Time-consuming service recovery and long-time service interruption

- Difficult to manage backup data and accurately search for data to be recovered
- Low return on investment (ROI) of backup systems, where backup data cannot be used to support other services for higher system utilization
- Difficult to predict the storage capacity required by backup data and unable to predict potential risks due to low reliability of backup systems

Business developments and technical innovations pose higher requirements on the construction and protection of customers' IT infrastructure, and choosing an efficient and reliable data protection system with extensive backup data repurposing capabilities is a crucial part in building a highly reliable and modern IT infrastructure.

Adhering to the ultimate protection and efficient utilization design, Huawei OceanProtect data backup features leverage the core technical capabilities of Huawei-developed backup software to implement native backup capabilities and provide stable, high-performance backup and ultimate reliable services, meeting the requirements of modern intelligent IT systems for backup systems. In addition, live mounts of copies, service migration back, global indexing, data anonymization, and data repurposing improve enterprises' ROI on backup systems.

Huawei OceanProtect data backup features are next-generation high-performance data protection product capabilities for mission-critical applications of financial institutions, carriers, and public safety that integrate backup software, backup servers, backup storage, backup data management, and data security capabilities, providing customers with efficient, flexible, and secure data protection capabilities.

- **Backup capability ecosystem compatibility**
  - Host backup: supports backup of filesets on physical machines and VMs.
  - NAS file backup: compatible with various vendors
  - Database backup: supports backup of multiple databases, such as Oracle, SQL Server, MySQL, MariaDB, DB2, SAP Oracle, SAP HANA, Exchange, GaussDB T, openGauss, Dameng, PostgreSQL, Enmotech (openGauss), and VAST Data (openGauss).
  - Big data backup: supports backup of HDFS, HBase, Hive, Elasticsearch, Redis and ClickHouse on Cloudera CDH/CDP big data platform, open-source Hadoop HDFS, Huawei FusionInsight, and MapReduce Service.
  - Virtualization/Cloud/Container backup: supports virtualization and cloud platform data backup, such as VMware, Hyper-V, RHV (Red Hat Virtualization), FusionCompute, Huawei Cloud Stack, and Kubernetes +FlexVolume.
  - Huawei data warehouse backup: supports the backup of Huawei Data Warehouse Service (DWS).
- **Concurrent data flows and high backup bandwidth**
  - Concurrent transmission: supports concurrent transmission of backup data on dual controllers. Backup transmission network channels are selected based on application and controller enclosure loads to maximize the network transmission capability. The backup bandwidth of a single controller enclosure can reach 20 TB/hour.
  - Linear expansion: supports expansion to a maximum of 32 controller enclosures.

- **Dynamic capacity expansion to provide large storage capacity**
  - Ultra-large capacity: The maximum usable capacity of a single controller enclosure is 1 PB.
  - The storage capacity of disks and disk enclosures can be scaled out online and automatically included in the backup storage space to meet the storage requirements of massive backup data.
- **Flexible copy mounting**
  - Backup copies can be mounted in real time to recover and take over production services within seconds, ensuring the service connectivity. After the production system recovers, services can be migrated back without any service interruption or data loss.
  - Mounted copies are automatically updated based on the mount update policies, reducing manual operations and greatly improving the efficiency of data analysis and automatic development and test.
  - The all-SSD system provides up to 100,000 IOPS @ 1 ms, ensuring efficient read and write of mounted copy data.
- **Global ultra-fast data search**
  - Global data search: Protected resources and copy data can be globally searched at file-level based on copy data content, and file-level recovery simplifies operations and makes data recovery more accurate.
  - The ultra-large index capacity supports storage of tens of billions of data records and retrieval within seconds, greatly improving the recovery efficiency.
- **Database anonymization and ransomware detection, ensuring cyber resilience**
  - Data anonymization for production databases and backup copies: Anonymized copy data can be used for testing and auditing, ensuring data security.
  - Flexible data anonymization policies: Various built-in data anonymization policy templates allow you to replace copy data on demand, meeting various security regulations.
  - Ransomware detection: Ransomware features are scanned in multiple copies based on machine learning, and copies suspected of being infected by ransomware will be marked to ensure copy security.
  - Compliance WORM is supported, preventing copy tampering and ensuring copy storage security.
  - End-to-end (E2E) encryption: Data transmission encryption and data storage encryption are supported.

Huawei OceanProtect data backup features meet the key protection requirements of enterprise-level applications such as databases, VMs, physical servers, files, and big data. One set of OceanProtect backup storage device protects multiple production systems and supports linear expansion of capacity and performance, helping finance, carrier, medical care, manufacturing, and other industries improve data protection efficiency, reduce data protection investment, and simplify the management process.

## 2.1.3 Customer Benefits

Based on end-to-end acceleration and the active-active high-reliability architecture, OceanProtect backup storage features rapid backup, rapid recovery, efficient reduction, and high reliability. With the fastest recovery speed, it can help users achieve efficient backup and recovery and greatly reduce the TCO. It is widely used in industries such as government, finance, carrier, healthcare, and manufacturing. It has the following advantages:

### Open Ecosystem and Flexible Application

Built-in data protection software for big data, databases, and virtualization provides core protection capabilities such as rapid backup, forever incremental backup, and rapid restoration, as well as the ransomware protection solution. In addition, the open ecosystem enables integration with mainstream backup software. All-scenario data protection solutions are provided.

### Rapid Backup, Rapid Recovery

Flash-based storage media deliver high performance.

The whole process is accelerated. The front end offloads the network protocol stack to the NIC to release CPU resources. The back-end CPU multi-core parallel scheduling enables dedicated cores through grouping and task partitioning, improving the node processing capability.

Based on backup service characteristics, multiple sequential data flows are aggregated to significantly improve bandwidth performance.

Mainstream backup software can be used to support instant recovery and reuse of data.

### Efficient Reduction

Industry-leading algorithms are adopted to identify data flow features for precise chunking. Inline variable-length deduplication, adaptive compression, and byte-level compaction are used to improve effective capacity and reduce the TCO.

In backup data preprocessing, backup data is aggregated and then deduplicated to improve the data reduction ratio.

The system investment is significantly reduced through on-demand expansion, small-scale initial configuration, and small incremental steps.

### High Reliability

The active-active hardware architecture is adopted. Therefore, if a single controller is faulty, running backup jobs can be switched over to the functioning controller within seconds without being interrupted.

RAID 5, RAID 6, and RAID-TP are supported. Simultaneous failures of a maximum of three disks can be tolerated.

Silent data consistency check is used to ensure data integrity and validity.

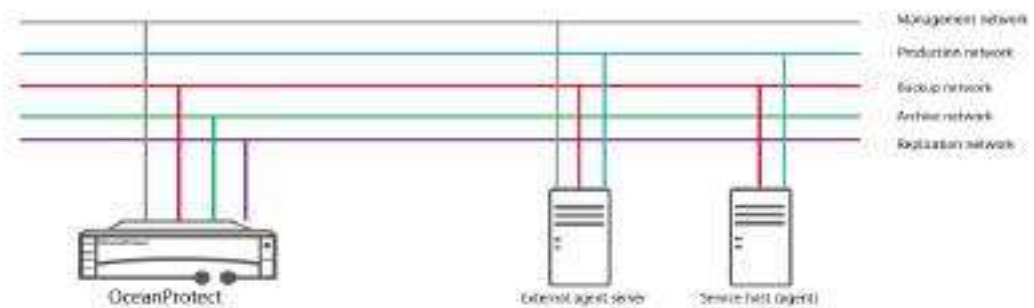
Proactive O&M, intelligent prediction, timely notification, and mobile O&M are available.

## 2.1.4 Network Planning

To enable the OceanProtect data backup capability, you need to configure at least one management IP address and at least four backup network plane IP addresses for a single node which are used to allocate IP addresses to internal component containers. During backup, the backup and recovery signaling is mainly transmitted. This section describes network plane division and network planning in typical scenarios after the data backup feature is enabled.

### 2.1.4.1 Network Plane Division

**Figure 2-1** OceanProtect network plane division



When OceanProtect data backup is enabled, there are logically management, production, backup, archive, and replication networks.

Management network is mainly used to transmit a small amount of management data, such as device management and service configuration data.

Production network transmits production data on the customers live network.

Backup network is an independent network used to transmit backup data in the customer's environment.

Archive network is a data transmission network where archive storage media resides.

Replication network is a private network or private line dedicated for data transmission between multiple centers.

Devices displayed in the figure:

OceanProtect: primary data protection device

External agent server: In VMware, HCS, Kubernetes+FlexVolume, big data, and DWS scenarios, the OceanProtect data backup feature requires additional agent servers.

Service host: To back up database or files, you need to install the backup agent software package on the service host where the database or files to be backed up reside.

OceanProtect network access:

Network Plane	Description
Management network	Used for OceanProtect backup service configuration, device management, and interconnection in cloud environment backup where the management network is involved. BMC device management Backup service network configuration Interconnection with vCenter for VMware backup Interconnection with the storage device management network for NAS backup
Backup network	Transmits signaling and data during backup and recovery and controls signaling transmission during replication.
Replication network	Transmits data during replication.
Archive network	Transmits signaling and data during archiving.

External agent server (Proxy) network access:

Network Plane	Description
Management network	The external agent server is a necessary expansion server independently configured by OceanProtect in VMware, HDFS/HBase/Hive, and HCS backup scenarios. The management network is mainly used for device management and backup environment interconnection. BMC device management Interconnection with vCenter for VMware backup, with NameNode for HDFS/HBase/Hive backup, and with ManageOne for HCS backup
Production network	Reads backup data and writes recovery data.
Backup network	Writes backup data and reads recovery data.

Service host (Agent) network access:

Network Plane	Description
Backup network	Writes backup data and reads recovery data.

2.1.4.2 Configuration Principles for Network Interface Card (NIC)

Figure 2-2 Configuration Principles for NIC



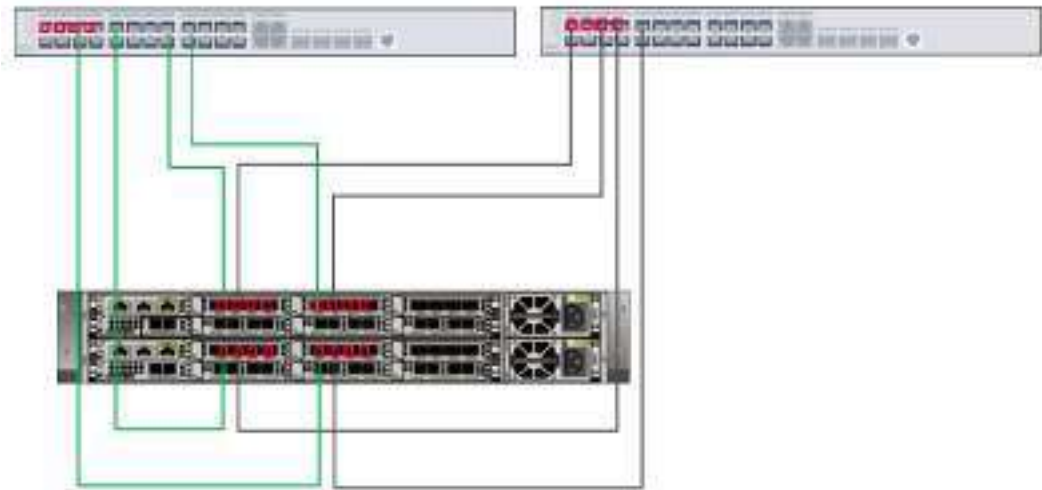
Interface Module	Configuration Requirements	Description
0#	Mandatory	Ports for transmitting backup data and replication data. 10GE/25GE optical ports and 10GE electrical interface modules are supported.
1#	Mandatory	Built-in container communication port for transmitting signaling during backup and recovery and transferring archived data. 10GE/25GE optical ports and 10GE electrical interface modules are supported.
2#	Optional, configured in four-controller scenarios.	Four controllers on a single device, cross-controller data forwarding, direct connection between controllers, and scale-out interface modules.
3#	Mandatory	When built-in containers access the storage pool, this module is used for data exchange between cards. No switch is required. Only 25GE RoCE is supported.

Interface Module	Configuration Requirements	Description
4#	Optional, configured in disk-to-disk-to-tape (D2D2T)/ encrypted replication scenarios	When backup data is archived to a tape library, read and write operations are performed on the tape library. 8 Gbit/s, 16 Gbit/s, and 32 Gbit/s Fibre Channel ports are supported.  Configured when the backup data is replicated to other OceanProtect devices and the replication encryption function is used. The Hi1823 encryption module is supported.  Note: D2D2T and encrypted replication cannot be configured on a device at the same time.
5#	Optional, configured in large-capacity scenarios	Configured when there are more than four disk enclosures to read and write data in added disk enclosures. 12 Gbit/s SAS ports are supported.

### 2.1.4.3 Typical Network Connection and IP Address Configuration

#### 2.1.4.3.1 Backup Networking

Figure 2-3 Backup networking

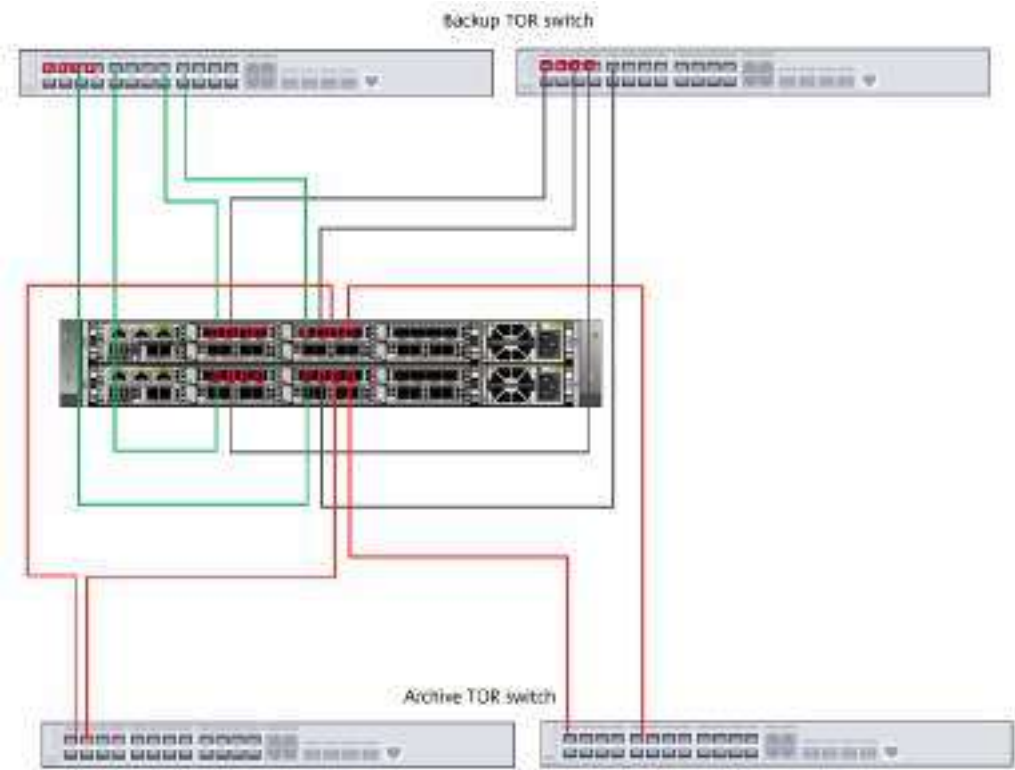


※ For the same device, the cable connection and configuration of controller B must be consistent with those of controller A.

Expansion Module	Port	Connection	IP Address Planning	Description
0#	P0	Mandatory	It is recommended that the P0 and P1 be bonded and share the same IP address.	Configured when the network bandwidth does not meet backup performance requirements.
	P1	Mandatory	It is recommended that the P1 and P0 be bonded and share the same IP address.	
	P2	Optional	It is recommended that the P2 and P3 be bonded and share the same IP address.	
	P3	Optional	It is recommended that the P3 and P2 be bonded and share the same IP address.	
1#	P0	Mandatory	No bond port is required. The backup software automatically forms a failover group consisting of this port and P1. Two IP addresses are shared.	※ This interface module is used for control message communication between internal containers and external backup clients. You need to configure one IP address for each of the two internal containers.
	P1	Mandatory	No bond port is required. The backup software automatically forms a failover group consisting of this port and P0. Two IP addresses are shared.	
	P2	No required		
	P3	No required		

2.1.4.3.2 Backup + Archiving Networking

Figure 2-4 Backup + archive network connection



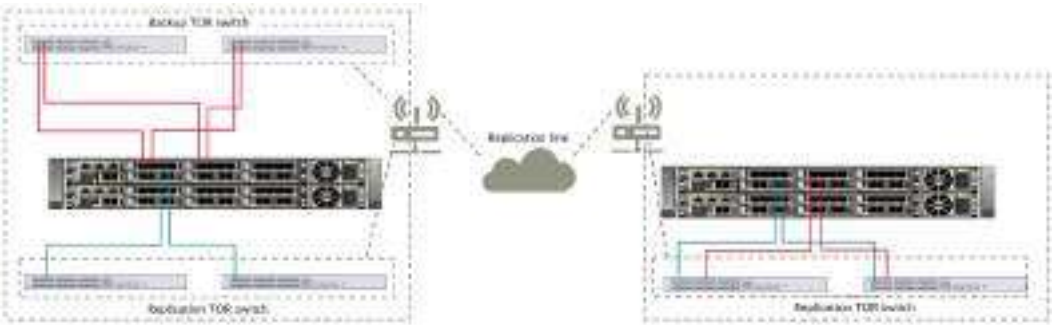
※ For the same device, the cable connection and configuration of controller B must be consistent with those of controller A.

Expansion Module	Port	Connection	IP Address Planning	Description
0#	P0	Mandatory	It is recommended that the P0 and P1 be bonded and share the same IP address.	
	P1	Mandatory	It is recommended that the P1 and P0 be bonded and share the same IP address.	
	P2	Optional	It is recommended that the P2 and P3 be bonded and share the same IP address.	Configured when the network bandwidth does not meet backup performance requirements.
	P3	Optional	It is recommended that the P3 and P2 be bonded and share the same IP address.	

Expansion Module	Port	Connection	IP Address Planning	Description
1#	P0	Mandatory	No bond port is required. The backup software forms a port failover group consisting of the port and P1. Two IP addresses are shared.	※ This interface module is used for control message communication between internal containers and external backup clients. You need to configure one IP address for each of the two internal containers.
	P1	Mandatory	No bond port is required. The backup software forms a port failover group consisting of the port and P0. Two IP addresses are shared.	
	P2	Mandatory	No bond port is required. The backup software forms a port failover group consisting of the port and P3. One IP address is shared.	※ If the IP address assigned to the P0/P1 port is in the same subnet as the archiving device, the P2/P3 port is not required.
	P3	Mandatory	No bond port is required. The backup software automatically forms a port failover group consisting of the port and P2. One IP address is shared.	

2.1.4.3.3 Backup + Replication Networking

Figure 2-5 Backup + replication networking



- For the same device, the cable connection and configuration of controller B must be the same as those of controller A.
- The container IP addresses (IP addresses of ports P0/P1 in 1# at the primary and secondary ends) at both ends must be logically reachable.

Cable connection and IP address configuration on the primary end:

Expansi on Module	Port	Con necti on	IP Address Planning	Description
0#	P0	Man dato ry	It is recommended that the P0 and P1 be bonded and share the same IP address.	IP address for backing up service data
	P1	Man dato ry		
	P2	Man dato ry	It is recommended that the P2 and P3 be bonded and share the same IP address.	IP address for replicating service data
	P3	Man dato ry		
1#	P0	Man dato ry	No bond port is required. The backup software forms a port failover group consisting of the port and P1. Two IP addresses are shared.	※ This interface module is used for control message communication between internal containers and external backup clients. You need to configure one IP address for each of the two internal containers.
	P1	Man dato ry		
	P2	Opti onal		Planned based on the archiving requirements at the local end
	P3	Opti onal		

Networking and IP address configuration of the secondary end

Expansi on Module	Port	Con necti on	IP Address Planning	Description
0#	P0	Opti onal	Planned based on backup service requirements	Planned when backup or recovery is required at the secondary end
	P1	Opti onal		
	P2	Man dato ry	It is recommended that the P2 and P3 be bonded and share the same IP address.	IP address for replicating service data

Expansi on Module	Port	Con nect ion	IP Address Planning	Description
	P3	Man dato ry		
1#	P0	Man dato ry	No bond port is required. The backup software automatically forms a port failover group consisting of the port and P1. Two IP addresses are shared.	※ This interface module is used for control message communication between internal containers and peer containers. You need to configure one IP address for each of the two internal containers.
	P1	Man dato ry		
	P2	Opti onal		Planned based on the archiving requirements at the secondary end
	P3	Opti onal		

# 3 Hardware Architecture

OceanProtect employs the SmartMatrix architecture. All field replaceable units (FRUs), such as front-end interface modules, controllers, back-end interface modules, power modules, BBUs, fan modules, and disks, are redundant and protected against single points of failure. All FRUs are hot-swappable.

The RDMA high-speed network implements shared access to the global cache at a low latency.

- 3.1 Hardware
- 3.2 Fully Interconnected Controllers Across Controller Enclosures
- 3.3 High Availability of Management Nodes in Multiple Clusters

## 3.1 Hardware

OceanProtect products provide two forms. See [Table 3-1](#).

Table 3-1 OceanProtect products

Item	OceanProtect X6000 (All-Flash)	OceanProtect X6000 (HDD)	OceanProtect X8000 (All-Flash)	OceanProtect X8000 (HDD)
Node	2 U, 25 slots, disk and controller integration	2 U, 25 slots, disk and controller integration	2 U, 25 slots, disk and controller integration	2 U, 25 slots, disk and controller integration
Architecture	Active-active	Active-active	Active-active	Active-active
Disk enclosure type*	2 U 25-slot 2.5-inch SAS disk enclosure	2 U 25-slot 2.5-inch SAS disk enclosure 4 U 24-slot 3.5-inch SAS disk enclosure	2 U 25-slot 2.5-inch SAS disk enclosure	2 U 25-slot 2.5-inch SAS disk enclosure 4 U 24-slot 3.5-inch SAS disk enclosure

Item	OceanProtect X6000 (All-Flash)	OceanProtect X6000 (HDD)	OceanProtect X8000 (All-Flash)	OceanProtect X8000 (HDD)
Disk type	2.5-inch SAS SSD	2.5-inch SAS SSD 3.5-inch NL-SAS HDD	2.5-inch SAS SSD	2.5-inch SAS SSD 3.5-inch NL-SAS HDD

 NOTE

- \*: In HDD form, 2 U 2.5-inch SAS disk enclosures contain cache disks.

### 3.1.1 OceanProtect X8000

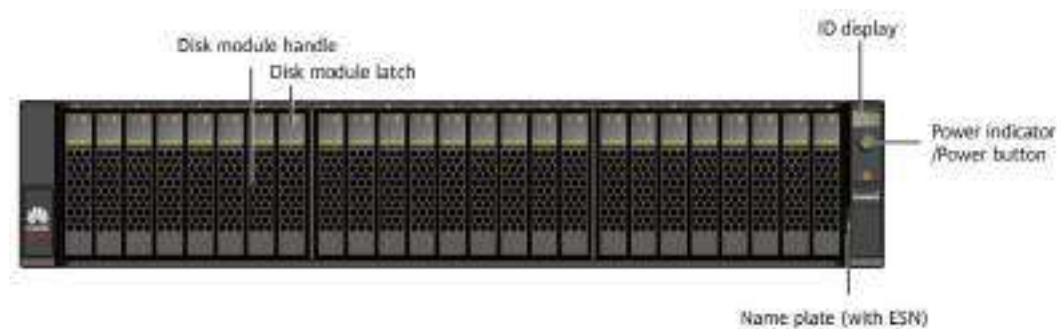
OceanProtect X8000 uses a 2 U controller enclosure that has two controllers and 25 slots. All key components are FRUs and redundant and can be replaced online. It provides unified backup storage services for SAN and NAS. It supports front-end protocols such as Fibre Channel and iSCSI for SAN services and NFS, CIFS, and NDMP for NAS services. Each controller enclosure supports up to 12 hot-swappable interface modules.

- Front end:  
4-port 8 Gbit/s, 16 Gbit/s, and 32 Gbit/s Fibre Channel interface modules, 4-port 10GE and 25GE interface modules, as well as 2-port 40GE and 100GE interface modules
- Cluster connection:  
Scale-out interface modules are 4-port 25 Gbit/s RDMA interface modules for a 4-controller direct-connection network.
- Back end:  
4-port 12 Gbit/s SAS interface modules are used.

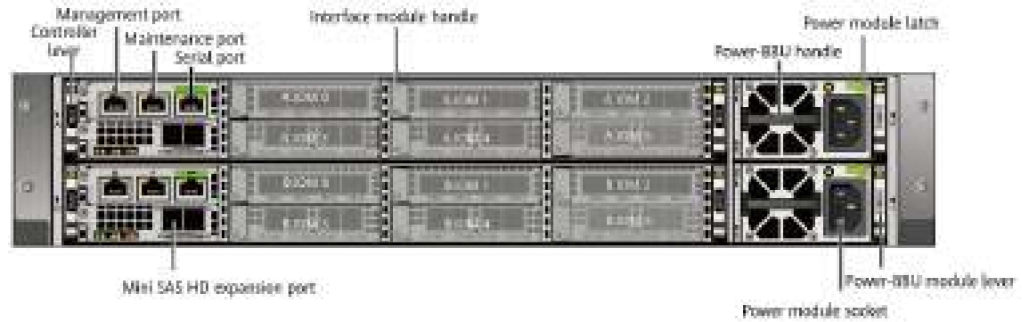
The OceanProtect mid-range device supports SAS disk enclosures.

The two controllers in a controller enclosure of the OceanProtect mid-range device are interconnected through RDMA mirror channels, and multiple controller enclosures can be directly connected through the scale-out interface modules. Each controller has two GE management and maintenance ports and one serial port. [Figure 3-1](#) and [Figure 3-2](#) show the front and rear views of the OceanProtect mid-range device.

**Figure 3-1** Front view of the mid-range device

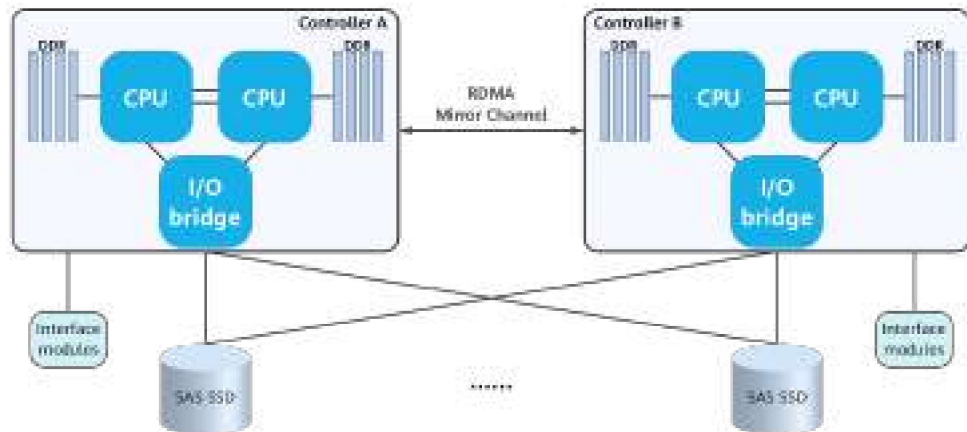


**Figure 3-2** Rear view of the mid-range device



The OceanProtect mid-range device has a 2 U controller enclosure with two controllers and integrated disks. The two controllers are symmetrically deployed in active-active mode and interconnected through RDMA mirror channels. The two controllers support service takeover upon faults and load balancing in normal cases. **Figure 3-3** shows the logical architecture.

**Figure 3-3** Logical architecture



### 3.1.2 OceanProtect X6000

OceanProtect X6000 uses a 2 U enclosure that has two controllers and a symmetric active-active architecture. 25-slot 2.5-inch SAS and 12-slot 3.5-inch SAS are supported. Key modules are FRUs and redundant and can be replaced online.

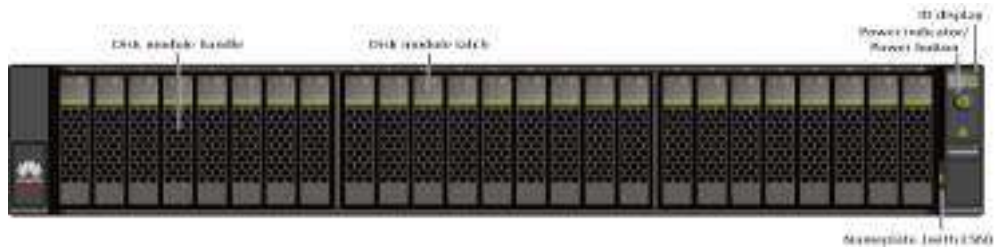
Each controller enclosure supports a maximum of 6 hot-swappable I/O modules.

- Front end:  
4-port 8 Gbit/s, 16 Gbit/s, and 32 Gbit/s Fibre Channel interface modules, 4-port 10GE and 25GE interface modules, as well as 2-port 40GE and 100GE interface modules
- Cluster connection:  
Scale-out interface modules are 4-port 25 Gbit/s RDMA interface modules.

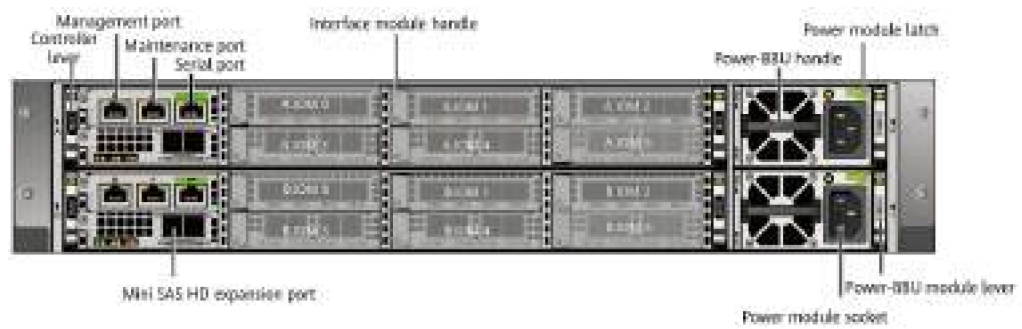
- Back end:  
Back-end interface modules are 4-port 12 Gbit/s SAS interface modules (for connecting to SAS disk enclosures).

Each controller has four onboard GE ports, four onboard 10GE ports, two GE management/maintenance ports, and one serial port. Two controllers in an enclosure are mirrored through RDMA channels. The scale-out interface module can implement directly connection of four controllers without switches. Onboard back-end SAS ports are provided, and SAS disk enclosures are supported.

**Figure 3-4** Front view of a 25-slot SAS enclosure

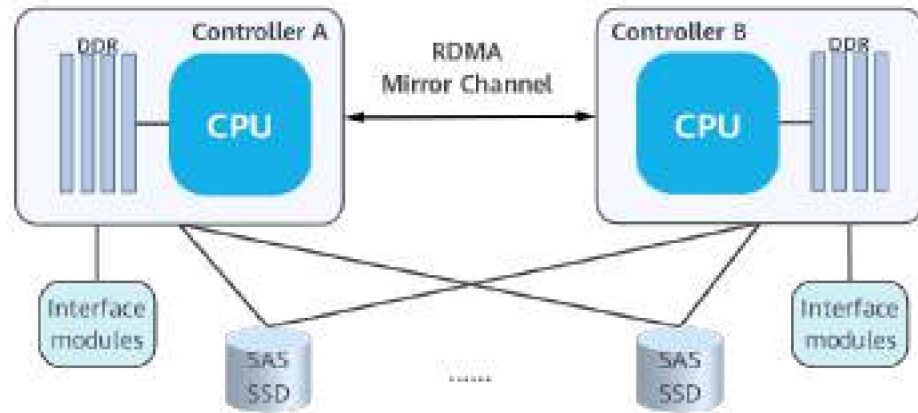


**Figure 3-5** Rear view of a 25-slot SAS enclosure



In the symmetric active-active architecture, the two controllers support service takeover upon a fault and load balancing when they are working properly. The two controllers are mirrored through RDMA channels. [Figure 3-6](#) shows the logical architecture.

**Figure 3-6** Logical architecture



### 3.1.3 SAS Disk Enclosures

The backup storage provides two types of SAS disk enclosures: 2 U 2.5-inch SAS disk enclosures and 4 U 3.5-inch SAS disk enclosures. 2 U 2.5-inch SAS disk enclosures support only 2.5-inch dual-port SAS SSDs, and do not support 3.5-inch NL-SAS HDDs. 4 U 3.5-inch SAS disk enclosures support only 3.5-inch NL-SAS HDDs, and do not support 2.5-inch SSDs.

The back-end SAS loop supports SAS disk enclosure cascading. A maximum of four SAS disk enclosures are supported. It is recommended that 2 U 2.5-inch SAS disk enclosures and 4 U 3.5-inch SAS disk enclosures use independent SAS loops to reduce the impact of mixed large and small I/Os on the latency.

#### 3.1.3.1 2 U 2.5-Inch SAS Disk Enclosures

The 2 U 2.5-inch SAS disk enclosure uses the SAS 3.0 protocol and supports 25 2.5-inch SAS SSDs. A controller enclosure connects to a SAS disk enclosure through the onboard SAS ports or SAS interface modules.

**Figure 3-7** Front view of a 2 U 2.5-inch 25-slot SAS disk enclosure

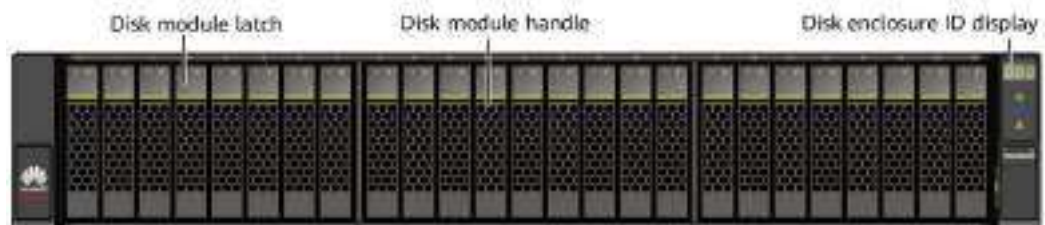
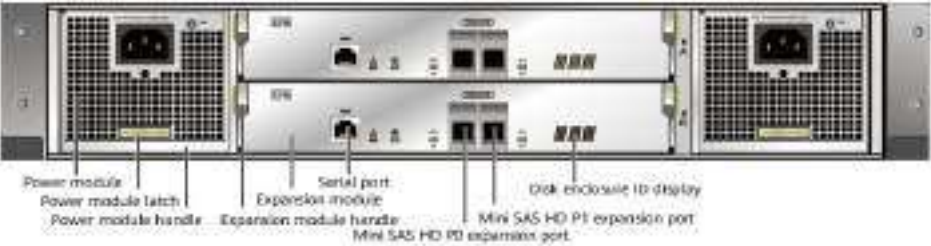


Figure 3-8 Rear view of a 2 U 2.5-inch 25-slot SAS disk enclosure



### 3.1.3.2 4 U 3.5-Inch SAS Disk Enclosures

The 4 U 3.5-inch SAS disk enclosure uses the SAS 3.0 protocol and supports 24 3.5-inch NL-SAS HDDs. A controller enclosure connects to a SAS disk enclosure through the onboard SAS ports or SAS interface modules.

Figure 3-9 Front view of a 4 U 3.5-inch 24-slot SAS disk enclosure

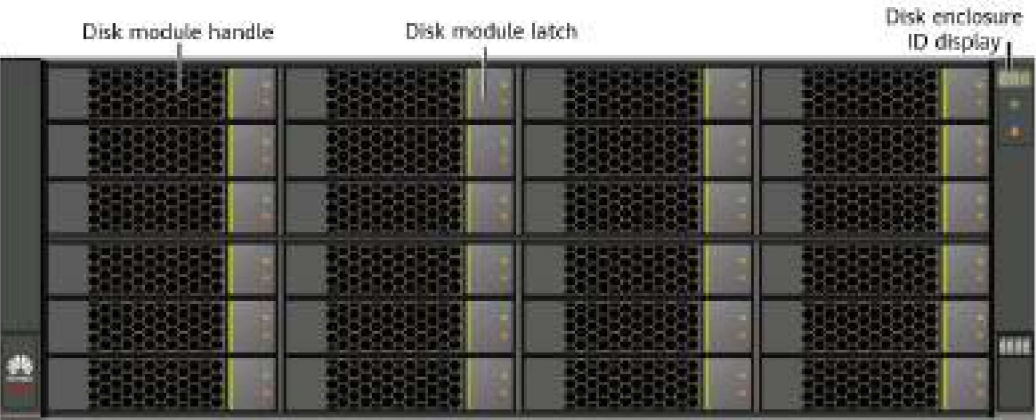
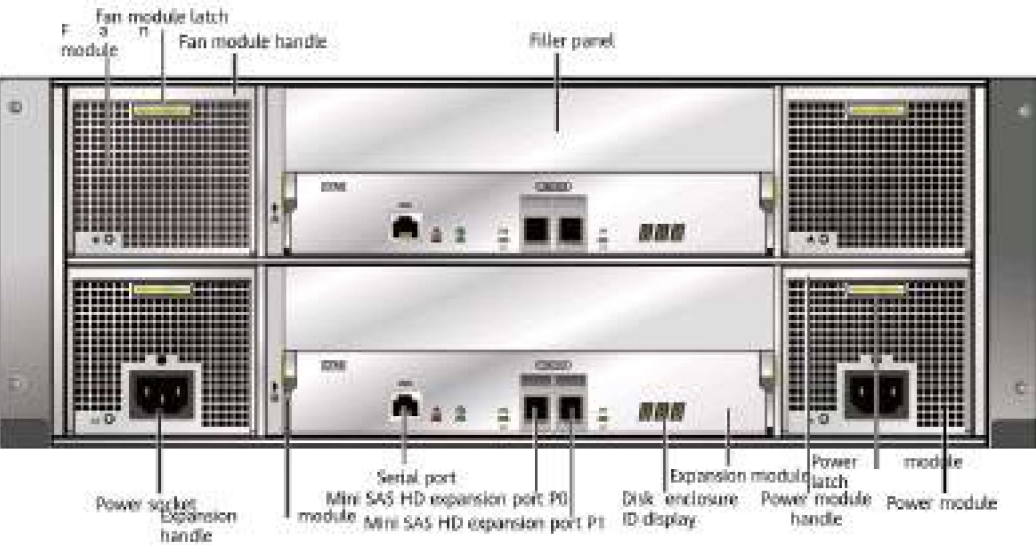


Figure 3-10 Rear view of a 4 U 3.5-inch 24-slot SAS disk enclosure



### 3.1.4 Power Consumption and Heat Dissipation

OceanProtect uses the following designs to meet the requirements for energy conservation and environment protection:

- Titanium power modules with the industry's highest conversion efficiency
- Proportional-integral-derivative fan speed adjustment algorithm for higher heat dissipation efficiency
- Staggered power-on design, avoiding peak load on the power supply

The energy-efficient design reduces power supply and heat dissipation costs.

#### Efficient Power Module

OceanProtect uses 80 Plus Platinum and Titanium power modules, which provide up to 94% power conversion efficiency and 98% power factor when the power load is 50%, reducing power loss. The Titanium power module can even reach 96% conversion efficiency, which is 2% higher than Platinum power modules, and over 90% conversion efficiency when the load is light, which is about 10% higher than common PSUs, minimizing power loss. The power modules have passed the 80 Plus certification (the certificate can be provided).

80 Plus efficiency requirements:

80 Plus Power Module Type	Power Conversion Efficiency (230 V Input)			
	10%	20%	50%	100%
80 Plus Bronze	---	81%	85%	81%
80 Plus Silver	---	85%	89%	85%
80 Plus Gold	---	88%	92%	88%
80 Plus Platinum	---	90%	94%	91%
80 Plus Titanium	90%	94%	96%	91%

#### High-Voltage DC Power Input

OceanProtect supports high-voltage direct current (HVDC) or AC/DC hybrid power input for better power supply reliability. This also reduces the UPS footprint and equipment room construction and maintenance costs. The HVDC power supply cuts down the intermediate processes of power conversion, improving the power efficiency by 15% to 25%. This saves millions of electricity fees for large data centers every year. In comparison, when low-voltage DC (12 V/48 V) is supplied to high-power devices, thick cables must be used to increase the current, which causes trouble in cable layout. This problem is solved when HVDC is used.

#### PID Fan Speed Adjustment Algorithm

OceanProtect uses the proportional integral derivative (PID) algorithm to adjust the fan speed, which solves problems such as slow response of fan speed

adjustment, high fan power consumption, great fan speed fluctuation, and loud noise. The PID algorithm allows the system to adjust the fan speed quickly, save energy, and reduce noise.

- The PID algorithm increases energy efficiency by 4% to 9% and prevents fan speed fluctuation.
- The PID algorithm increases the fan response speed by 22% to 53% and significantly reduces the noise.

## Staggered Power-On

Staggered power-on of disks prevents the electrical surge that would occur when multiple devices are powered on simultaneously, eliminating the risk on the power supply of the equipment room.

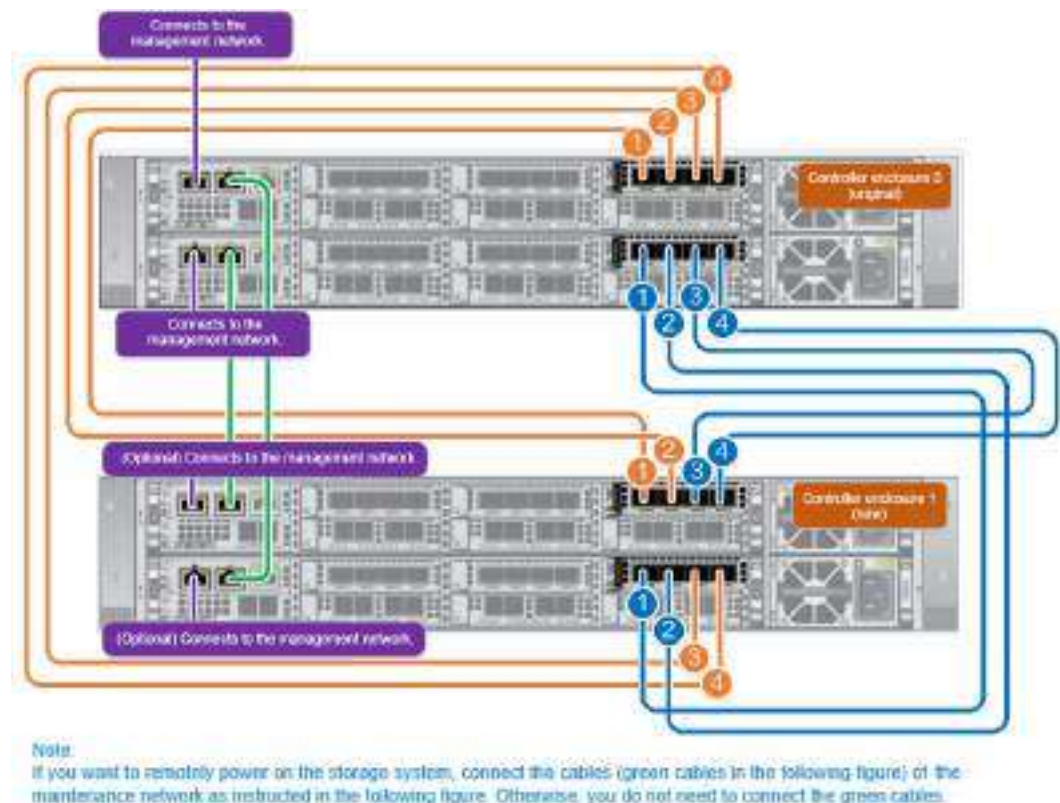
## Energy Conservation Certification

The product has passed the RoHS energy efficiency certification.

## 3.2 Fully Interconnected Controllers Across Controller Enclosures

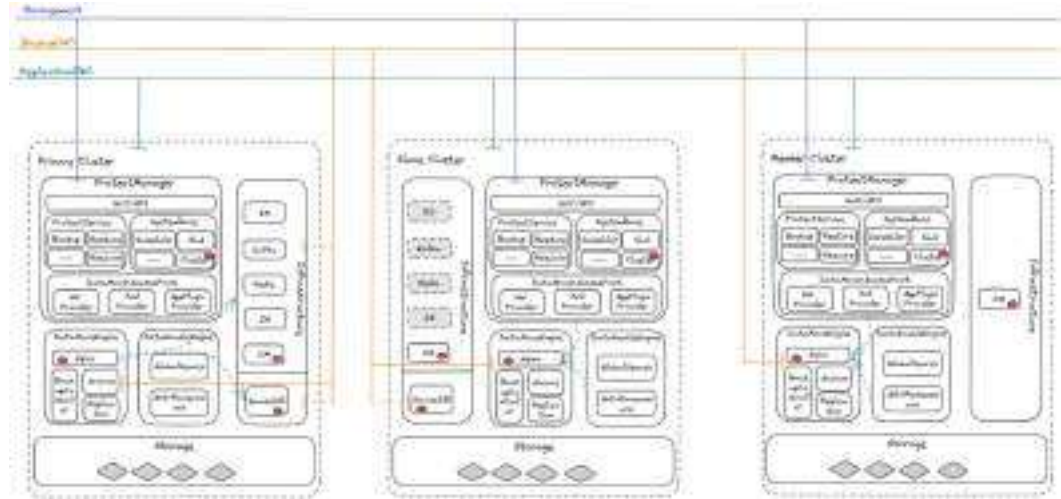
OceanProtect supports scale-out to four controllers. Each controller uses four 25 Gbit/s RDMA ports to connect to controllers of the other controller enclosure, as shown in the following figure.

**Figure 3-11** Scale-out to four controllers



### 3.3 High Availability of Management Nodes in Multiple Clusters

**Figure 3-12** HA principles of management nodes in multiple clusters



Multiple independent OceanProtect backup storage systems can be used to form a scale-out and federated backup cluster with multiple domains and K8s clusters using global management data share. Each OceanProtect backup storage system is an independent backup domain where backup services are autonomous. Physically, an OceanProtect backup storage system is a K8s cluster with multi-controller redundancy.

Cluster high availability (HA) can be deployed. GaussDB and Elasticsearch in the active cluster provide globally shared management data services and management data instances in the active and standby clusters are in active-passive (AP) mode. Management data is synchronized from the active cluster to the standby cluster so that the standby cluster can quickly take over services to prevent single points of failure if an active cluster failure occurs.

Cluster HA is implemented based on Huawei-developed high availability component OMM HA, which monitors system resources and arbitrates active/standby nodes based on resource running status, weights, and data integrity to achieve failover, breakdown switchover, dual-active, and dual-standby recovery for nodes.

# 4 Storage System Software Design

## 4.1 Software Architecture

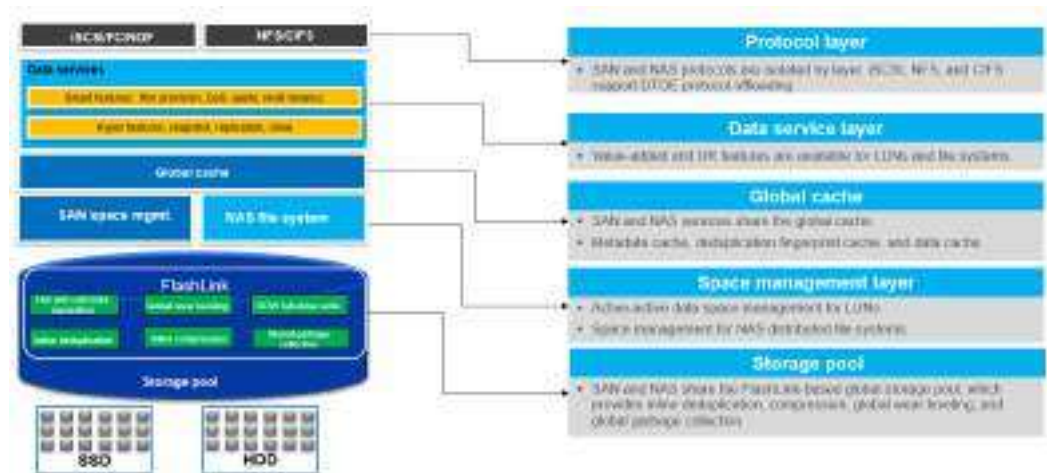
### 4.1 Software Architecture

OceanProtect uses the symmetric active-active software architecture to implement load balancing and optimizes the system based on characteristics in backup scenarios, fully utilizing backup storage system performance. All OceanProtect models use a unified software architecture and support interworking.

#### 4.1.1 SAN+NAS Parallel Architecture

OceanProtect integrates SAN and NAS. The space management software of SAN and that of NAS are at the same software layer. This solves the problem of performance differences and resource competition between LUNs and file systems when SAN is supported based on the NAS architecture or when NAS is supported based on the SAN architecture. CIFS and NFS are supported for file access and Fibre Channel and iSCSI are supported for block access. SAN and NAS share physical layer resources. Resources are allocated on demand and do not need to be reserved, simplifying space management and improving resource utilization. Both SAN and NAS support scale-out to multiple controllers. Hosts can access any LUN or file system from a front-end port on any controller.

**Figure 4-1** OceanProtect software architecture



The OceanProtect system architecture consists of the following subsystems:

- **Storage pool:** globally unifies storage pool services. It uses redirect-on-write (ROW) to allocate space for LUN and file system data, deduplicates and compresses data, and identifies and distributes metadata and user data to SSDs. It also provides fast reconstruction with RAID 2.0+ and global garbage collection in the background.
- **Space management:** allocates and reclaims space for LUNs and file systems with thin provisioning.
- **Global cache:** provides the read/write and metadata caches for LUNs and file systems.
- **Data service layer:** provides disaster recovery capabilities such as remote replication configuration for LUNs and file systems. It provides unified data replication and manages the replication configuration and networks.
- **Protocol layer:** provides protocol parsing, I/O receiving and sending, and error processing for LUNs and file systems.

Storage pools of OceanProtect directly allocate space for file systems and LUNs, which directly interact with the underlying storage pools for parallel SAN and NAS architecture.

Parallel architecture streamlines storage with the shortest I/O paths for LUNs and file systems. In addition, space management of LUNs and file systems is independent of each other, enhancing reliability.

#### 4.1.1.1 Active-Active Logical Architecture for SAN

In an asymmetrical logical unit access (ALUA) architecture, each LUN is owned by a specific controller. Customers need to plan the owning controllers of LUNs for load balancing. However, it is difficult for an ALUA architecture to implement load balancing on live networks because service pressures vary with LUNs and vary in different periods for a same LUN.

OceanProtect uses a symmetric active-active software architecture. The balancing algorithm is used to distribute the capacity of a single LUN to all controllers to balance the read and write requests received by each controller. In this way,

system performance can be fully utilized. Global cache allows that LUNs have no ownership. Each controller processes received read and write requests. (In an ALUA architecture, read and write requests must be processed by the LUN's owning controller.) This balances loads among controllers and simplifies O&M. During O&M, you only need to create LUNs that meet capacity requirements. You do not need to consider the impact of new LUN ownership on controller load balancing and whether real-time service load changes of controllers will cause the controllers to become performance bottlenecks. RAID 2.0+ evenly distributes data to all disks in a storage pool to implement load balancing.

#### 4.1.1.1 Global Load Balancing

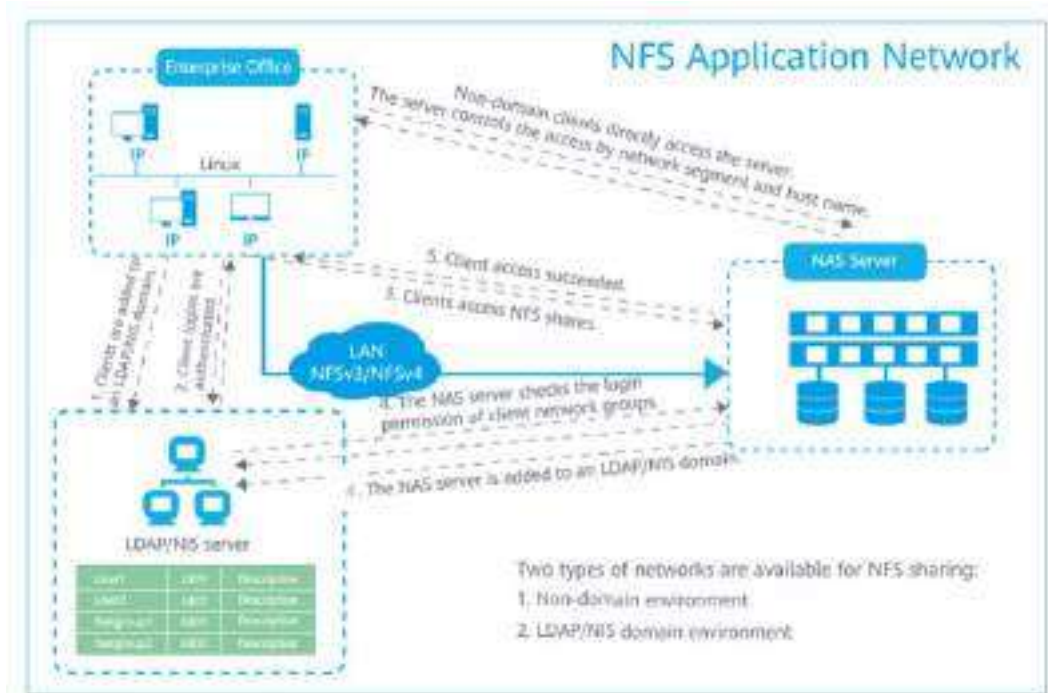
OceanProtect hashes the logical block addressing (LBA) of each host read/write request to determine the controller that processes the request. Huawei-developed multipathing software UltraPath, front-end interconnect I/O modules (FIMs), and controllers negotiate the same hash method and parameters to implement intelligent distribution of read and write requests. UltraPath and FIMs work together to directly distribute a read/write request to the optimal processing controller, avoiding forwarding between controllers. If UltraPath and FIMs are not used, after receiving a host request, a controller forwards the request to the corresponding processing controller based on the hash result of request's LBA, ensuring that host requests are balanced between controllers.

#### 4.1.1.2 Distributed Logical Architecture for NAS

OceanProtect adopts the distributed NAS architecture. When a file system is created, it is evenly distributed to different controllers. During normal system running, all directories and files in the file system are distributed among all cores of a CPU. In this way, the cross-CPU access latency is reduced, and performance is improved in scenarios such as read and write, directory traversal query, attribute traversal query, and batch attribute setting. With the global data block distribution technology of RAID 2.0+, data in each file system is distributed to all disks in the storage pool, improving the write bandwidth of large files. Based on the NAS protocol architecture, the OceanProtect backup appliance provides live mount and instant recovery services.

##### 4.1.1.2.1 NAS Protocols

## NFS



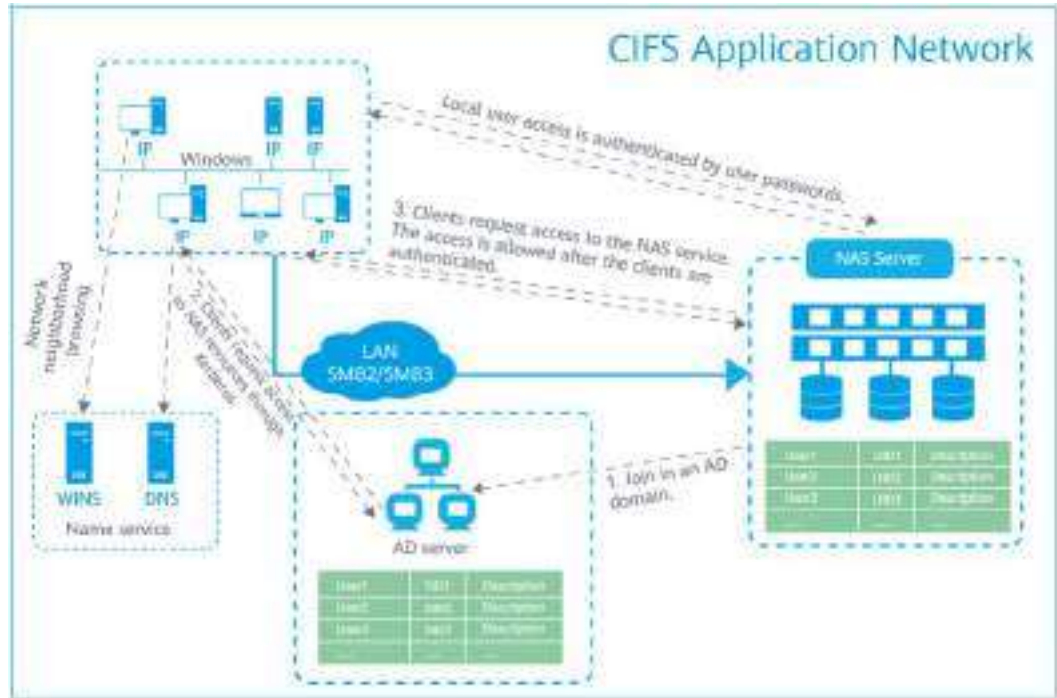
Network File System (NFS) is a common network file sharing protocol used in Linux and Unix environments.

NFS is mainly used for:

- Backup software installed on Unix and Unix-like OSs such as Linux, AIX, and Solaris

OceanProtect supports NFSv3 and NFSv4.1 protocols and usage in local user environments (non-domain environments) and LDAP/NIS domain environments. Users can import LDAP certificates for secure domain transmission with LDAPS. In a multi-tenancy environment, the LDAP/NIS service can be configured separately for each vStore.

## CIFS



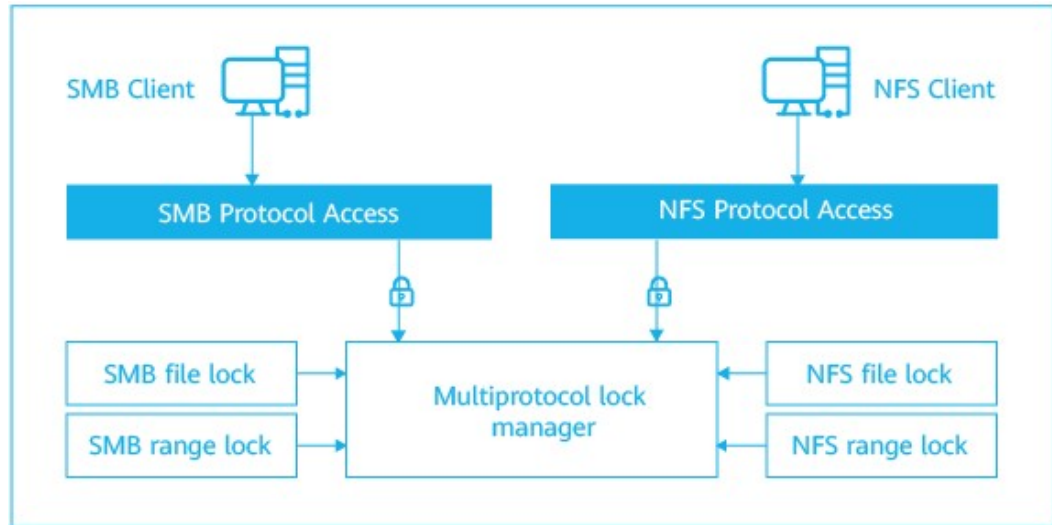
Server Message Block (SMB), also called Common Internet File System (CIFS), is a network file sharing protocol widely used in Windows environments.

SMB is mainly used for:

- Backup software installed on Windows  
OceanProtect supports SMB 2.0 and SMB 3.0, and can be used in local user environments (non-domain environments) and AD domain environments. Kerberos and NTLM authentication by AD domain is also supported. The AD domain environments can be individual, parent-child, or trusted domains. In a multi-tenancy environment, an independent AD domain can be configured for each vStore.

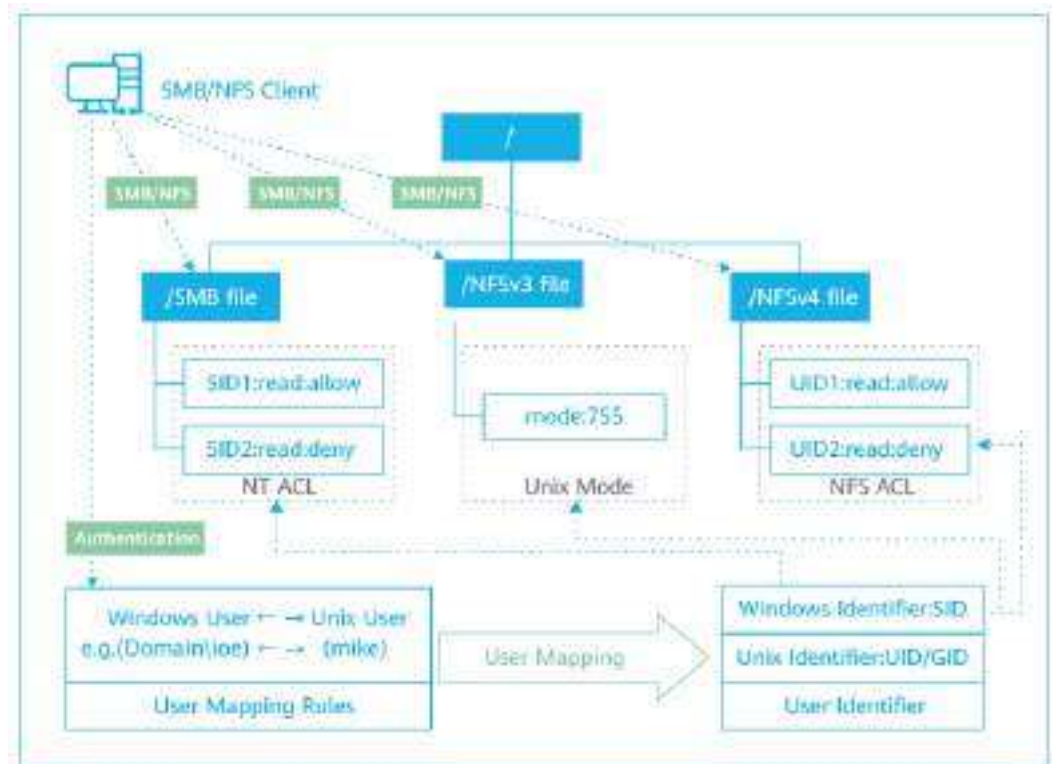
## Multi-Protocol Access

OceanProtect supports access across the NFS and SMB protocols. Both NFS and SMB shares are configurable for a file system. The system uses a multi-protocol lock manager with NFS and SMB for exclusive file access without data corruption or inconsistency.



OceanProtect supports the following security modes for multi-protocol access:

- NT mode: File attributes and ACL permissions can only be set on Windows clients with SMB. The system automatically maps the file permissions of SMB users to the NFS users on Linux clients based on the user mapping relationship for successful authentication of NFS users during file access. NFS users are prohibited from setting the mode or ACL permissions for files.
- Unix mode: File permissions can only be set on Linux/Unix clients with NFS. The system automatically maps the file permissions of NFS users to the SMB users based on the user mapping relationship for successful authentication of SMB users during file access. SMB users are prohibited from setting the ACL permissions of files.

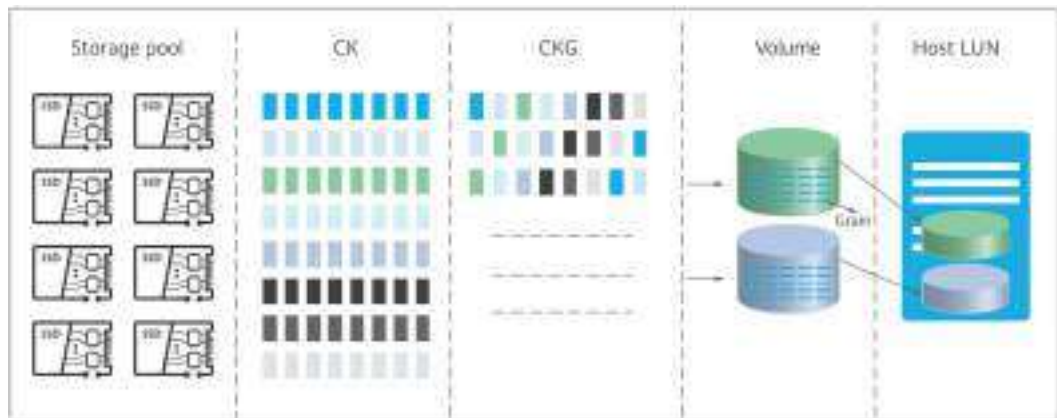


### 4.1.1.3 RAID 2.0+

If data is not evenly stored on SSDs, some heavily loaded SSDs may become the system bottleneck. OceanProtect uses RAID 2.0+ to implement fine-grained division of SSDs and evenly distributes data to all LUNs or file systems on each SSD, balancing loads among disks. RAID 2.0+ is implemented as follows:

- Multiple SSDs form a storage pool.
- Each SSD is divided into fixed-size chunks (typically 4 MB per chunk) to facilitate logical space management.
- Chunks from different SSDs constitute a chunk group based on the RAID policy set by the user.
- A chunk group is further divided into grains (typically 8 KB per grain), which are the smallest unit for volumes.

**Figure 4-2 RAID 2.0+ mapping**



# 5 Data Backup Software Design

This chapter describes key functions and technical principles of the data backup feature built in the OceanProtect X6000/X8000 backup storage.

## 5.1 Software Architecture

### 5.2 Highlights

### 5.3 Protection Ecosystem for Production Systems

### 5.4 Copy Lifecycle Management

### 5.5 Copy Data Anonymization

## 5.1 Software Architecture

This chapter describes the software architecture of Huawei OceanProtect data backup features and the containerized distributed balanced scheduling mode.

### 5.1.1 Backup Software Architecture

For OceanProtect data protection system, storage and data protection software are separately deployed. To be specific, the basic storage system runs on the host operating system (OS) to ensure the read/write I/O switching performance. The data protection software is deployed in containerized and microservice-based mode. Resources are isolated between the storage system and data protection system. Data protection systems are isolated by application resource, greatly narrowing down the fault domain. From the perspective of architecture, the OceanProtect data protection system is divided into the following layers:

Figure 5-1 Software architecture



- **Data protection client (ProtectAgent):** A backup agent is required for backup of host files and databases on hosts. The agent software to be installed is determined based on the type of applications to be protected, and application data protection is implemented on the hosts. The agent is not required for agent-free application backup.
- **Data protection layer:** This layer processes data protection services, supports container-based deployment, and uses a distributed architecture. Protection for host files, database ecosystems, virtualization/cloud/container ecosystems, and big data ecosystems is supported. End-to-end copy data ransomware protection solutions and copy data flow capabilities are provided.
- **Data storage layer:** This layer is the basic storage platform for backup data. It provides backup data transmission protocols, backup data storage and layout management, and deduplication. In addition, it provides basic capabilities including storage, snapshot, and replication for the backup system.
- **Hardware layer:** It includes basic hardware and related drivers.
- **Infrastructure:** It supports tool-based installation and deployment, and provides O&M tools. Different network management platforms of the Huawei Storage Product Line are supported, and third-party network management platforms can be interconnected using standard protocols.

## 5.2 Highlights

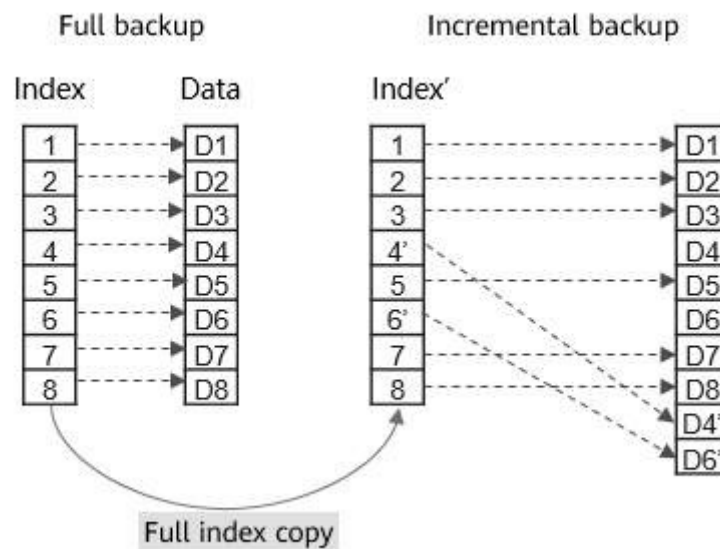
### 5.2.1 Forever Incremental Backup (Synthetic Full Backup)

Most applications supported by OceanProtect store incremental data in forever incremental backup mode. Different from traditional synthetic forever incremental backup, the backup software uses redirect-on-write (ROW) for synthesis. Backup data is written in the original file format or a fixed file format. During incremental backup, the system analyzes the differential data blocks and offset positions between the incremental data and the last backup data, overwrites the differential

data to the file to obtain the latest full synthetic copy. After the copy is synthesized, a snapshot is generated to record the relationship between the incremental data and the last backup data. During the next incremental backup, new data fragments are redirected and written to a new location based on the previous file system snapshot, improving the backup and recovery efficiency.

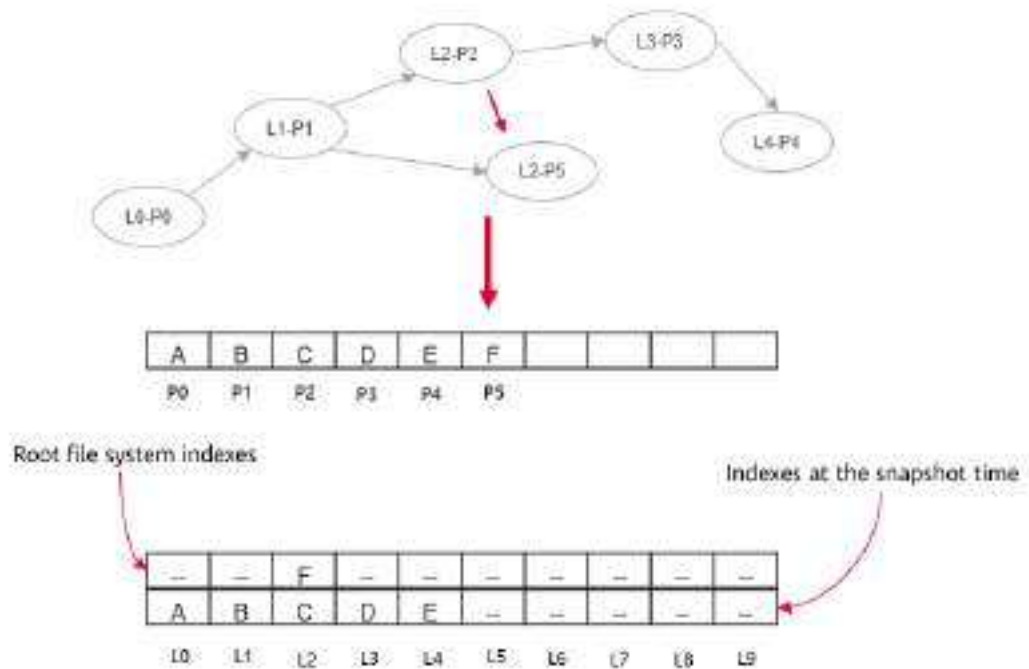
In addition, the full backup synthesis technology can be used together with applications to implement application-aware source deduplication. During full backup, only differential block data is transmitted and synthesized into a complete full backup copy on the server.

**Figure 5-2** Forever incremental backup principles of traditional vendors



When implementing forever incremental backup, the backup storage software of traditional vendors fully copies the index file of the last backup copy and then modifies the corresponding offset index based on the incremental information to obtain the full index file of the forever incremental backup copy. Alternatively, the backup storage software copies real data and writes the data into a new file. In such mode, the index file needs to be copied once and be modified for multiple times, or the entire file content needs to be copied and written, which brings great I/O load to the storage layer.

**Figure 5-3** Principles of OceanProtect ROW-based forever incremental backup

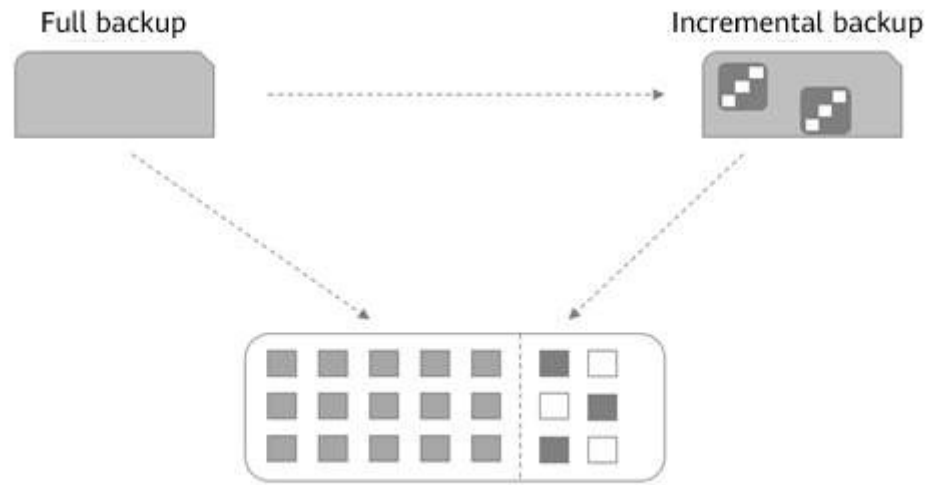


The OceanProtect forever incremental backup technology uses the ROW snapshot technology for index files. Only the index information of changed data blocks and the index dependency of data blocks are added. This greatly reduces storage I/O overhead and improves backup performance.

## 5.2.2 Backup in Native Format

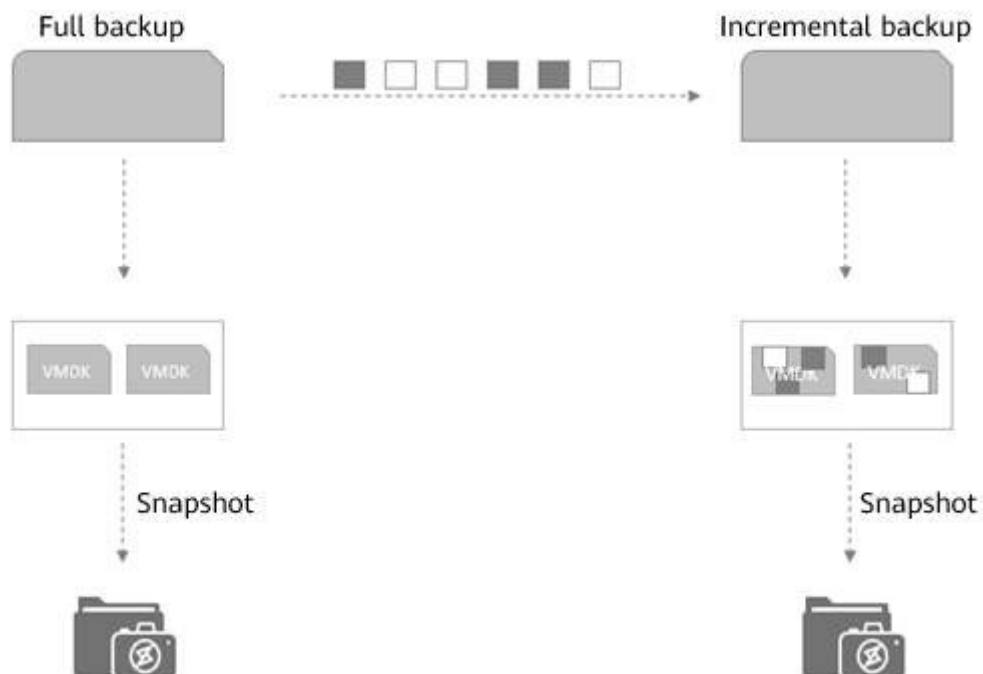
In the development of traditional backup products, backup software and storage media are deployed separately. To facilitate deduplication and encryption, backup data is split and stored in private formats, which cannot be identified by the original application software. This kind of backup is called backup in non-native format. For backup copies in non-native format, data must be completely written back to the production environment to recover services. The larger the data volume, the larger the RTO. For backup in native format, backup data is stored at the storage layer in a format that can be identified by applications. During incremental backup, incremental data and full data are integrated to form a complete backup copy that can be identified by applications, greatly improving copy recovery efficiency.

**Figure 5-4** Data storage format of traditional backup software



For backup in native format, OceanProtect combines data into a file that can be identified by applications, creates snapshots for backup files, and provides live mount and instant recovery capabilities for applications.

**Figure 5-5** OceanProtect retains data in the same native format as the production end.

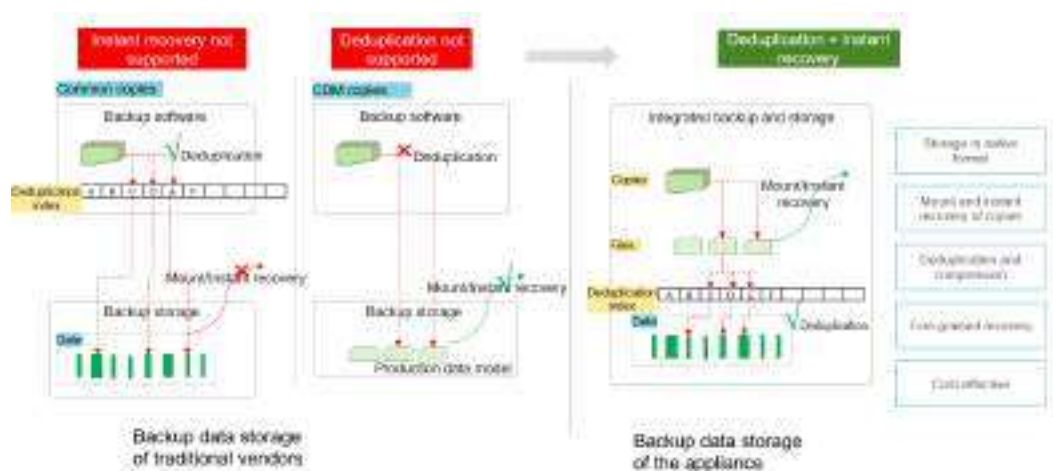


Copies in native format must be stored in a way that can be identified by common protocols so that backup copies can be identified by production applications.

Therefore, standard protocols such as SCSI, NFS, CIFS, HDFS, and S3 must be supported. The deduplication layer of traditional backup software is above the preceding protocol layers. Therefore, deduplicated data cannot be identified by standard protocols or stored in standard protocol formats. As a result, deduplication, compression, and encryption cannot be performed. The mounting capability and the deduplication and compression capability are mutually exclusive.

OceanProtect is vertically integrated with the data backup architecture. The deduplication capability is built under the file system protocol, and the standard NFS and CIFS protocols are supported. Fast recovery capabilities such as mounting and instant recovery are combined with deduplication, compression, and encryption capabilities to meet customer requirements.

**Figure 5-6** Both copy mounting and deduplication and compression are supported.

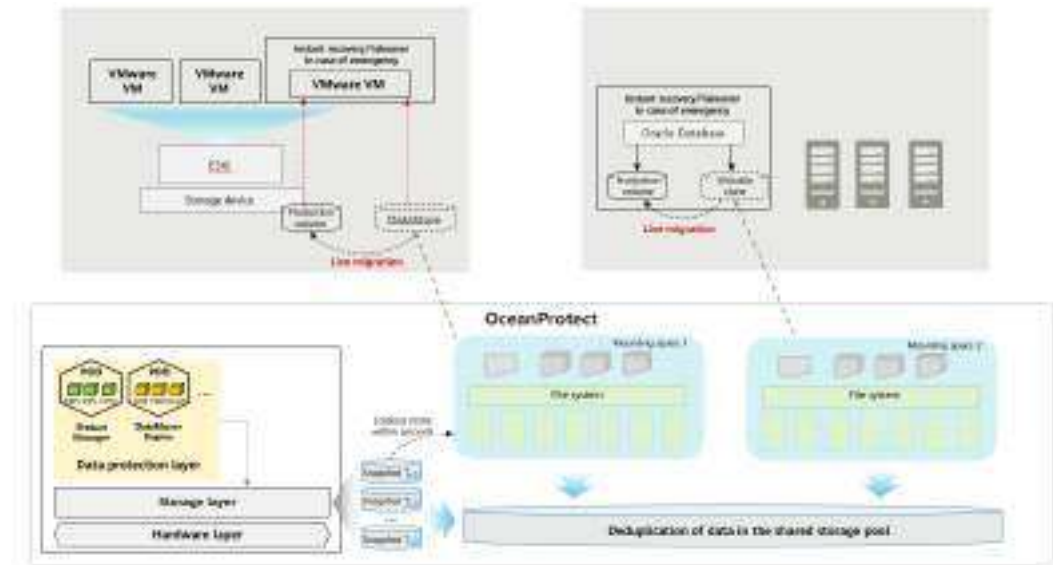


- Deduplication and compression in native backup format:** Due to software architecture restrictions, the deduplication and compression features of traditional backup software are built on the software and the deduplicated and compressed data is stored in the backup storage. However, applications run on operating systems and can identify only data on common NAS storage, SAN storage, or object storage, and deduplicated backup data cannot be identified. Therefore, live mount and instant recovery of copies cannot be performed for many applications. The OceanProtect has built-in backup software and storage systems. It implements deduplication applicable to backup scenarios under the storage layer and supports copy mounting and instant recovery through common NAS protocols.

### 5.2.3 Instant Availability of Copies

To help users view copies in real time or use backup copies for development, testing, and data analysis, the OceanProtect supports live mount of backup copies and replication copies of Oracle, VMware, NAS, MySQL, and files as well as back-end data migration of Oracle and VMware. This section describes the mounting capability and its principles.

Figure 5-7 Live mount



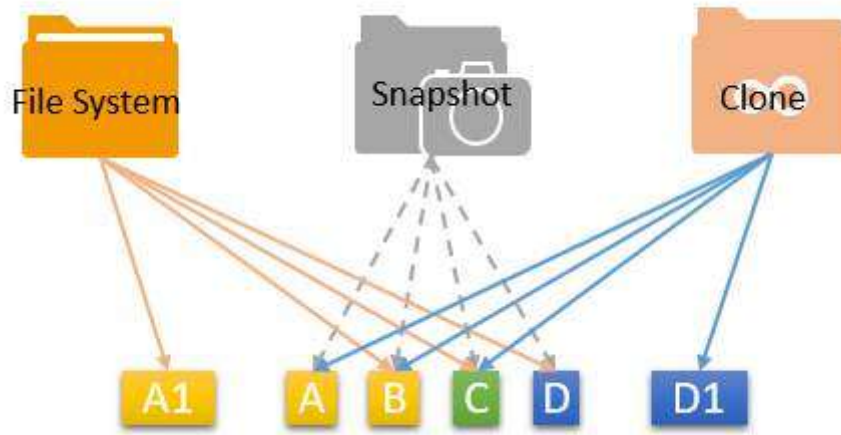
Live mount is suitable for the following scenarios:

- Manually select a copy and mount it.
- Configure a mount policy, based on which OceanProtect automatically replaces and mounts copies.
- Take over services by mounting a copy. After the services run for a period of time, they can be switched back to the production center.

For applications that support copy mounting, OceanProtect retains a read-only snapshot whose content and format are the same as those of the production application. Before performing a live mount, the system creates a lossless writable clone file system for the specified read-only copy and mounts the writable clone file system to a VM or an application. This enables the application or VM to be quickly started.

If the started application involves data changes, the new data is written to the OceanProtect and stored using ROW. Unchanged data of the original snapshot is shared, and deduplication is performed based on the same pool. After services are mounted, backup data and new service data can be seamlessly migrated back to the production storage by using VM vMotion or database storage switchover.

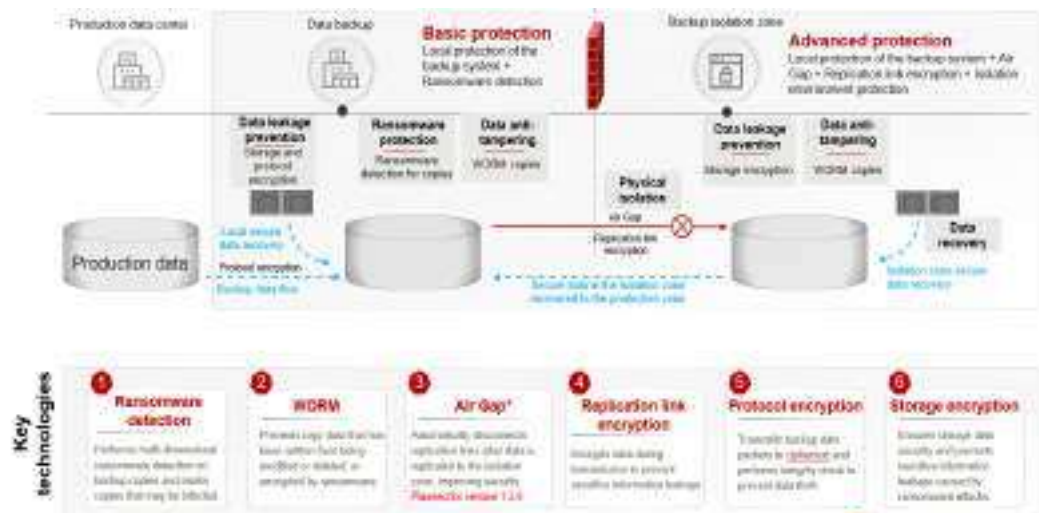
Figure 5-8 Second-level lossless clone



## 5.2.4 E2E Ransomware Protection

The OceanProtect data backup features provide an E2E ransomware protection solution for copies. In this solution, data to be backed up is encrypted and transmitted to a backup storage device, and integrity check and encryption are performed on data in links. After receiving the data, the backup storage encrypts the data and stores it to disks. For copies that pass the ransomware detection, compliance WORM attributes are configured to prevent data tampering during storage. Backup data can be replicated to and stored in a remote secure zone through encrypted links, and Air Gap is used to control the opening and closing of replication links, reducing the risk of network attacks and information theft.

Figure 5-9 E2E ransomware protection solution



### CAUTION

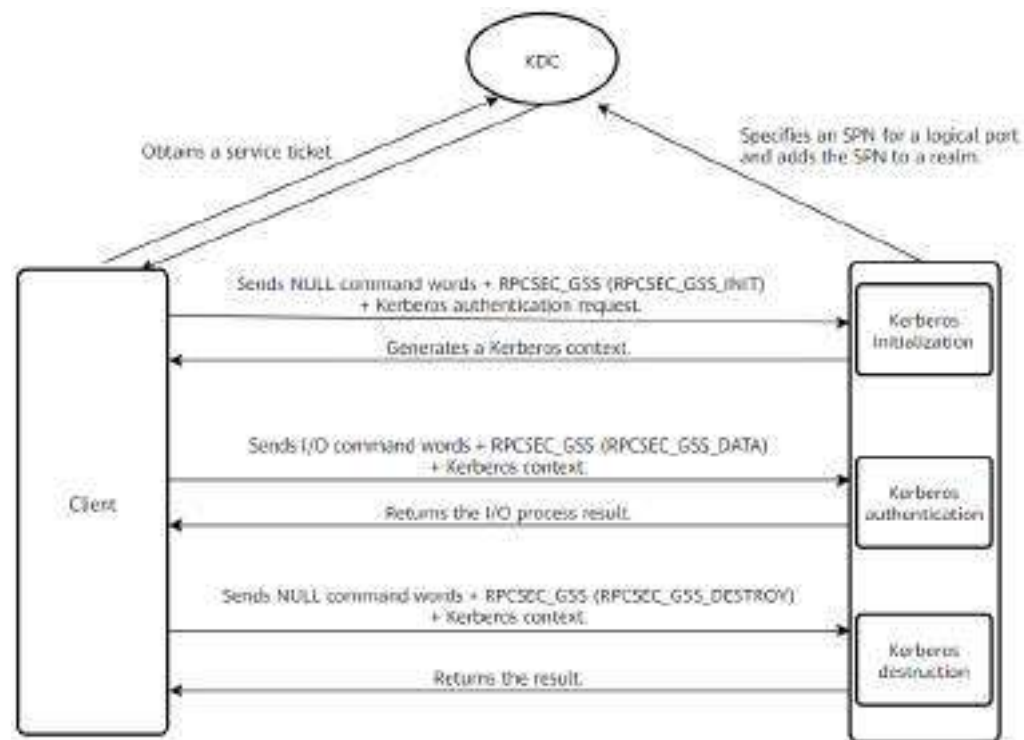
Air Gap is supported by OceanProtect 1.3.0 and later versions.

### 5.2.4.1 Encrypted Transmission

The data backup features use the NFS protocol to transmit data in the production system to backup devices for storage. The NFS protocol uses client IP addresses for authentication, which poses security risks in complex network environments. The backup features use Kerberos to authenticate and encrypt backup links. Kerberos is a client/server structure used to provide secure transactions on the network. It provides powerful user authentication, integrity, and confidentiality. Authentication is to verify the identities of the sender and receiver of network transactions. Integrity and confidentiality are guaranteed by checking the validity of the transmitted data and encrypting the data during transmission. The Kerberos service enables you to securely access computers, execute commands, exchange data, and transfer files. Encrypted data transmission provides higher security but causes performance deterioration, which should be enabled or disabled based on the actual application scenario.

NFSv4 Kerberos authentication is implemented based on the RPCSEC\_GSS security mechanism. The RPCSEC\_GSS security mechanism consists of three parts: Kerberos context creation, RPC data exchange, and Kerberos context destruction.

**Figure 5-10** Transmission encryption using Kerberos



1. The storage system adds the logical port to the realm by specifying the SPN. The KDC generates the SPN machine account. The SPN and its key are generated in the keytab file of the storage system.
2. The NFS client obtains the SPN ticket information from the realm controller.
3. When the NFS client is mounted (using the SPN of the logical port added to the realm), it sends the following information to the NFS server for Kerberos initialization: the NULL command word, RPCSEC\_GSS, RPC\_GSS\_INIT, and the ticket obtained from KDC. The NFS server verifies the ticket information of the

GSS request based on the SPN and key in the keytab file. After the verification is successful, the NFS server returns the Kerberos context to the client. The context file is used for verification when NFS Kerberos services that use the same mount point are accessed.

4. After the NFS client passes service authentication, it can access data, such as reading, writing, creating, and deleting files. During I/O data access, the client carries the authenticated Kerberos context and security mode (Krb5, Krb5i, or Krb5p). The NFS server performs Kerberos authentication based on the context. Data can be processed only after the Kerberos authentication is successful for security purposes.
5. After the NFS client completes service access, the mount point is unmounted. After unmounting, the client instructs the NFS server to clear the Kerberos cache information by sending the NULL command, RPC\_GSS\_Destroy, and the Kerberos context. After receiving the clearance request, the NFS server clears Kerberos resources based on the Kerberos context.

The backup features use Krb5p to perform integrity check and data encryption on data packets, providing the highest level of security.

#### 5.2.4.2 Encrypted Storage

To ensure the security of backup data, the data backup features support storage system-based array encryption. The built-in encryption engine of the controller processor is enabled, and internal or external key manager is configured to implement static encryption of backup data.

- Internal key manager is a built-in key management application of the storage system with the NIST SP 800-57 best practices for key lifecycle management. Being easy to deploy, configure, and manage, it is recommended if FIPS 140-2 certification and State Cryptography Administration (SCA) certification are not required and the key management system is only used by the storage systems in a data center.
- External key manager is a third-party key management system that complies with FIPS 140-2 certification and SCA certification. It uses the standard Key Management Interoperability Protocol (KMIP) and Transport Layer Security (TLS) to communicate with storage systems. The external key manager is recommended if FIPS 140-2 certification is required or multiple systems in a data center require centralized key management.

Backup data encryption uses the built-in encryption engine of the controller processor to implement encryption and decryption. The independent built-in encryption engine leverages the encryption algorithm of Arm hardware to offload encryption workloads. The encryption and decryption algorithms are offloaded to the hardware for execution, without involving software. During data encryption at the backup storage layer, the encryption subsystem generates a data encryption key (DEK) on each disk, and the key manager provides an authentication key (AK). The AK is used to encrypt the DEK. After service I/Os are delivered, encryption and decryption are offloaded to the built-in encryption engine for execution. The encryption engine supports the AES-256-XTS and SM4-128-XTS (only for the Chinese mainland) algorithms. The algorithm used by the key manager must match that used by the encryption engine.

The internal key manager is the storage system's built-in key management system based on NIST SP 800-57 best practices. It generates, updates, backs up, restores,

and destroys keys, and provides hierarchical key protection. Being easy to deploy, configure, and manage, it is recommended if FIPS 140-2 certification is not required and the key management system is only used by the backup systems in a data center.

The internal key manager provides the following six features:

- **Key protection**  
The internal key manager supports three-layer key design, including the root key, master key, and work key.
- **Root key:** first-layer key in the multi-layer key system, used to encrypt the master key
- **Master key:** encryption key, used to encrypt the work key
- **Work key:** working key, used to encrypt and decrypt data
- **Key backup**  
The internal key manager enables you to manually or automatically back up key ciphertext to an external Secure File Transfer Protocol (SFTP) or File Transfer Protocol (FTP) server.
- **Key recovery**  
The internal key manager enables you to restore the backup key using SFTP or FTP.
- **Key generation**  
When the encryption function is enabled on the OceanProtect backup storage system, the block device management subsystem of the OceanProtect backup storage system applies for and obtain an encryption key from the internal key manager.
- **Key update**  
If the lifecycle of an encryption key ends or the key is disclosed, the internal key manager allows the block management subsystem to apply for updating the encryption key.
- **Key destruction**  
If the lifecycle of an encryption key or a disk ends, you need to destroy the encryption key of the disk, which is supported by the internal key manager.

#### NOTE

The internal key manager supports the AES algorithm rather than SM4 algorithm.

The security capability of the internal key manager depends on the storage system. For details about how to build the security capability of the storage system, see the relevant security technical white paper.

The external key manager stores encryption keys outside the storage system and supports key generation, update, destruction, backup, and restoration. It uses KMIP and TLS protocols for key transmission with storage controllers to ensure key security. In addition, two external key managers can be deployed, in which case keys are synchronized between the two external key managers in real time to ensure key reliability.

#### NOTE

The external key manager supports the AES algorithm and SM4 algorithm (for details about the supported algorithms, see *OceanProtect Backup Storage System Disk Encryption User Guide*) and meets the requirements of FIPS 140-2 certification and SCA certification. To use the external key manager, contact the MO to check its availability.

### 5.2.4.3 WORM

To ensure the storage reliability of backup data and prevent data from being tampered accidentally or intentionally during storage, the data backup features support WORM locking for backup copies. A backup copy written to the storage system and in protected state can only be read.

The data backup features add file-related attributes to the file system at the storage layer to control file access and modification. The WORM attributes of each file are saved to prevent data tampering.

During the retention period, WORM copies cannot be modified or deleted. In addition, the WORM function can work with the ransomware detection function. After ransomware detection, WORM properties are set only for uninfected copies to protect the backup data in the storage system.

#### NOTICE

Huawei OceanProtect data protection devices comply with SEC Rule 17a-4(f) to provide data protection.

### 5.2.4.4 Ransomware Detection

Ransomware is a new type of computer virus that spreads in the form of emails, Trojans, and Trojan-infected URLs. Ransomware brings immeasurable loss to users. It uses various encryption algorithms to encrypt files. Users can only decrypt files by using the private key.

During backup, traditional vendors only migrate data without considering the security of copies. When the production environment is infected by ransomware, infected files are copied during backup. As a result, the copies are also infected by ransomware, which cannot be discovered by users.

The OceanProtect data backup features have the built-in intelligent ransomware detection capability. After a backup job is complete, feature check and behavior analysis are performed on the copies. The baseline feature model is used to perform basic check on the backup copies. If the copies are suspected of being infected by ransomware, the AI machine learning check model is started for in-depth analysis. When a copy is highly suspected of being infected by ransomware, the copy is marked and an alarm is generated.

Figure 5-11 Intelligent ransomware detection

The ransomware protection system mainly uses features to detect copy content. The system uses application statistics methods and data visualization to collect statistics on basic change characteristics, explore the differences between normal file system changes and changes caused by ransomware, and establish a baseline model to filter out suspicious file system changes for further detection and determination.

Table 5-1 Basic change characteristics

Basic characteristics	Description	Detail
Number of extension type changes	Number of extensions of changed files before and after the change – Number of same extensions of changed files before and after the change	File name extensions are seldom changed but ransomware usually adds specific or random extensions. If many extension types are changed or if there are few extension types after the change, ransomware encryption may exist.
Number of extension types after change	Number of extension types after file change	
Number of st_size changes	Number of file size changes	The change in the size of multiple files may be caused by ransomware encryption.

Basic characteristics	Description	Detail
Number of changed file header types	Type of the file header after the change – Type of the file header before the change	Files of the same type start with the same file headers so that there are few file header types. However, most ransomware also encrypts file headers and causes more different file headers. Some ransomware adds the same file header to all encrypted files, according to which you can determine whether the file is infected by ransomware.
Number of file tail types after the change	Number of file tail types after the change	Most ransomware adds the same file tail to the encrypted files. Therefore, you can determine whether ransomware infection exists by checking whether the number of file tail types decreases by a certain value and whether the number of file tail types after the change is less than a certain value.
Number of changed file tail types	Number of file tails after the change – Number of file tails before the change	
Number of new files with the same name	Number of new files with the same name	Ransomware usually places ransomware notes in each directory. If a certain number of new files have the same file name, ransomware intrusion may exist.
Ratio of new file header types	Number of new file header types/Number of new files	If ransomware generates encrypted files by creating, copying, encrypting, and deleting original files, and the file headers of the encrypted files are different, the proportion of new file headers is high.

If any of the preceding basic features reaches the corresponding threshold shown below, the backup copy is marked as a copy suspected of being infected by ransomware.

- Number of extension type changes  $\geq$  [Threshold]
- Number of extension types after change  $\leq$  Threshold and number of changed extension types  $\neq 0$
- Number of st\_size changes  $\geq$  [Threshold]
- Number of changed file header types  $\geq$  [Threshold]
- Number of file tail types after the change  $\leq$  [Threshold] and number of changed file tail types  $\leq$  [Threshold]
- Number of new files with the same name  $\geq$  [Threshold]
- Ratio of new file header types  $\geq$  [Threshold]

For copies suspected of being infected by ransomware based on the baseline model, the OceanProtect backup storage system collects the full change

features of the dataset, uses the machine learning classification model to determine whether the copies are infected by ransomware, and marks the infected copies.

**Table 5-2** Full change features

Feature	Description
Modified file ratio	Number of modified files/Total number of original files
Number of changed extension types	Number of extension types after change - Number of extension types before change
Extension diff value before and after the change	Number of extensions of changed files before and after the change – Number of same extensions of changed files before and after the change
Number of extension types after change	Statistics on the number of extension types after the files are changed
File name change ratio	Statistics on the types and number of diff files whose file names are changed. Same-name change mode/Changed file quantity
st_mode change ratio	Number of files whose st_mode is changed/ Total number of changed files
st_size change range	Maximum and minimum size changes
Standard deviation of st_size change	Dispersion degree of size change value
Number of modes in st_size change values	Number of modes in size change values
Ratio of modes in st_size change values	Number of modes in size change values/Total number of changed files
Ratio of st_mtime that is not changed in the changed files	Number of unchanged st_mtime in the changed files/Total number of changed files
st_mtime range	Maximum st_mtime – Minimum st_mtime
Standard deviation of st_mtime variation	Dispersion degree of st_mtime of changed files
Ratio of file header change	Number of files with changed file headers/ Total number of changed files
Ratio of file header type after change	Number of file header types of changed files/ Total number of changed files
Number of changed file header types	Type of the file headers after change – Type of the file headers before change

Feature	Description
Total entropy of file headers after change	Combine the changed file headers and calculate the total entropy.
Variation of the total entropy of file headers	Total entropy of file headers after change – Total entropy of file headers before change
Number of file tails after change	Number of files with the same 4-byte tail after change
Number of file tail types of changed files	Number of types of 4-byte tails of changed files
Ratio of file tail types of changed files	Number of file tail types of changed files/Total number of changed files
Number of changed file tail types	Number of file tails after the change – Number of file tails before the change
Total entropy of changed file tails	Combine the changed file tails and calculate the total entropy.
Change of total entropy of file tails	Total entropy of changed file tails – Total entropy of file tails before change
Number of new files with the same name	Number of new files with the same name
Check whether the name of the new files with the largest number contains special words.	Check whether the new file name with the largest number contains special words, such as crypt, recover, restore, and readme.
Number of new extension types	Number of new file extension types and initial extension type difference
Number of new extensions	Number of files with the same file extension which is mostly shared among new files
st_mtime range of new files	Maximum st_time - minimum st_time of new files
Standard deviation of new file st_mtime	Dispersion degree of st_time of new files
Ratio of new file header types	Number of new file header types/Number of new files
Number of new file tails	Number of files with the same file tail which is mostly shared among new files
Backup fle deletion ratio	Number of deleted backup files/Total number of backup files

### NOTE

Currently, ransomware detection is supported only for VMware and NAS backup copies.

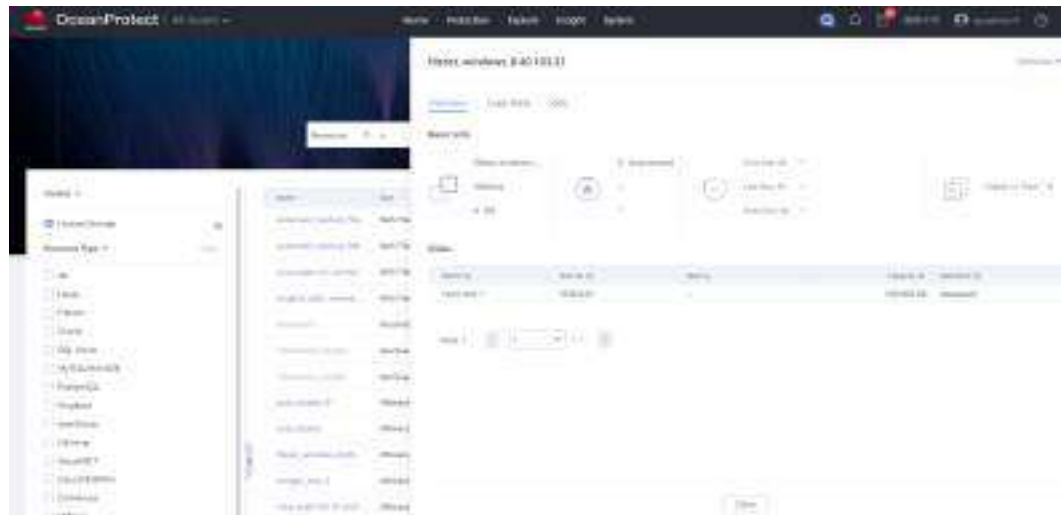
## 5.2.5 Global Search

The OceanProtect data backup feature provides the capability of searching for protected resources in the entire system. It can quickly search for all objects (such as VMware vCenter VMs) in the protected environment, search for directories and files in VMware VM backup copies and host file backup copies in seconds, and download target files and perform fine-grained recovery.

Both online and offline index creation are supported. Online indexes are created immediately after a backup copy is generated. Offline indexes are manually created for a specific copy after the backup is complete. After the backup is complete, a job that creates indexes for copy data is automatically or manually triggered on the GUI. Note: The job creates indexes for VMware VM backup copies. The host file-level index data is directly exported from the metadata during the backup. The system creates indexes as follows: 1. Mounts a copy in native format to a node. 2. Identifies and mounts the file system in the copy to the controller node. 3. Traverses the directories and files in the file system to generate index information of the copy content.

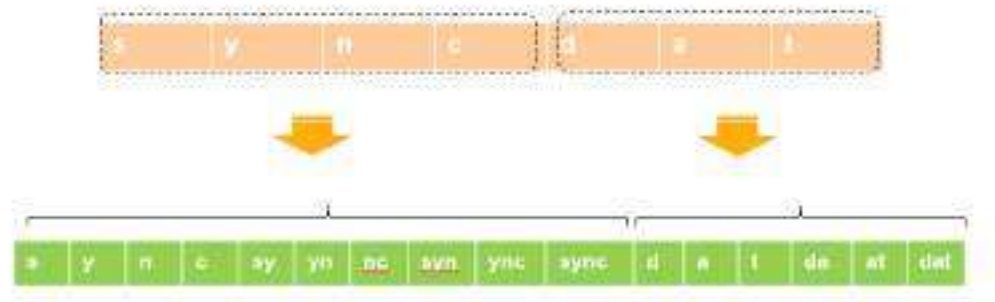
The global search function quickly and intuitively displays the content of copies, helping users quickly detect the change process of any file in multiple backup copies and providing basic capabilities for file level restore.

Figure 5-12 Global search



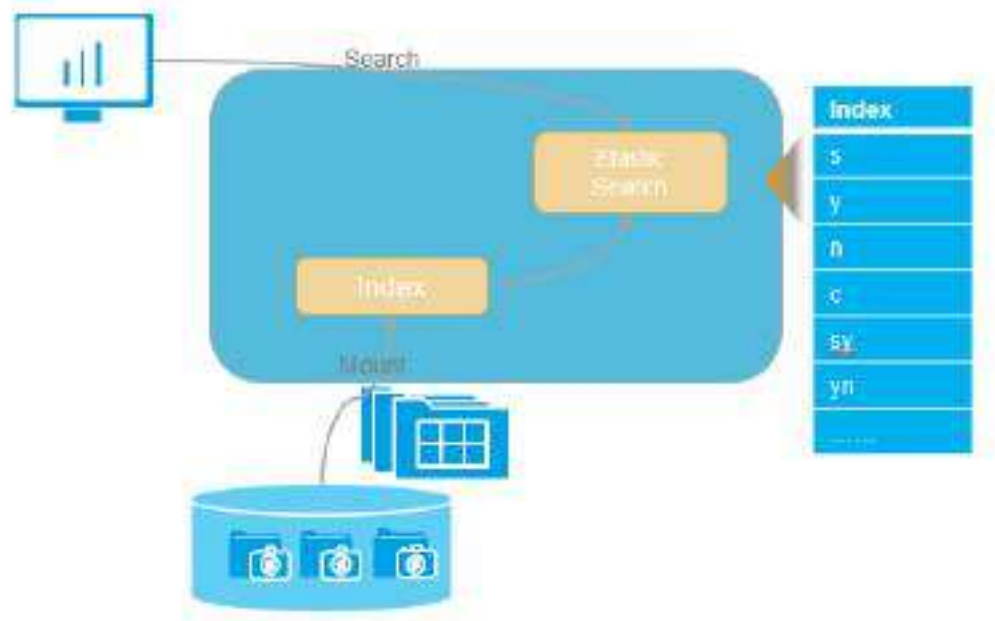
The retrieval system splits a file or directory name in a copy into four-byte parts to create index information. This balances the index information storage space and improves the search accuracy. For example, the system splits the file name **syncdat** into **sync** and **dat** to create an index. The index is a random combination of consecutive characters. For example, the index for the **syncdat** file is as follows:

**Figure 5-13** Creating an index



When an index is created, OceanProtect implements the following operations: 1. Mounts a copy to the index container. 2. Traverses the directories and files of the copy and creates indexes for them based on the preceding index creation policy. 3. Records the weights and word segmentation results to the distributed real-time search engine Elasticsearch. 4. Ranks and scores the results based on the weights and displays the results to the user when users search for data.

**Figure 5-14** Copy content indexing



**NOTE**

Currently, content-level file retrieval is supported for VMware VM, Hyper-V, and RHV VM copies, big data HDFS copies, NAS massive small file backup copies, and host file-level copies. For other applications, only the information of the protected objects can be searched for.

## 5.3 Protection Ecosystem for Production Systems

Huawei OceanProtect data backup features provide rich application ecosystem compatibility. It can protect: Hosts, NAS servers, databases, big data, virtualization, cloud platforms, and containers.

Application backup functions:

### 5.3.1 Host/NAS Backup

OceanProtect provides host and NAS backup capabilities, including the backup of host files, volumes, operating system status, and NAS files. To back up host files, you need to install a data protection client on the host and back up the files in LAN-Base mode. NAS storage does not require an agent and NAS file backup is implemented in Server-Free mode.

**Host volume/OS status backup** is based on the snapshot (such as VSS and LVM) and host changed block tracking (CBT) technologies to track changed data and can achieve higher backup performance.

- Snapshot-based consistency protection: The snapshot-based consistency backup and the host CBT technology are used to improve incremental backup performance.

**Host file system backup** is based on the quick scanning. The file/directory association difference quick scanning technology and small-file aggregation technology are used to improve the scanning and backup performance.

- Quick backup solution based on quick scanning and aggregation: Quick scanning based on file differences and small-file aggregation technologies are used, improving backup performance.

Based on the production storage type, there are three technical solutions for **file backup of NAS storage systems**:

- Ultra-fast backup solution based on SnapDiff and block-level incremental backup: OceanStor storage products perform scanning-free SnapDiff snapshot difference comparison. In addition, the intra-file block-level incremental backup technology is used to provide ultra-fast backup capabilities three times higher than those of peer vendors.
- Quick backup solution based on quick scanning and aggregation: The quick scanning technology based on file differences and small-file aggregation technologies are used to improve backup performance.
- NDMP-based quick incremental backup of NAS files.

Highlights

- **Agent-free backup:** No backup agent is required in any scenario for the backup of massive NAS small files.
- **Forever incremental backup (synthetic full backup):** Based on the ROW snapshot technology, full copies are synthesized during backup, improving recovery performance. After the recovery dependency between the copies is canceled, no extra space is occupied, and any copy can be deleted.

- **Concurrent backup and recovery:** A single job can be split into multiple subjobs which are concurrently executed on multiple nodes and controllers.
- **File aggregation:** Backup files can be aggregated to improve backup performance in KB-level small file scenarios and prevent disk I/Os from becoming a performance bottleneck.
- **Copy mount:** Backup copies can be mounted and backup data can be used.
- **Global search and fine-grained recovery:** Global search enables the search of any files or directories to be recovered in seconds, improving the recovery efficiency.
- **Data compression and deduplication:** Storage layer deduplication and source deduplication of backup data are supported. In addition to deduplication, advanced functions such as copy mount and instant recovery are supported.
- **Backup copy replication and archiving**  
Backup copies can be replicated to OceanProtect backup storage with data backup features or be archived to object storage or tape libraries. A replication copy can be used in a similar way as a backup copy. Archive copies can be used to restore data.
- **Traffic control**  
OceanProtect supports flow control of backup jobs, properly plans the transmission time and bandwidth of backup objects, and reduces the impact on the running of the production system.

#### NOTE

For details, see the *Huawei OceanProtect Technical White Paper for NAS Backup*.

## 5.3.2 Database Ecosystem Backup

Huawei OceanProtect data backup features provide forever incremental backup, periodic full backup, and log backup for databases. Based on user configurations, OceanProtect backs up production databases at specified time points and stores the backup data in the backup storage pool. When production data is damaged, backup data can be used to quickly recover the production data or temporarily take over services.

The supported databases include Oracle, SQL Server, MySQL, MariaDB, DB2, SAP HANA, SAP Oracle, Exchange, GaussDB T, openGauss, Dameng, PostgreSQL, Enmotech (openGauss), and VAST Data (openGauss).

Database backup has the following features:

- **Full backup, incremental backup, differential backup, and forever incremental backup**  
OceanProtect supports flexible backup policies, including forever incremental backup, periodic full backup, differential backup, and log backup.
- **Instant recovery**  
Backup data in native format can be used to quickly start applications before production data recovery is performed, thereby shortening service interruption and decreasing RTO (recovery time objective).
- **Data recovery to any point in time**

Log backup enables production data to be recovered to any time point, shortening RPO.

- **Data recovery to the original host, a different host, the original location, or a specified location**
- **Data backup is performed on the secondary node preferentially**, reducing the impact on production services.
- **High reliability of backup jobs**: The active and standby roles are identified during each backup job. The backup job will not fail in the case of master/slave database switchover or node faults.
- **Data deduplication and compression**  
Backup data is deduplicated at the storage layer and source end. In addition to data deduplication, advanced functions such as copy mount and instant recovery are supported.
- **Data encryption**  
Backup data can be encrypted during transmission or on storage.
- **Backup copy mounting**  
Backup copies can be quickly mounted to take over production services.
- **Backup copy replication and archiving**  
Backup copies can be replicated to an OceanProtect backup storage device with data backup features or be archived to object storage or tape libraries. A replication copy can be used in a similar way as a backup copy. Archive copies can be used to restore data.
- **Traffic control**  
OceanProtect supports flow control of backup jobs, properly plans the transmission time and bandwidth of backup objects, and reduces the impact on the running of the production system.

#### NOTE

The functions may vary with databases. For details about the functions and technical features of databases, see the *Huawei OceanProtect Backup Technical White Paper for Databases*, *Huawei OceanProtect Backup Technical White Paper for Exchange*, and *Huawei OceanProtect Backup Technical White Paper for SAP*.

### 5.3.3 Virtualization, Cloud, and Container Ecosystem Backup

For the system backup in virtualization, cloud, and container scenarios supported by data backup software, the data protection agent needs to be deployed on a physical machine or the VM in some scenarios to control the backup process and data exchange because Huawei backup storage uses the full-Arm architecture.

The features of data protection in virtualization environments are as follows.

Supported application types: FusionCompute, Huawei Cloud Stack, Hyper-V, Red Hat Enterprise Virtualization, Kubernetes+FlexVolume, and VMware.

- **Full backup and forever incremental backup**
- **Automatic environment scanning** and backup of VMs, hosts, datastores, CNA nodes, container namespaces, and StatefulSet
- **Concurrent backup and recovery**: A single job can be split into multiple subjobs which are concurrently executed on multiple nodes and controllers.

- **Block-level incremental backup:** Block-level incremental backup is supported, greatly reducing the amount of data to be backed up.
- **Data deduplication and compression**  
Storage layer deduplication and source deduplication of backup data are supported. In addition to deduplication, advanced functions such as copy mount and instant recovery are supported.
- **Instant recovery**  
Instant recovery can quickly recover production services within seconds.
- **Live mount**  
Live mounts do not affect the original production environment. In the new environment, service VMs are started in seconds, providing an environment for quick development and test as well as data analysis. Live mounts also allow users to configure automatic mount update policies, eliminating the need for periodic manual configuration for copy mounting.
- **User-defined pre-processing and post-processing scripts**  
You can customize scripts to run before a job or after a job succeeds or fails, adapting to various scenarios.
- **Service migration from VMs based on live mounts**  
VMs that have run for a period of time can be migrated back to the production environment.
- **Data recovery to the original or a specified location**  
Data can be restored to the original, other VMs, and new VMs.
- **File-level recovery**  
Fine-grained recovery of files or directories in copies is supported. Desired files or directories can be accurately recovered.
- **Retrieval of files or directories in copies**  
You can quickly search for files or directories in the copies by name.
- **Backup copy replication and archiving**  
Backup copies can be replicated to another OceanProtect backup storage device with backup software capability enabled or be archived to object storage or tape libraries. A replication copy can be used in a similar way as a backup copy. Archive copies can be used to restore data.
- **Traffic control**  
OceanProtect supports flow control of backup jobs, properly plans the transmission time and bandwidth of backup objects, and reduces the impact on the running of the production system.

#### NOTE

The actual functions vary in different production environments. For details about the functions and technical features of virtualization, cloud, and container environments, see the *Huawei OceanProtect Virtualization Backup Technical White Paper*.

## 5.3.4 Big Data Ecosystem Backup

The data backup software supports backup of HDFS in the Cloudera CDH/CDP big data platform, Hadoop HDFS, Huawei FusionInsight, and MapReduce Service. Multi-node concurrent stream backup is used to back up data from multiple

backup agent nodes. Compared with file-level differential data comparison, the OceanProtect backup storage HDFS backup supports incremental data backup based on file content segments, greatly reducing the amount of data to be backed up during incremental backup. In addition, the ROW technology is used to implement real-time full copy synthesis in the background, reducing RTO. The deduplication technology at the storage layer is used to provide backup in native format, global search, and file-level recovery capabilities in addition to data deduplication and compression.

The following component types are supported: HDFS, HBase, Hive, Elasticsearch, Redis and ClickHouse.

HDFS backup has the following features:

- **Flexible agent deployment:** Backup agents can be deployed in a non-intrusive manner or on big data platform nodes.
- **Multi-component protection:** Components such as HDFS, HBase, Hive, Elasticsearch, Redis and ClickHouse are supported.
- **Forever incremental backup:** Based on the ROW snapshot technology, full copies are synthesized during backup, improving recovery performance. After the recovery dependency between the copies is canceled, no extra space is occupied, and any copy can be deleted.
- **Fine-grained backup:** Database backup in namespace and table level, distributed file system directory-level and file-level backup, as well as search engine index-level backup is supported.
- **Fine-grained recovery:** Specific contents can be recovered by using backup copies.
- **Concurrent backup and recovery:** A single job can be split into multiple subjobs which are concurrently executed on multiple nodes and controllers.
- **Global search and fine-grained recovery:** Global search enables the search of any files or directories to be recovered in seconds, improving the recovery efficiency.
- **Block-level incremental backup:** Block-level incremental backup based on file content is supported. This function applies to HDFS file append writing and modification scenarios, greatly reducing the amount of data to be backed up.
- **Recovery to any point in time:** Table data can be recovered to any point in time based on WAL-based HBase backup.
- **Data compression and deduplication:** Backup data is deduplicated at the storage layer and source end. In addition to data deduplication, advanced functions such as fine-grained recovery are supported.
- **Data consistency:** HDFS backup is based on snapshots and supports point-in-time consistency of files within copies.
- **Backup copy replication and archiving**  
Backup copies can be replicated to another OceanProtect backup storage device with backup software capability enabled or be archived to object storage or tape libraries. A replication copy can be used in a similar way as a backup copy. Archive copies can be used to restore data.
- **Traffic control**

OceanProtect supports flow control of backup jobs, properly plans the transmission time and bandwidth of backup objects, and reduces the impact on the running of the production system.

#### NOTE

For details about the capabilities of components in the big data environment, see the *OceanProtect Big Data Backup Technical White Paper*.

## 5.3.5 Data Warehouse Backup

GaussDB (DWS) is an online data processing database that uses Huawei Cloud or Huawei Full-Stack Cloud infrastructure to provide scalable, fully-managed, and out-of-the-box analytic database service that frees you from database management and monitoring. It is a native cloud service based on the Huawei converged data warehouse GaussDB, and is fully compatible with the standard ANSI SQL 99 and SQL 2003, as well as the PostgreSQL and Oracle ecosystems. GaussDB (DWS) provides competitive solutions for PB-level big data analysis in various industries.

Based on the DWS cloud infrastructure, Huawei OceanProtect data backup feature provides database backup and recovery capabilities for DWS. ProtectAgent and RoachClient are deployed together to provide interfaces for RoachClient to invoke and write backup data to storage devices.

GaussDB OLAP and DWS are supported.

Technical highlights

- **E2E O&M:** A complete DWS backup management interface enables convenient management operations.
- **Distributed concurrent stream backup** enables multiple devices to concurrently receive backup data, significantly reducing the backup time window.
- **External backup agent can be deployed**, without any impact on the production system.
- **The system supports scale-out** as production services increase.
- **Data deduplication and compression**  
Backup data is deduplicated and compressed at the target end at block level.
- **Fine-grained recovery**, shortening RTO
- **Backup copy replication and archiving**  
Backup copies can be replicated to another OceanProtect backup storage device with backup software capability enabled or be archived to object storage or tape libraries. A replication copy can be used in a similar way as a backup copy. Archive copies can be used to restore data.
- **Traffic control**  
OceanProtect supports flow control of backup jobs, properly plans the transmission time and bandwidth of backup objects, and reduces the impact on the running of the production system.

#### NOTE

For details about more DWS backup technologies, see sections about DWS backup in the *OceanProtect Backup Technical White Paper for Databases*.

## 5.4 Copy Lifecycle Management

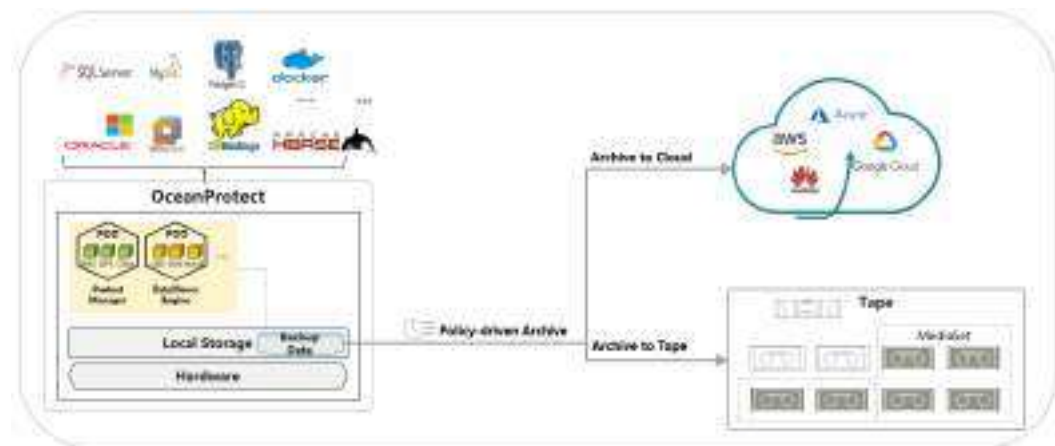
Huawei OceanProtect data backup features provide various copy lifecycle management and copy data flow capabilities for completed backup copies. It can archive copy data to object storage, such as OceanStor Pacific (S3), FusionStorage OBS, OceanStor 9000 (S3), Huawei Cloud OBS, Amazon AWS, Microsoft Azure, Google Cloud, and tape storage media, and replicate copy data to another OceanProtect X8000 device.

### 5.4.1 Copy Archiving

Enterprise users often need to retain backup data for a long time (generally five or ten years, or a longer period of time) due to laws and regulations or service requirements. To reduce construction costs and eliminate data silos, multi-cloud data environments can be used. Huawei OceanProtect data backup and archiving features allow users to archive full or partial copies of protected objects to the cloud environment or offline storage (such as tape media). This feature is further divided into cloud archiving and tape archiving based on storage media.

Both cloud archiving and tape archiving are triggered and executed based on SLA policies. Restoration using the archived copies takes a long time. If the recovery time objective (RTO) of backup data is not required, the data can be archived to external archive media.

Figure 5-15 Copy archiving



#### 5.4.1.1 Cloud Archiving

Copy cloud archiving is developed based on SDKs or APIs provided by cloud vendors. Archiving jobs of a single copy are concurrently executed on multiple controller nodes based on job splitting. In addition, the BBR acceleration capability is used to provide efficient data cloud archiving.

The archive service splits an archive job into multiple sub-jobs and sends the sub-jobs to the communication framework cache queue. Then, the archive service obtains sub-jobs from the cache queue together with the archiving services on other nodes. After obtaining sub-jobs, the archive service queries the detailed

information about the copy and the segment differential bitmap for incremental archiving from the original service where the sub-jobs are located based on the job type. Based on the differential bitmap information, the archive service reads the corresponding data block, splits the large block or aggregates the small blocks into the 4 MB data blocks, and sends the 4 MB data blocks to the object storage after compression.

During restoration using an VMware or database archive copy, data in the object storage is read to the local host, and then mounted to the restoration target for instant recovery. A part of backup storage space is occupied during the restoration and will be released after the restoration. For recovery using archive copies of other applications, data can be directly read from the object storage and recovered to the production center.

Long-term Retention of Copies (Copy Archiving)		
Copy archiving	RPO	Minimum value: Copies are archived immediately after backup. (The minimum RPO for periodic archiving is 1 hour.)
	Archive mode	<ul style="list-style-type: none"><li>- Copies are archived immediately after the backup is complete.</li><li>- Copies are archived periodically.</li><li>- Copies are archived after XX.</li></ul>
	Archive granularity	Specified copies can be archived (specified copies are archived based on policies).
Archive object	Backup copies	Full backup and forever incremental backup
	Replication copies	Full backup and forever incremental backup
Archive target	Target	AWS Azure Google Cloud OceanStor Pacific (S3) FusionStorage OBS OceanStor 9000 (S3) Huawei Cloud
	Maximum number of archive targets	4 (A maximum of 1:4 archive network is supported.)
Data reduction	Storage space reduction	Data compression
	Transmission link reduction	Data compression

Archive network	Network quality	BBR-based transmission protocol optimization, tolerating the worst network quality with a TTL $\leq 100$ ms and a 5% packet loss rate
	Bandwidth limit	Supported

### 5.4.1.2 Tape Archiving

Copy cloud archiving is applicable to online backup data for long-term offline retention. Concurrent archiving of multiple drives is implemented and a media set is formed by multiple tapes which can be overwritten cyclically. The specifications are as follows:

Long-term Retention of Copies (Copy Archiving)		
Copy archiving	RPO	Minimum value: Copies are archived immediately after backup. (The minimum RPO for periodic archiving is 1 hour.)
	Archive mode	- Immediately archive after backup – Periodically archive – Archive after the copy is retained for <i>XX</i> days
	Archive granularity	Specified copies can be archived (specified copies are archived based on policies).
Archive object	Backup copies	Full backup and forever incremental backup
	Replication copies	Full backup and forever incremental backup
Archive target	Disk-to-disk-to-tape (D2D2T)	LTO5, LTO6, LTO7, and LTO8
	Maximum number of archive targets	4 (A maximum of 1:4 archive network is supported.)
Data reduction	Storage space reduction	Data compression
	Transmission link reduction	Data compression

#### NOTICE

For details about archiving and replication technologies, see the *OceanProtect Technical White Paper for Copy Archiving and Replication*.

## 5.4.2 Copy Replication

Huawei OceanProtect data backup allows users to replicate copies by protection resource to another OceanProtect device with the data backup feature enabled. This enables services to be quickly recovered at the remote site in the event of a disaster at the source data center. The following table lists the copy replication specifications.

Copy Replication		
Copy Replication	RPO	Minimum value: Copies are replicated immediately after being backed up. (The minimum RPO for periodic replication is 1 hour.)
	Replication mode	- Immediately replicate after the backup is complete - Periodically replicate
	Replication granularity	Copies are replicated in terms of protection resources. Some copies of a single protection resource cannot be replicated.
	Copy replication object	Full, incremental, differential, and forever incremental backup
Replication Network	Maximum number of replication targets	4 (A maximum of 1:4 replication network is supported.)
	Maximum number of replication sources	16 (A maximum of 16:1 replication network is supported.)
	Bidirectional replication	Supported
Data Reduction	Link transmission data reduction	Data compression
	Storage space data reduction	Deduplication and compression
Replication Network	Bandwidth limit	Supported
Recovery Using Replication Copies	Data recovery to the replication target end	Supported
	Data recovery to the replication source	Replication copies are reversely replicated to the source end for recovery.

Archive Network	Network quality	BBR-based transmission protocol optimization, tolerating the worst network quality with a TTL ≤ 100 ms and a 5% packet loss rate
	Bandwidth limit	Supported

OceanProtect allows users to replicate copies immediately or periodically. OceanProtect supports link compression, target-end data compression and deduplication, and bandwidth limit for replication jobs.

**Figure 5-16** Copy replication process



The backup copy replication function of Huawei OceanProtect with data backup features applies to scenarios such as copy-level disaster recovery and remote application migration. Only copies are replicated. Other information such as archive logs is not replicated. Replication copies at the target end can be used in the same way as those at the source end. The underlying remote replication capability of storage is used to implement asynchronous replication of file systems. This mechanism reduces the overall replication performance loss caused by backup software read and write operations and improves the replication bandwidth.

#### NOTE

The replication jobs use asynchronous replication of file systems.

## 5.5 Copy Data Anonymization

To ensure the security of enterprises' backup data, Huawei OceanProtect data backup features provide the copy data anonymization capability. This section describes the data anonymization principle.

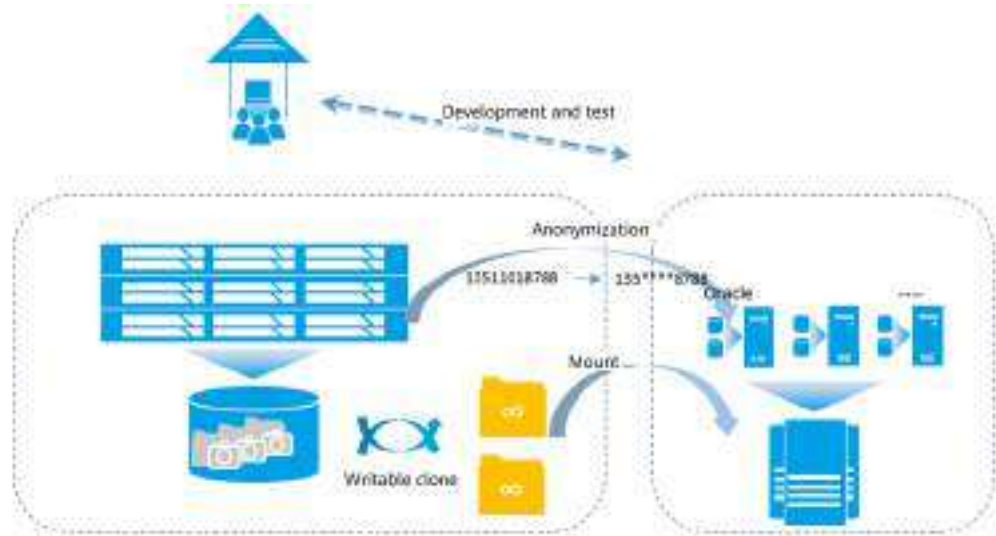
### 5.5.1 Data Anonymization

Live mount greatly improves the usage efficiency of enterprise development and test data. If data in a copy identical with the original production data is used in development and test scenarios, privacy data may be leaked. To prevent privacy data leakage, Huawei OceanProtect data backup feature provides the data anonymization function for database backup copies. It can anonymize data in

copies offline and mount the anonymized copies for tests. This ensures that private data in copies from being leaked and user data is secure and reliable.

Offline data anonymization is implemented as follows: 1. Mount a copy to be anonymized to a database. 2. Log in to the target database and anonymize the data based on the anonymization rule. 3. Verify the anonymization result. If the anonymization result is correct, generate an anonymization report. The persistent data in the target copy has been anonymized.

**Figure 5-17** Offline anonymization principles



To facilitate sensitive data anonymization, Huawei OceanProtect data backup feature provides the following built-in typical anonymization rules:

Type	Description	Original Data	Anonymized Data
Noise-Adding	Increases or decreases the original value by a random number.	10,000	1,053,000
Fixed-Number	Replaces the original data with a fixed number.	8088	9527
Partial-Mask1	Masks some contents.	15935124518	159****4518
Partial-Mask2	Masks some contents.	50024317820701 125x	500*****25x
Shuffling	Randomly change positions.	I am ok.	ok. I am
Full-Mask	Masks all of the original data.	Secret	*****

Numeric-Range	Replaces the original data with a random number within a specified range.	8088	1111
PII-Type	Overwrites the original data with the PII type.		
Format-Preserving	Uses an encryption algorithm to encrypt data. The length of the ciphertext is the same as that of the original text.	Wonderful	Da2CB2VxM

# 6 System Performance Design

To meet the requirements of high throughput and data reduction ratio, the number of CPU cores of an OceanProtect backup storage controller is the largest in the industry, providing powerful computing capabilities.

OceanProtect implements host accessing balancing and software optimization on all I/O paths, including the front-end network, CPU, and back-end network, and fully utilizes its powerful hardware capabilities to provide customers with ultra-high bandwidth. [Table 6-1](#) describes the key performance design based on the I/O delivery process from the host to SSDs/HDDs for addressing the current problems and pain points.

**Table 6-1** Key performance design

I/O Processes	Challenge	Key Design	Performance Design Principle
Front end	The native Ethernet protocol involves multiple layers and consumes a large number of CPU resources, leading to limited maximum performance.	Direct TCP/IP Offloading Engine (DTOE), a technology optimized by Huawei, saves CPU resources for other jobs, improving the maximum system throughput.	<ul style="list-style-type: none"><li>I/Os bypass the kernel mode to reduce cross-mode overheads.</li><li>Protocols are offloaded to hardware, reducing CPU usage.</li></ul>

I/O Processes	Challenge	Key Design	Performance Design Principle
Controller	How to make full use of the computing capability of multi-core CPUs	The intelligent multi-core technology reduces the resource conflict probability and CPU scheduling overheads, improves the I/O processing efficiency of CPU resources, thereby improving the maximum system throughput.	<ul style="list-style-type: none"><li>• I/Os are distributed among CPUs by CPU group to reduce the latency of cross-CPU scheduling.</li><li>• A CPU is divided into different partitions based on services to reduce service interference.</li><li>• No lock is designed in the service partition to reduce lock conflicts.</li></ul>
	Huawei-developed efficient fingerprint algorithm	With the multi-channel concurrency technology, multiple CPU instructions are invoked simultaneously.	Multi-buffer technology: The algorithm performs two SHA1+CRC32 calculations at the same time and two buffers are provided. Two instruction streams are mixed together to perform the two calculations concurrently.
Backend	Write amplification causes short SSD life time and low write performance.	The multi-stream technology reduces write amplification of garbage collection and improves SSD write bandwidth.	Hot and cold data is separated to reduce write penalty within disks.
		ROW full-stripe write reduces write amplification of EC verification and improves system write bandwidth.	The ROW full-stripe write design reduces random write amplification.

I/O Processes	Challenge	Key Design	Performance Design Principle
Optimization for backup scenarios	Deduplication is not user-friendly for sequential read/write of HDDs.	Fingerprint deduplication is separated from address indexing to ensure that the data layout on disks after deduplication and compression is consistent with the LBA of the file.	The ID-centric three-layer metadata architecture solves the problem of poor disk read performance caused by disordered data distribution on HDDs due to deduplication.
	After full backup is performed for multiple times, the latest full backup is scattered, affecting recovery performance of the latest backup job.	Data locality rearrangement is performed for large data blocks scattered due to chunking and deduplication.	Data locality rearrangement against data dispersion

- 6.1 Front-end Network Optimization
- 6.2 Intra-Controller Optimization
- 6.3 Back-end Network Optimization
- 6.4 Backup Software Performance Optimization
- 6.5 Backup Media Performance Optimization
- 6.6 Backup Performance Monitoring

## 6.1 Front-end Network Optimization

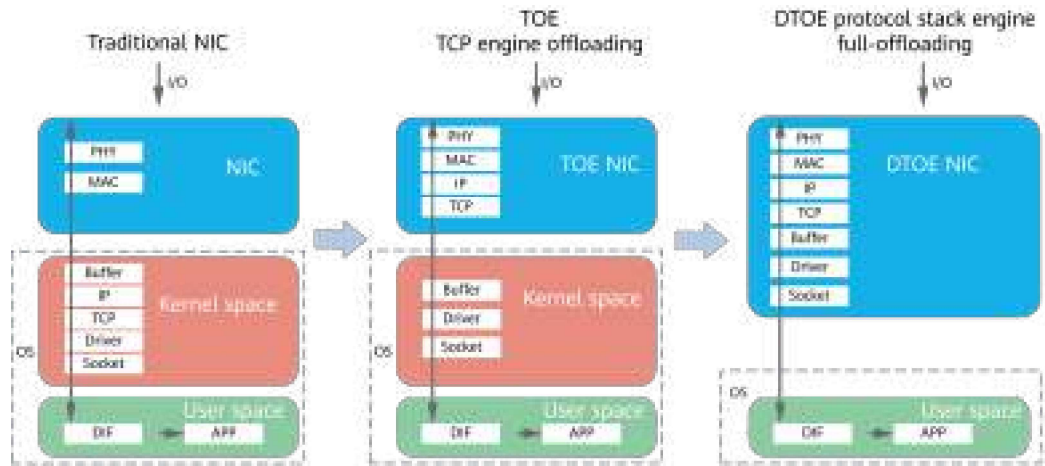
Front-end network optimization mainly refers to the optimization of latency between applications and storage devices, including protocol offloading optimization in NAS and iSCSI scenarios and scheduling optimization in common scenarios.

### Protocol Offloading

The performance bottleneck of network adapters (NAS/iSCSI) lies in the long I/O paths. The overhead of TCP and IP protocols is extremely high. Huawei uses the

deeply optimized user-mode TCP/IP and iSCSI protocol stacks and TOE-enabled network adapter passthrough to offload protocol computing to interface modules. This releases computing resources of CPUs for other jobs and improves maximum system performance, as shown in **Figure 6-1**:

**Figure 6-1** DTOE technology



If controllers use traditional NICs, network protocol stacks processed by the controllers have deep layers. As a result, every time a data packet is processed, an interruption is triggered, causing high CPU overhead.

By using the TOE technology, NICs offload the TCP and IP protocols. An interruption is triggered after an application implements a complete data processing, which significantly reduces the interruption overhead. However, in this case, some drivers are running in the kernel mode, resulting in the latency caused by the overhead of system calls and thread switchover between user mode and kernel mode.

The DTOE technology adopted by OceanProtect offloads IP data paths to NICs. Transport layer processing, including the DIF check function, is moved to the NIC microcode customized by Huawei, eliminating the CPU overhead. In addition, the working thread in the system checks the receiving queue of the interface module periodically in polling mode. If there is a request, the thread processes the request immediately. This reduces the latency overhead caused by waking up the working thread after a request is received.

## 6.2 Intra-Controller Optimization

### 6.2.1 Intelligent Multi-Core Technology

OceanProtect contains more CPUs and cores than any other backup products in the industry to provide powerful computing capabilities. The intelligent multi-core technology reduces the performance deterioration caused by cross-CPU access in multi-core and multi-CPU architectures, fully utilizes the powerful computing capability of CPUs, and implements linear increase of system performance as the number of CPUs increases. The intelligent multi-core technology consists of CPU grouping, service grouping, and lock-free design between cores.

### 6.2.1.1 vNode Processing Domain

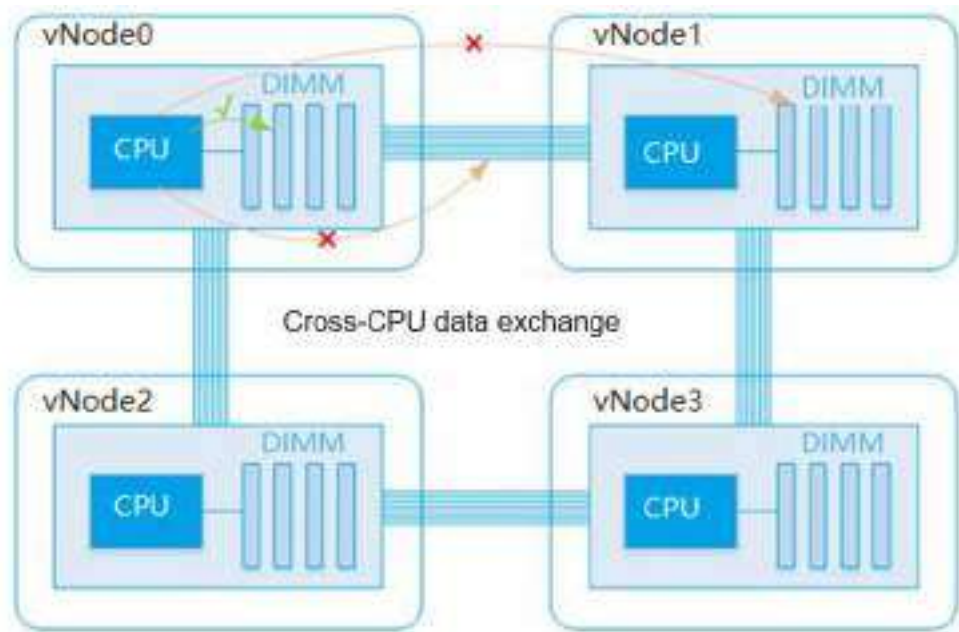
#### vNode Domain

For better scalability, a storage system is divided into multiple vNodes. Each vNode is a logical resource object and corresponds to a CPU as well as memory resources directly accessed by the CPU. Each CPU and its local memory are allocated to the same vNode so that the CPU can efficiently access the memory.

A vNode is also a service processing unit. A service should be processed within one vNode, and the relationship between cache mirroring copies is also established between vNodes.

#### No Cross-CPU Access

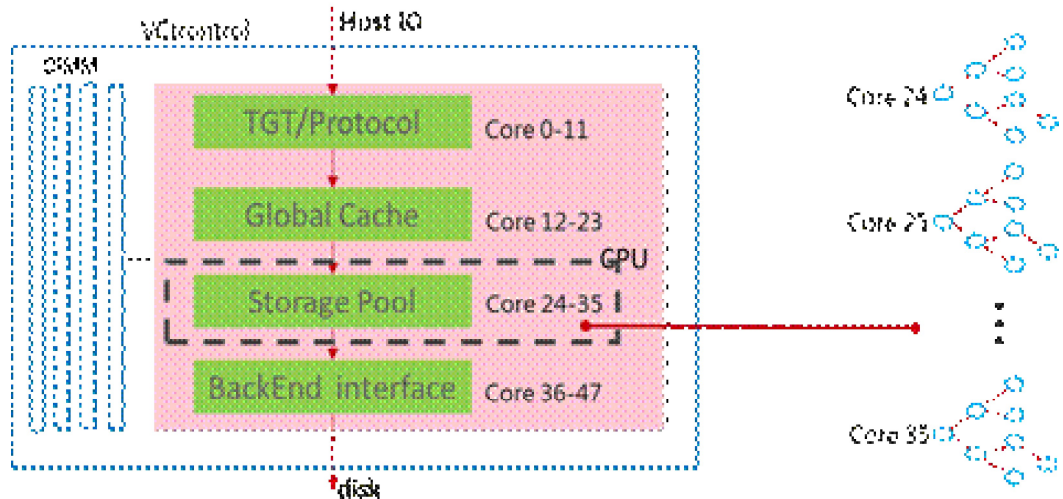
A vNode contains physical resources such as the CPU and memory. The memory is the local memory that can be directly accessed by the CPU. This means that the CPU does not need to access remote memory (which would involve higher latency). I/O requests from hosts are distributed to vNodes based on the intelligent distribution algorithm and are processed in the vNodes from end to end. This eliminates the overhead of communication across CPUs and accessing the remote memory, as well as conflicts between CPUs, allowing performance to increase linearly with the number of CPUs. In the following figure, vNode0 minimizes its access to the memory of other vNodes, and requests of vNode0 are rarely forwarded to other vNodes through the communication channels between CPUs.



### 6.2.1.2 Lock-free Design Between Cores

In a service group, each core uses an independent data organization structure to process service logic. This prevents the CPUs in a service group from accessing the same memory structure, and implements lock-free design between CPU cores.

The following figure shows an example. The CPU cores 24-35 match the storage pool service group and run only the storage pool service logic. In the storage pool service group, services are allocated to different cores, which use independent data organizations to prevent lock conflicts between the cores.



### 6.2.1.3 Intelligent Dynamic Load Balancing of CPUs

The traditional CPU grouping technology can solve the collision domain problem of each service, but also brings the problem of unbalanced resource usage between CPU groups in different scenarios. The dynamic load balancing technology of OceanProtect defines differentiated scheduling policies based on the computing overheads of jobs. In this way, jobs are balanced among the CPU core groups. High-density computing jobs are distinguished from common density computing jobs and are used as scheduling units for load balancing among CPU groups. This prevents scheduled jobs from interfering with each other, improving job execution efficiency.

Figure 6-2

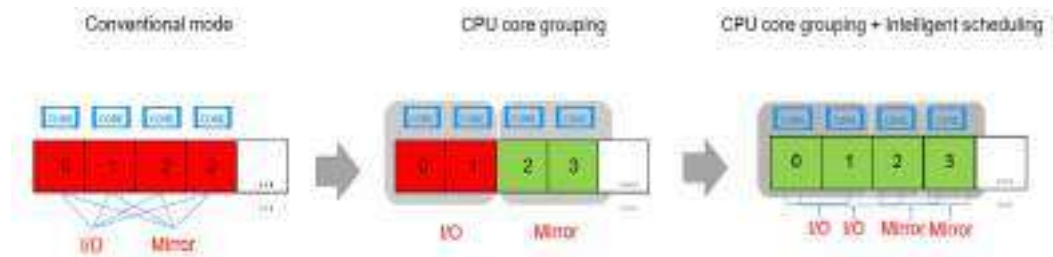
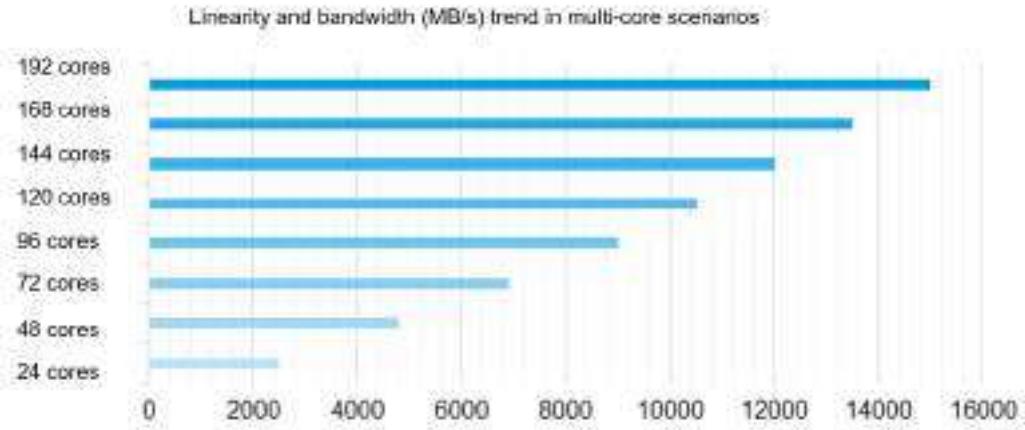


Table 6-2 Scheduling mode comparison

	Traditional Mode	Fixed CPU Core Grouping	CPU Core Grouping + Intelligent Scheduling
Advantages	All CPU cores can be fully scheduled and utilized, making this mode suitable for scenarios where services change frequently.	Job grouping prevents interference between different jobs, reduces frequent resource switching across CPU cores, and improves CPU processing efficiency.	Jobs are allocated to proper CPU cores according to the load status of CPU cores and the group attributes of the jobs. In this way, load balancing among CPU cores is implemented and CPU resource waste is obviously reduced.
Disadvantages	Different jobs compete for time slices of different CPU cores, and high lock conflicts cause frequent switching of data I/Os between different cores, leading to low CPU computing efficiency, high latency, and unstable performance.	CPU resource groups are bound to jobs, which cannot adapt to scenarios where services change frequently. As a result, some CPU resources are idle and wasted.	Development is difficult.

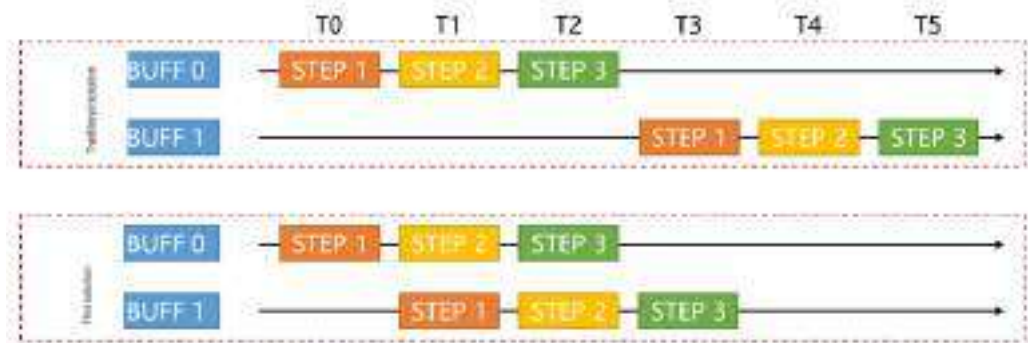
The vNode, service grouping, lock-free, and intelligent dynamic CPU load balancing technologies enable system performance to increase in quasi-linear mode with the number of controllers, CPUs, and CPU cores. The following figure shows the tested performance as the number of CPU cores increases.



## 6.2.2 Huawei-developed Efficient Fingerprint Algorithm

The fingerprint NEON instruction is used, the code logic is optimized, and multiple instruction streams are mixed through software and hardware to calculate fingerprints at the same time.

Multi-buffer technology: Two types of hash calculations are performed at the same time and two buffers are provided. Two instruction streams are mixed together to perform the two calculations concurrently. In this case, arithmetic logic units (ALUs) in CPUs can be more fully used, improving performance. The multi-buffer technology achieves nearly twice computing bandwidth with the same computing resources.



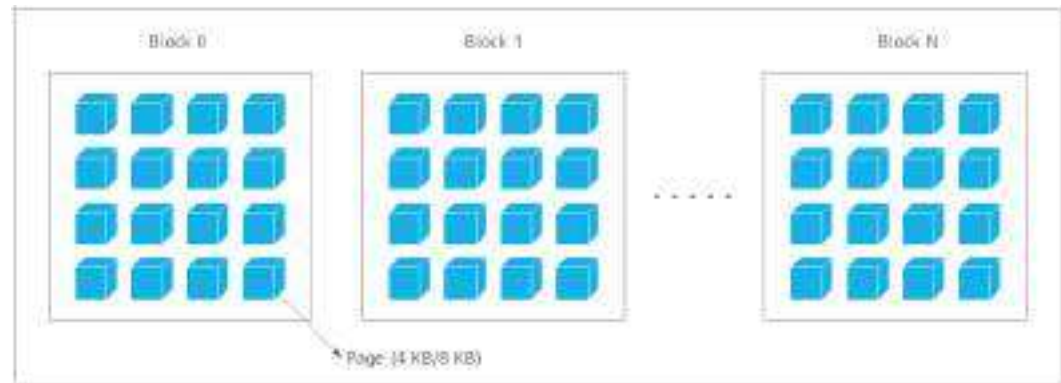
## 6.3 Back-end Network Optimization

### 6.3.1 Multistreaming

SSDs use NAND flash. The following figure shows the logical structure. Each SSD consists of multiple NAND flash chips, each of which contains multiple blocks. Each block further contains multiple pages (4 KB or 8 KB). The blocks in NAND flash chips must be erased before being written. Before erasing data in a block, the system must migrate valid data in the block, which causes write amplification in the disk.

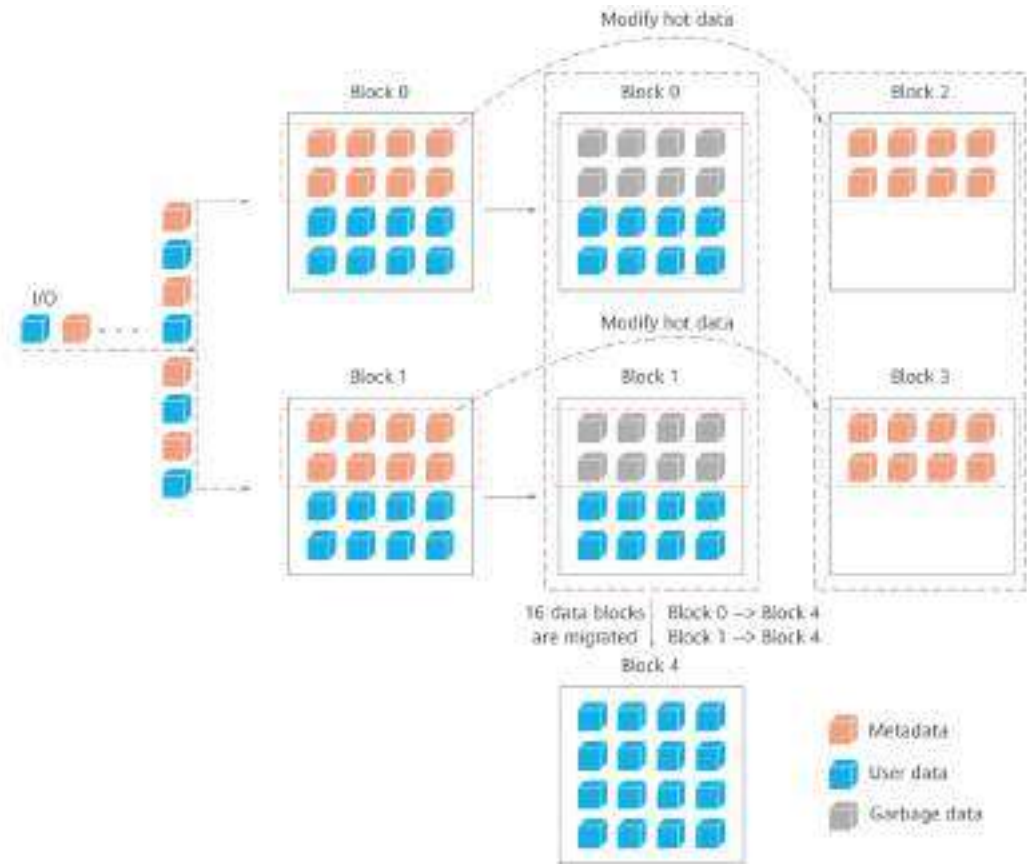
The multistreaming technology classifies data and stores different types of data in different blocks. There is a high probability that data of the same type is valid or garbage data at the same time. Therefore, this technology reduces the amount of data to be migrated during block erasure (one read and one write for each block migration) and minimizes write amplification, improving the performance and service life of SSDs.

**Figure 6-3** Logical structure of an SSD

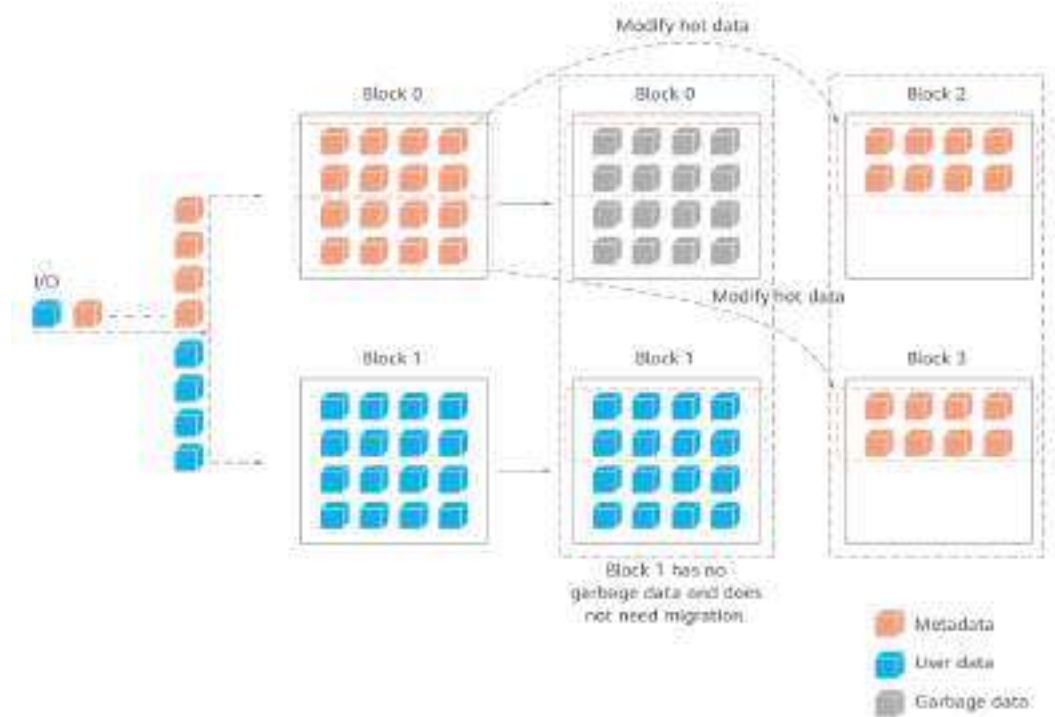


In an ideal situation, garbage collection would expect all data in a block to be invalid so that the whole block could be erased without data movement. This would minimize write amplification. But this is rarely the case. Usually different data in a storage system is updated less or more frequently. This is also referred to as cold and hot data. For example, metadata (hot) is updated more frequently and is more likely to cause garbage than user data (cold). The multistreaming technology enables SSD drivers and controllers to work together to store hot and cold data in different blocks. This increases the possibility that all data in a block is invalid, reducing valid data to be migrated during garbage collection and improving SSD performance and reliability. [Figure 6-4](#) shows data migration for garbage collection before separation of hot and cold data, in which a large amount of data needs to be migrated. [Figure 6-5](#) shows data migration for garbage collection after separation of hot and cold data, in which less data needs to be migrated.

**Figure 6-4** Data migration for garbage collection before separation of hot and cold data



**Figure 6-5** Data migration for garbage collection after separation of hot and cold data



OceanProtect separates hot and cold data into three types: metadata, newly written data, and valid data that must be migrated for garbage collection. Metadata is the hottest. Newly written data is migrated for garbage collection if it is not modified for a long period. The migrated data has the lowest probability of being modified and is thus considered the coldest. Data separation reduces SSD write amplification, greatly streamlines garbage collection, improves write performance, and reduces the amount of data written to SSDs to extend the SSD lifespan.

### 6.3.2 Large-Block Sequential Write

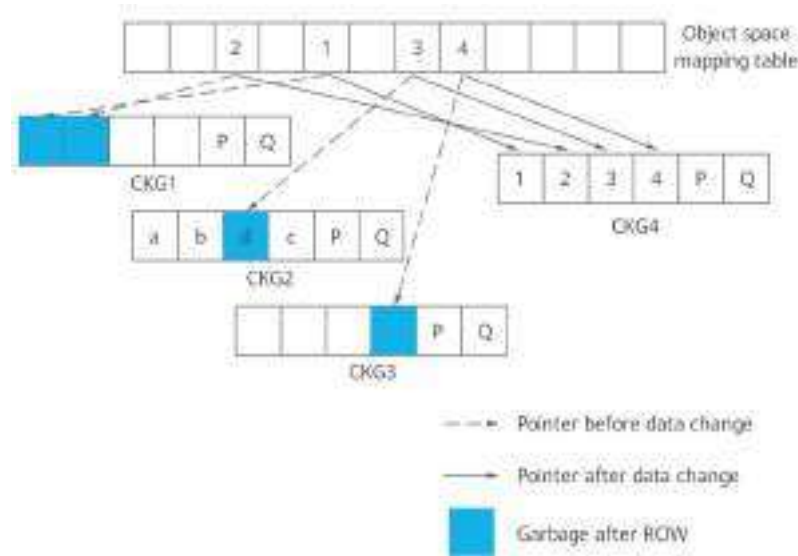
Most I/Os of backup jobs are sequential write I/Os. However, in scenarios where backup data is used to directly start services, most I/Os of applications are random write I/Os. For the random write model, traditional RAID write processes generate several times of write amplification penalties, causing write performance to deteriorate greatly.

OceanProtect uses the Redirect On Write (ROW) large-block sequential write for all writes including new data writes and data modifications, avoiding RAID write penalty caused by data reads and modification write verification required in traditional RAID. This greatly reduces the overhead on controller CPUs and read/write loads on SSDs in write processes. In addition, ROW allocates a new flash chip for each write to balance the number of erasure times of each flash chip. In this way, flash chips do not wear out due to repeated erasures in a location, avoiding fast faults on an SSD.

Compared with the traditional RAID overwrite mode, the ROW large-block sequential write mode eliminates the RAID write penalty and enables various

RAID levels to achieve high performance. When using OceanProtect, users only need to consider data reliability requirements of services and do not need to consider the impact of different RAID groups on performance. In this way, performance of RAID-TP can meet commercial requirements.

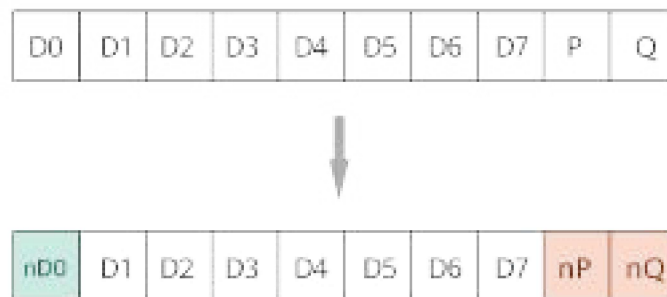
**Figure 6-6** ROW large-block sequential write



In the above figure, the system uses RAID 6 (4+2) and writes new data blocks 1, 2, 3, and 4 to modify existing data. In traditional overwrite mode, a storage system must modify every chunk group where these blocks reside. For example, when writing data block 3 to CKG2, the system must first read the original data block d and the parity data P and Q. Then it calculates new parity data P' and Q', and writes P', Q', and data block 3 to CKG2. In ROW full-stripe write, the system uses the data blocks 1, 2, 3, and 4 to calculate P and Q and writes them to a new chunk group. Then it modifies the logical block addressing (LBA) pointer to point to the new chunk group. During this process there is no need to read any existing data.

For traditional RAID, for example, RAID 6, when D0 is changed, the system must first read D0, P, and Q, and then write new nD0, nP, and nQ. Therefore, both the read and write amplifications are 3. Generally, the read and write amplification of random I/Os in traditional RAID ( $xD+yP$ ) is  $y+1$ .

**Figure 6-7** Write amplification of traditional RAID 6



The following table lists the write amplification statistics of various traditional RAID levels.

**Table 6-3** Write amplification of traditional RAID levels

RAID Level	Write Amplification of Random Write I/Os	Read Amplification of Random Write I/Os	Write Amplification of Sequential Write I/Os
RAID 6 (14D+2P)	3	3	1.14 (16/14)
RAID-TP (not available in traditional RAID)	-	-	-

Typically, RAID 5 uses 11D+1P, RAID 6 uses 22D+2P, and RAID-TP uses 21D+3P, where D indicates data columns and P, Q, and R indicate parity columns. The following table compares write amplification on OceanProtect using these RAID levels.

**Table 6-4** Write amplification in ROW large-block sequential write

	Write Amplification of Random Write I/Os	Read Amplification of Random Write I/Os	Write Amplification of Sequential Write I/Os
RAID 6 (22D+2P)	1.09 (24/22)	0	1.09
RAID-TP (21D+3P)	1.14 (24/21)	0	1.14

On OceanProtect, performance differences are within 5% between RAID 6 and RAID 5 and between RAID-TP and RAID 6.

OceanProtect uses larger RAID stripes to improve RAID space utilization while ensuring data reliability. The following figure compares the RAID space utilization of traditional backup storage and OceanProtect.

**Table 6-5** RAID space utilization comparison

	Traditional RAID		OceanProtect RAID	
RAID Level	Maximum RAID Stripe	Space Utilization	Maximum RAID Stripe	Space Utilization
RAID 6	12+2	85.7%	23+2	92%
RAID-TP	N/A	N/A	22+3	88%

OceanProtect can provide higher data reliability and space utilization.

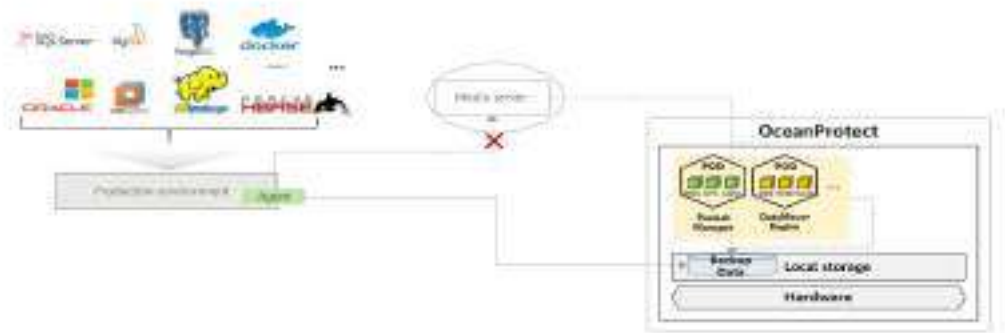
## 6.4 Backup Software Performance Optimization

### 6.4.1 Data Passthrough to Storage

In traditional backup, agents or proxies are usually used to capture production data to be backed up. The captured data is transmitted to the media server for encryption and slicing, and then the processed data is stored in the storage system. In such process, the I/O path is too long due to unnecessary processing performed, causing performance deterioration or large-scale expansion of media servers.

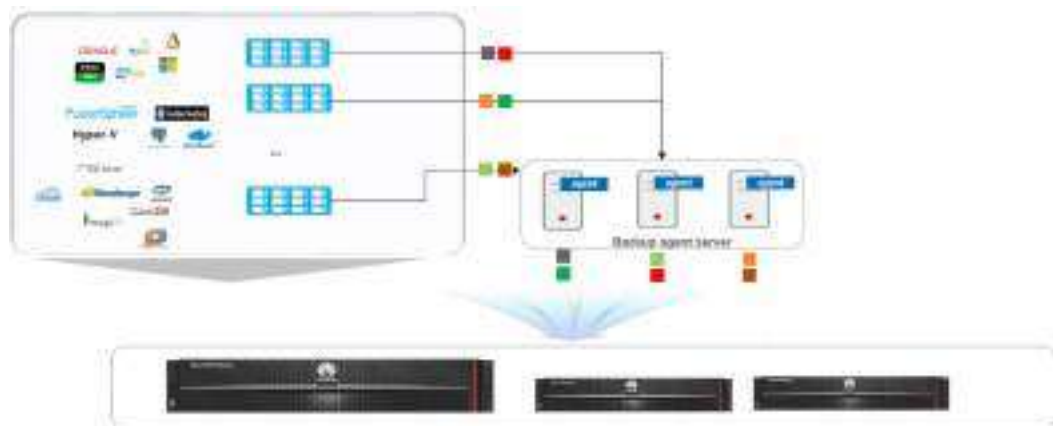
OceanProtect enables data passthrough to storage. Backup data is captured using agents or proxies. The captured data is written into the all-in-one storage device in native format. The storage device then encrypts and deduplicates the data. In this way, the number of data transmission times is reduced, eliminating unnecessary processing and improving backup efficiency.

**Figure 6-8** Data passthrough to storage



### 6.4.2 Distributed Concurrent Stream Backup

**Figure 6-9** Multi-node distributed concurrent stream backup, implementing backup and recovery for PB-level ultra-large-scale clusters



OceanProtect data backup features are designed for applications with a large amount of data. To improve the backup bandwidth, a job is split into multiple subjobs and allocated to multiple nodes. In this way, the computing and transmission capabilities of multiple nodes are fully utilized to meet the backup performance requirements of PB-level data.

Concurrent backup of virtualization, cloud, or container applications: A VM, container, or cloud host job is split into multiple disk subjobs and distributed to multiple backup agent nodes by disk subjob.

Concurrent backup of big data ecosystem applications: A job is split into multiple subjobs at the file level (HDFS) or table level (HBase/Hive) and allocated to multiple backup agent nodes by subjob.

DWS backup: Jobs are distributed to multiple OceanProtect devices by backup object generated by nodes.

## 6.5 Backup Media Performance Optimization

Based on the I/O model of multi-channel large-block sequential writes + large-block sequential reads, OceanProtect implements end to end optimizations within controllers. Large granularity data chunking reduces the amount of metadata to be stored per unit space, and improves the processing efficiency for metadata in large-bandwidth sequential reads and writes. The metadata processing overhead on the CPU reduces so that more overheads can be used for other services, improving performance. In the backup scenario with a high data reduction ratio, the proportion of the metadata space to the total storage space greatly decreases, saving much storage space.

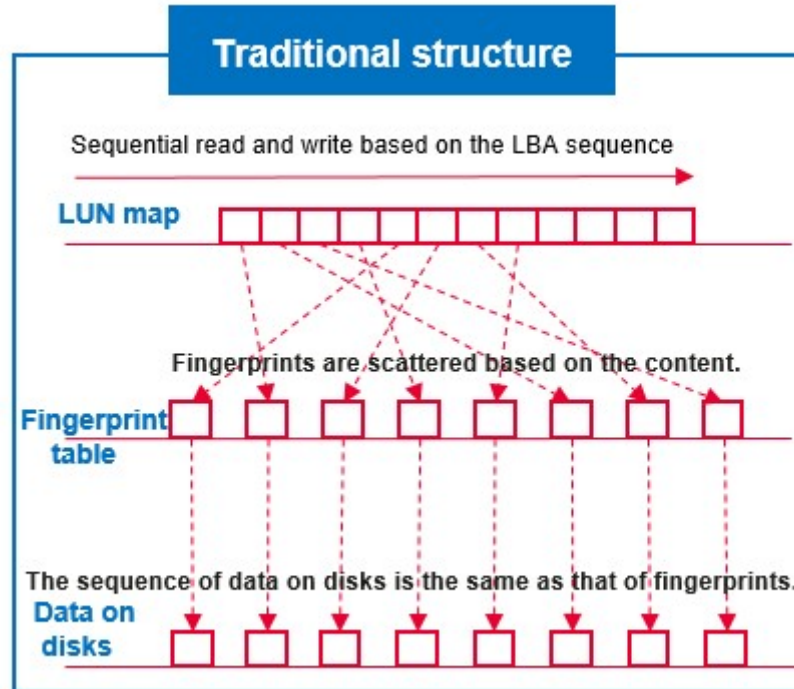
For the application of the variable-length chunking algorithm, larger continuous input data blocks help achieve better deduplication effect. However, the data deduplication ratio does not linearly increase with the block size. Based on much experimental data analysis, 4 MB data blocks are used for variable-length chunking of OceanProtect. Based on the sequential write characteristics of backup applications, sequential write service flow data can be aggregated by using the buffer effect of the write cache, and 4 MB data blocks aggregated in the cache are transmitted to the variable-length chunking module at a time. Then, the variable-length chunking module performs data chunking calculation to achieve a high data deduplication ratio. For non-sequential write requests, data is not aggregated into 4 MB data blocks. At the same time, writes of large data blocks are less than those of small data blocks within unit time, reducing overheads and improving the back-end and write performance.

### 6.5.1 Optimization for Backup and Recovery Jobs

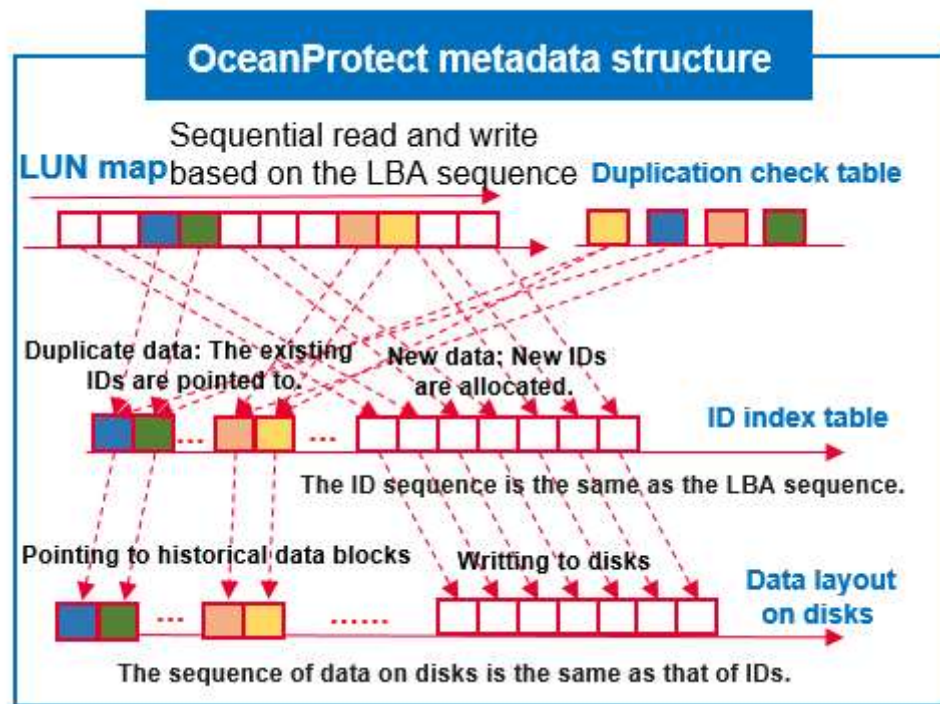
Generally, I/Os are written based on the logical address sequence of the files generated by a backup job. During recovery, most I/Os are read based on the logical address sequence of the files. The sequence of reading data is basically the same as that of writing data. OceanProtect backup storage provides a data layout architecture based on the feature of backup software.

In the traditional deduplication architecture, data is distributed based on the fingerprint table of data blocks. After being deduplicated, sequential data is scattered and stored based on fingerprints. Data on disks is stored based on

fingerprints, which is irrelevant to the LBA sequence. In this case, the sequence of service data is affected. This mode is suitable for the all-flash form. However, if HDDs are used as storage media, the random access performance is poor, which affects the deduplication and recovery performance of written data.



When a backup job is executed, data is written based on the logical address sequence of generated backup files. During recovery, data is read based on the logical address sequence of the backup files. The sequential read/write performance of HDDs is high, but the random read/write performance is low. The ID-centric three-layer metadata structure is designed to solve the problem of poor HDD read performance caused by disordered data due to deduplication.



Backup jobs based on backup software involve sequential writes, recovery jobs involve sequential reads. Therefore, ID metadata is introduced. ID metadata is applied for and allocated in batches to ensure that the ID allocation sequence is consistent with the LBA sequence. Fingerprint deduplication is separated from address indexing. The duplication check table and ID index table are provided. The ID and LBA are in the same sequence, and the data layout on disks is synchronized with IDs. This ensures that the metadata and data stored on HDDs are in the same sequence as that of host accesses.

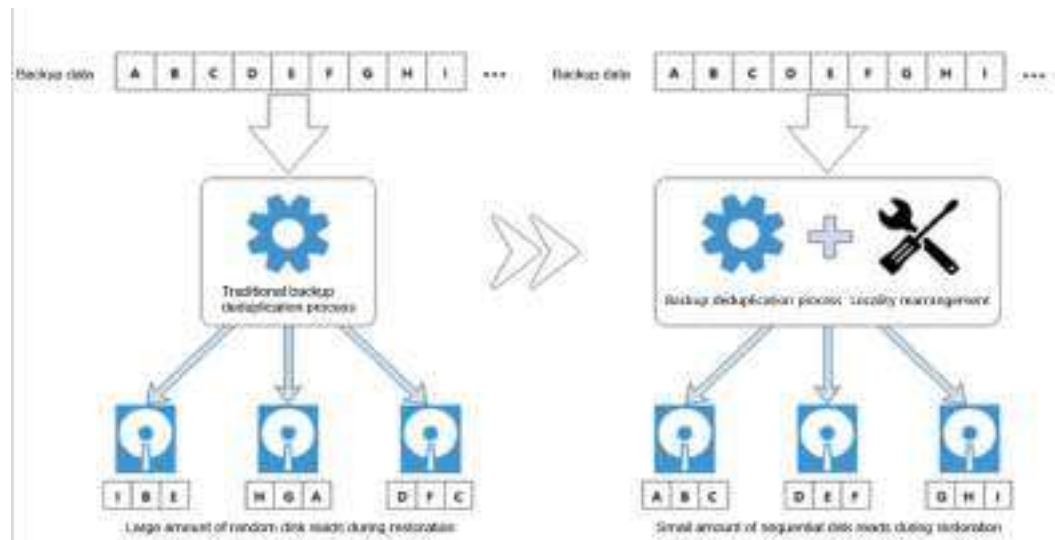
As shown in the preceding figure, after variable-length chunking is performed for file-level sequential write data, fingerprint calculation and duplication check is performed. For duplicate data, the LBA directly points to an existing ID. New data blocks are sorted based on the LBA address and then new IDs are allocated. The new IDs increase in sequence and are never reused. New data is compressed and then written to disks. New data is arranged on storage media based on the sequence of new data IDs. According to the end-to-end result, the new data blocks (sequential data written by the backup software after being deduplicated) are also saved in sequence on the storage media. This design has two important advantages: 1. Backup performance is improved. I/Os are sequentially written by the storage system to storage media. The high bandwidth of HDDs improves performance of the entire system. 2. The backup and recovery performance is optimized. The sequence of new data recovery by backup software is changed to sequential read of large data blocks on the storage media after ID conversion. A large segment of sequential data is read at a time because the data on the storage media is arranged based on the write sequence of the backup software. If excessive data is not required for the I/O read, it is stored in the read cache. However, as long as the recovery jobs of the backup software are not terminated, there is a high probability that the excessive data can be hit by the subsequent sequential read requests of the backup software within a short period of time, improving recovery performance.

## 6.5.2 Recovery Performance Improvement

As the number of backup times increases and due to the deduplication relationships during incremental backup and full backup, deduplicated data of the backup data that is written later is more scattered physically. As a result, recovery read performance deteriorates gradually. The latest backup data is most likely to be used for recovery due to the application scenarios of backup and recovery, but recovery performance is low due to scattered data on HDDs. Therefore, the recovery duration is unacceptable in the case of data loss or unexpected situations. Huawei has developed the locality rearrangement algorithm to build the performance advantages of OceanProtect and the instant recovery function is provided for user data.

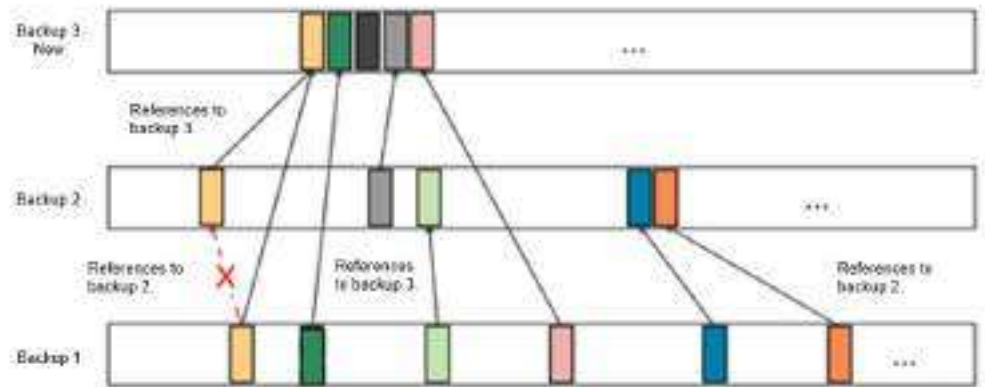
The latest backup data is usually deduplicated based on earlier backup data. Therefore, during recovery using the latest backup copy, data of the earlier backup copies is read randomly for multiple times. As a result, read performance of HDDs is low. Locality rearrangement aims to reduce the number of random read operations during backup and recovery, as shown in [Locality rearrangement effect](#).

**Figure 6-10** Locality rearrangement effect



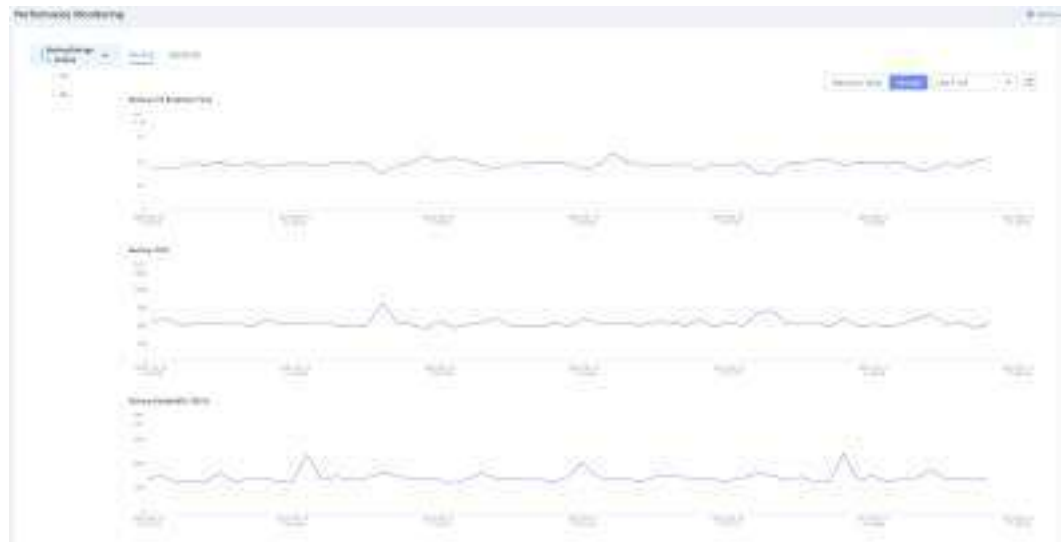
When backup data is written, the system determines whether to adjust the data distribution on disks based on how scattered the data is and how many random operations on disks are involved during recovery. Different from the traditional mode in which earlier backup data is referenced by the new backup data, locality rearrangement references earlier backup data to the new backup data and deletes the duplicate data blocks of the earlier backup data, as shown in [Figure 6-11](#). In this way, during recovery using the latest backup copy, only data of the latest backup copy and some data of earlier backup copies are sequentially read, reducing the impact of random reads on recovery performance.

**Figure 6-11** Data reference relationships after locality rearrangement



## 6.6 Backup Performance Monitoring

The data protection appliance supports monitoring of backup performance data.



# 7 System Reliability Design

---

OceanProtect provides a maximum of 99.9999% availability via data reliability and service availability designs. The active-active distributed architecture ensures that services are not interrupted and backup applications are not aware of faults of interface modules, controllers, and disks. This ensures that each backup job can be completed within a stable time window.

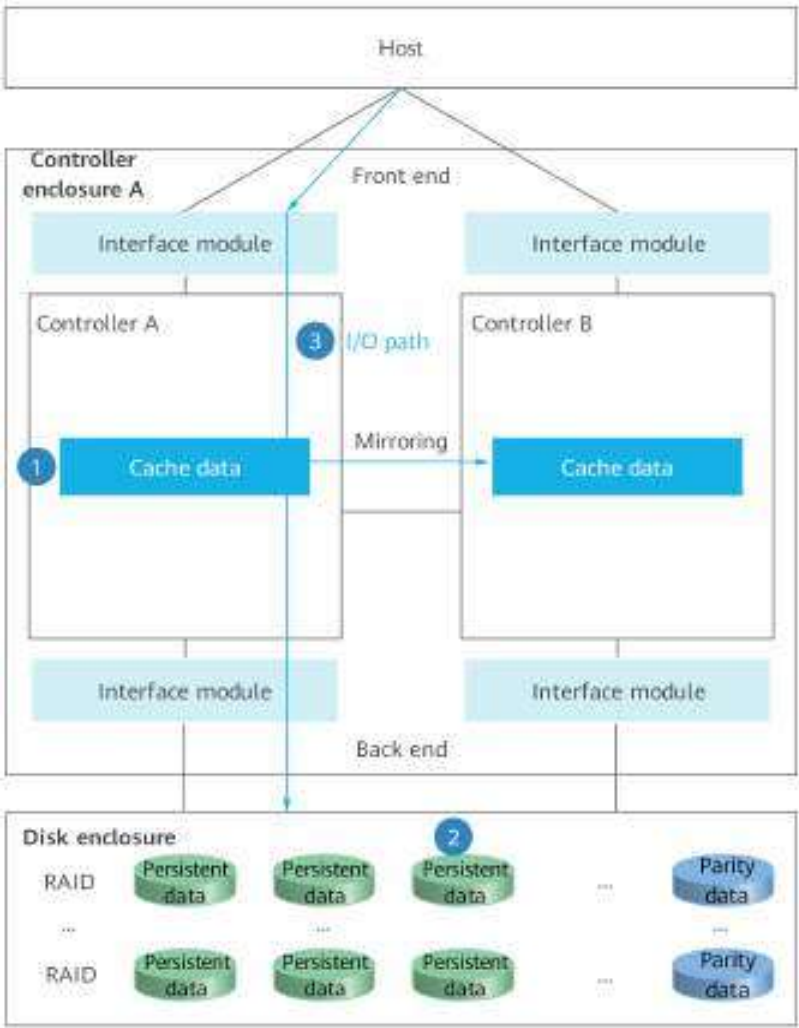
## [7.1 Data Reliability Design](#)

## [7.2 Service Availability](#)

## 7.1 Data Reliability Design

For data written by hosts to storage, OceanProtect undergoes three processes: data caching, data persisting on disks, and data transmitting on paths. The following describes data reliability measures in the three processes.

Figure 7-1 Data reliability panorama



## 7.1.1 Cache Data Reliability

To improve the data writing speed, OceanProtect provides the write cache mechanism. That is, after data is written to the memory cache of a controller and the mirrored copy, a success message is returned to the host and then cache data is written to disks in the background.

User data stored in the controller memory may be lost if the system is powered off or the controller is faulty. To prevent data loss, the system provides written data mirroring and power failure protection to ensure data reliability.

### 7.1.1.1 Written Data Mirroring

The full series of Huawei backup storage supports written data mirroring.

Data written by a host is first written to the battery-protected cache of the two controllers, and then a write success message is returned to the host. If a controller is faulty or reset, the mirrored dirty data on the other controller can be used to directly take over the dirty data and services of the faulty controller. This ensures that services are not interrupted and data is not lost.

### 7.1.1.2 Power Failure Protection

OceanProtect has built-in BBUs (backup power). If power outage occurs, BBUs in controllers provide extra power for moving cache data in the memory to the coffer. After the power supply is recovered, the system recovers the cache data in the coffer to the memory during system startup, preventing data loss. After detecting that a controller is removed, the software module on the controller uses the BBU (backup power) on the controller to back up user cache data to the coffer (space for storing data in the cache in case of power failures), ensuring data integrity. The process of moving cache data to the coffer upon power failures is implemented by the underlying system and does not rely on upper-layer software, thereby not affected by services and further improving user data reliability.

## 7.1.2 Persistent Data Reliability

OceanProtect uses the intra-disk RAID technology to ensure disk-level data reliability and prevents data loss. The RAID 2.0+ technology and dynamic reconstruction ensure system-level data reliability. As long as the number of faulty disks does not exceed that of the redundant disks, data will not be lost and the redundancy will not decrease.

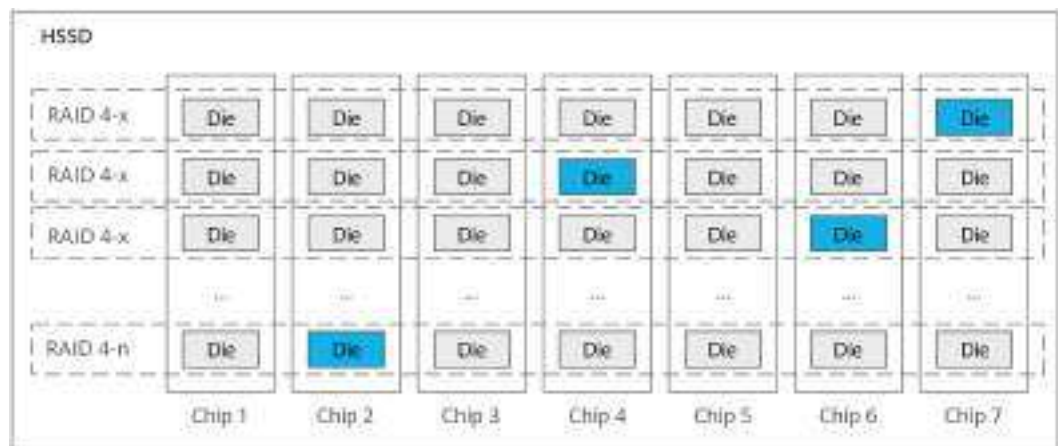
### 7.1.2.1 Intra-disk RAID

In addition to overall disk faults, regional damage may occur on the chips used for storing data. This is called the silent failure (bad block). These bad blocks do not cause the failure of an entire disk, but cause the data access failure on the disk.

Common bad block scanning can detect silently failed data in advance and recover the data. However, disk scanning occupies a large number of resources. To prevent impact on foreground services, the scanning speed must be controlled. Therefore, if the disk capacity and quantity are large, it takes weeks or even months to scan all disks. In addition, if both the bad block and disk failure occur in the interval between two scans, data may fail to be recovered.

Based on bad block scanning, OceanProtect uses HSSDs to provide the intra-disk RAID feature to prevent silent failures in scanning intervals. Specifically, RAID 4 groups are created for data on disks in the unit of dies to implement redundancy, tolerating the failure of a single die without any data loss.

**Figure 7-2** Intra-disk RAID on an SSD



### 7.1.2.2 RAID 2.0+

In a conventional RAID storage system that uses fixed physical disks in RAID groups, LUNs or file systems used by users are divided from the RAID groups. Because LUNs or file systems in a storage system are accessed at different frequencies, disks in some RAID groups may be heavily loaded and become hot spots, while those in other RAID groups may be idle. In addition, if a disk works for a longer time than others, its failure rate increases sharply and may be faulty in a shorter time than other disks. Therefore, hot disks in conventional RAID storage systems are at the risk of being overloaded.

With RAID 2.0+, OceanProtect divides each SSD into fixed-size chunks (CKs, generally 4 MB). CKs from different disks are joined into a chunk group (CKG) based on the RAID groups. RAID 2.0+ has the following advantages over traditional RAID:

- The service load is balanced to avoid hot spots. Data is evenly distributed to all disks in a storage resource pool, eliminating hotspot disks and lowering the disk failure rate.
- Fast reconstruction reduces the risk window. Faulty disks trigger data reconstruction on all the other disks in the storage pool. This many-to-many reconstruction is rapid and significantly reduces data vulnerability.
- All member disks in a storage resource pool participate in reconstruction, and each disk only needs to reconstruct a small amount of data. Therefore, the reconstruction process does not affect upper-layer applications.

#### 7.1.2.2.1 RAID for Disk Redundancy

RAID for disk redundancy leverages RAID 2.0+ to randomly select disks and evenly distribute data to selected disks. Each selected disk provides one chunk to form a chunk group. This ensures that no data is lost when a specific number of disks fails. Currently, the system supports RAID 5, RAID 6, and RAID-TP algorithms and allows users to customize the global hot spare disk mechanism.

- RAID 5 of a storage pool uses the EC-1 algorithm and generates one copy of parity data for each stripe. The failure of one disk in the storage pool can be tolerated.
- RAID 6 of a storage pool uses the EC-2 algorithm and generates two copies of parity data for each stripe. The failure of two disks in the storage pool can be tolerated.
- RAID-TP of a storage pool uses the EC-3 algorithm and generates three copies of parity data for each stripe. The failure of three disks in the storage pool can be tolerated.
- The global hot spare policy of a storage pool supports 0 to 8 disks.

#### NOTE

- The storage system uses RAID 2.0+ virtualization technology, so hot spare capacity is provided by all member disks in each storage pool. For ease of understanding, the hot spare capacity is expressed in the number of hot spare disks on DeviceManager.
- Even if the hot spare space is used up, the system can use the free space of the storage pool to rebuild data, ensuring storage system reliability.

### 7.1.2.3 Dynamic Reconstruction

Data reconstruction is not possible if the number of available member disks in a disk domain with conventional RAID is less than the number of member disks in a RAID group due to continuous disk faults or disk replacement. Guaranteeing user data redundancy is impossible without reconstruction. To cope with the preceding problems, OceanProtect uses dynamic RAID reconstruction. If the total number of available disks in a storage pool is less than the number of RAID member disks, the system retains the number of parity columns (M) and reduces the number of data columns (N) during reconstruction. After the reconstruction is complete, the number of member disks in the RAID group decreases, but the RAID redundancy level remains unchanged.

After the faulty disks are replaced, the system increases the number of data columns (N) based on the number of available disks in the storage pool, and new data will be written to the new N+M columns. Data that has been written during the fault will also be converted into the new N+M columns.

#### NOTE

Dynamic reconstruction reduces the total available capacity of the system. If multiple disks are faulty, handle the disk faults in time and pay attention to the storage pool usage.

### 7.1.2.4 Background Data Consistency Scanning

OceanProtect provides automatic and periodic background data consistency scanning. The scanning speed decreases as the host service pressure increase. The background data consistency scanning function uses the bitmap to scan the disk drive space recorded by the system in sequence. After data is read from the disk drive, the DIF verification, RAID stripe consistency verification, and matrix verification are performed. If any problem is detected, the system automatically restores the data in the background. If the recovery fails, a data damage alarm is reported. In this case, contact Huawei engineers to manually perform in-depth recovery.

This function is used to prevent silent failures that may occur during long-term data storage. You can enable or disable this function using related CLI commands.

## 7.1.3 Data Reliability on I/O Paths

During data transmission within a storage system, data passes through multiple components over various channels and undergoes complex software processing. Any problem during this process may cause data errors. If such errors cannot be detected immediately, error data can be written to persistent disks, calculated internally, or returned to the host, causing service exceptions.

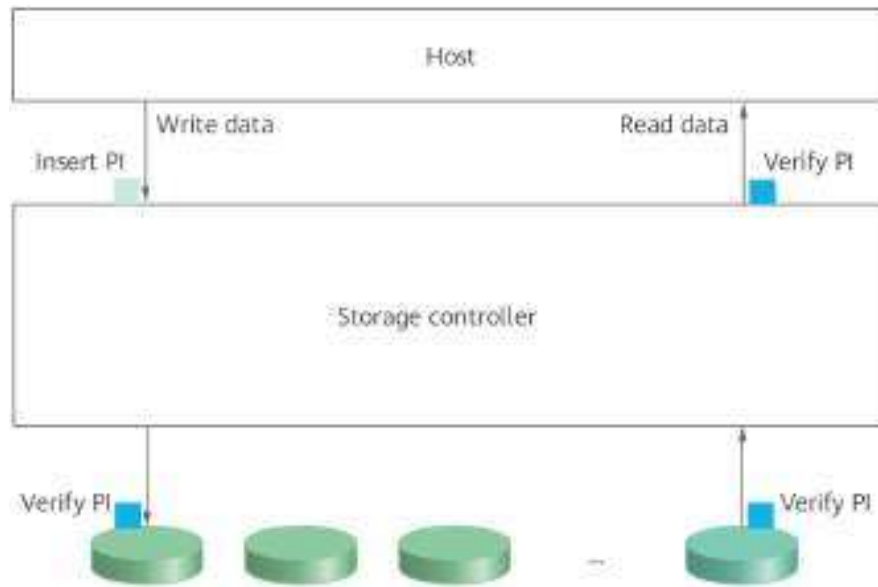
To resolve the preceding problems, OceanProtect uses the end-to-end protection information (PI) function to detect and correct data errors (internal changes to data) on the transmission path. The matrix verification function ensures that changes to the whole data block (the whole data block is overwritten by old data or other data) can be detected. The preceding measures ensure data reliability on I/O paths.

### 7.1.3.1 End-to-end PI

OceanProtect supports ANSIT10 PI. Upon reception of data from a host, the storage system inserts an 8-byte PI field to every 512 bytes of data before performing internal processing.

After data is written to disks, the disks verify the PI fields of the data to detect any change to the data between reception and flushing to the disks. In the following figure, the green point indicates that a PI is inserted to the data. The blue points indicate that a PI is calculated for the 512-byte data and compared with the saved 8-byte PI to verify data correctness.

**Figure 7-3** End-to-end PI

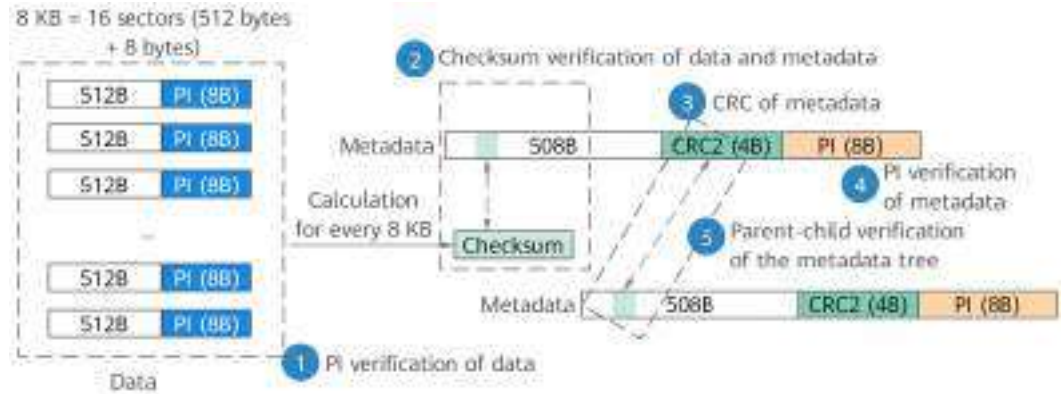


When the host reads data, the disks verify the data to prevent changes to the data. If any error occurs, the disks notify the upper-layer controller software, which then recovers the data by using RAID. To prevent errors on the path between the disks and the front end of the storage system, the storage system verifies the data again before returning it to the host. If any error occurs, the storage system recovers the data using RAID to ensure end-to-end data reliability from the front end to the back end.

### 7.1.3.2 Matrix Verification

Because the internal structure of disks is complex or the read path is long (involving multiple hardware components), various errors may occur due to software defects. For example, a write success is returned but the data fails to be written to disks; data B is returned when data A is read (read offset); or data that should be written to address A is actually written to address B (write offset). Once such errors occur, the PI check of the data is passed. If the data is still used, the incorrect data (such as old data) may be returned to the host.

**Figure 7-4 Parent-child verification**



OceanProtect provides matrix verification to cope with the write failure, read offset, and write offset that may occur on disks. In the preceding figure, each piece of data consists of 512-byte user data and 8-byte PI. Two bytes of the PI are used for cyclic redundancy check (CRC) to ensure reliability of the 512-byte data horizontally (protection point 1). The CRC bytes in 16 PI sectors are extracted to calculate the checksum, which is then saved in a metadata node. If offset occurs in a single or multiple pieces of data (512+8), the checksum of the 16 pieces of data is also changed and becomes inconsistent with that saved in the metadata. This ensures data reliability vertically. After detecting data damage, the storage system uses RAID redundancy to recover the data. This is matrix verification.

## 7.1.4 Automatic Cross-site Data Repair

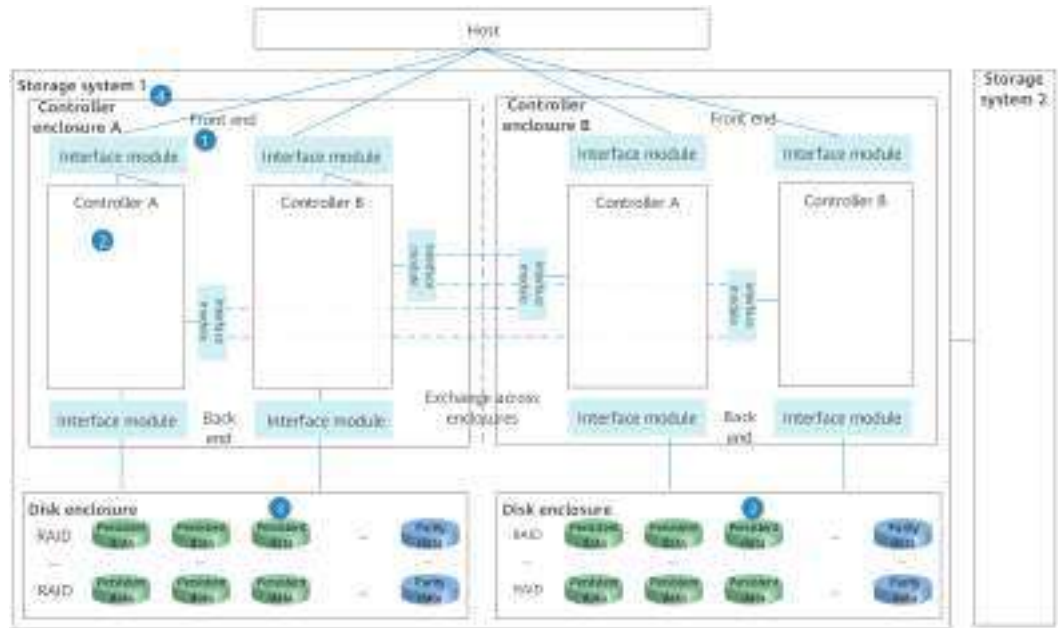
OceanProtect provides the remote replication function to ensure high data reliability. With remote data backup based on remote replication, OceanProtect can automatically restore local corrupted data by using remote backup data. This improves system data reliability and O&M efficiency.

When the end-to-end verification of a read request from a host fails or the background data consistency scanning detects a data damage and the data fails to be restored through local data redundancy, the system automatically reads backup data from the remote end using the remote replication function and returns the data to the host after the verification is successful. Then, the system uses the data read from the remote storage system to restore the damaged data on the local storage system. The entire process does not require manual intervention and does not affect host applications.

## 7.2 Service Availability

The storage system provides multiple redundancy protection mechanisms for the entire path from the host to the storage system. That is, when a single point of failure occurs on the interface module or link (1), controller (2), and storage media (3) that I/Os pass through, redundant components and fault tolerance measures can be used to ensure that services are not interrupted.

**Figure 7-5 Multi-layer redundancy and fault tolerance design**



## 7.2.1 Interface Module and Link Redundancy Protection

OceanProtect supports full redundancy. Link and interface module redundancy protection is provided for the front end for interconnection with the host, the back end for connecting disks, and the communication between controllers. The front-end interconnect I/O module (FIM) is connected to the host. If a controller is faulty or replaced, the connection between the host and the FIM is not interrupted and therefore, the connection will not be re-established. After the remaining controllers take over services, the FIM delivers retry or new I/Os to the controllers to ensure service continuity.

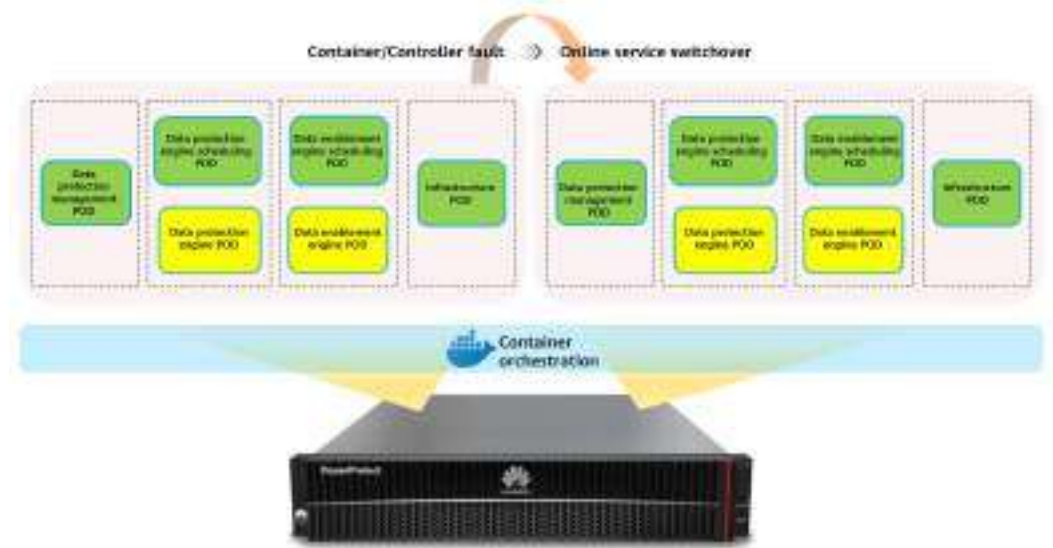
## 7.2.2 Controller Redundancy

OceanProtect provides redundant controllers to ensure reliability. In typical scenarios, cache data is stored on the current controller and the copy of cache data is stored on another controller. If a controller fails, services can be switched to the controller to which the cache data copy belongs, ensuring service continuity.

Container instance redundancy protection by using backup software

The backup service function of Huawei OceanProtect X6000 and X8000 is deployed in containers and redundant by controller. Backup service instances are deployed on each controller. When a service fault occurs in one controller, the backup service on the other controller takes over services.

**Figure 7-6 Node redundancy**



## 7.2.3 Storage Media Redundancy

OceanProtect not only ensures high reliability of a single disk, but also uses the multi-disk redundancy capability to ensure service availability if a single disk is faulty. That is, disk faults or sub-health is detected in a timely manner by using algorithms, and faulty disks are isolated in time to avoid long-term impact on services. Then, data of the faulty disk is recovered by using the redundancy technology. In this case, services can be continuously provided.

### 7.2.3.1 Fast Isolation of Disk Faults

When disks are running properly, OceanProtect monitors the in-position and reset signals. If a disk is removed or faulty, the storage system isolates it. New I/Os are written to other disks and a read success is returned to the host after I/Os are read by using RAID.

In addition, continuous, long-time operation causes disks to wear and increases the chance of particle failures. As a result, disks respond more slowly to I/Os, which can affect services. In this case, slow disks are detected and isolated in a timely manner so that they cannot further affect services.

A model that compares the average I/O service time of disks is built for OceanProtect based on common features of disks, including the disk type, interface type, and owning disk domain. With this model, slow disks can be detected and isolated within a short period of time, shortening the time when host services are affected by slow disks.

### 7.2.3.2 Disk Redundancy

OceanProtect supports three RAID configuration modes, which ensure service continuity in the event of disk failures. The system can tolerate simultaneous failure of at most three disks in a storage pool without any data loss or service interruption.

- RAID 5 uses the EC-1 algorithm and generates one copy of parity data for each stripe. The failure of one disk can be tolerated.
- RAID 6 uses the EC-2 algorithm and generates two copies of parity data for each stripe. The simultaneous failure of two disks can be tolerated.
- RAID-TP uses the EC-3 algorithm and generates three copies of parity data for each stripe. The simultaneous failure of three disks can be tolerated.

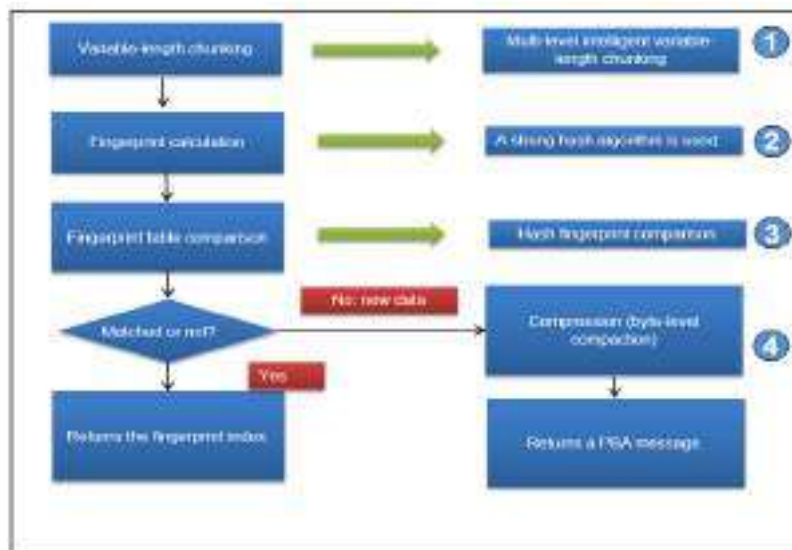
# 8 Data Reduction (SmartDedupe and SmartCompression)

SmartDedupe retains only one data block of the same data blocks, and other data directly references this data block to save storage space.

SmartCompression compresses data to be written to storage media using the lossless compression algorithm and then writes the compressed data blocks to the storage media to save storage space.

OceanProtect applies SmartDedupe for deduplication and then SmartCompression for compression. OceanProtect uses inline variable-length deduplication. In addition to Huawei's proprietary efficient compression algorithm, OceanProtect further compacts data by byte. This chapter describes the working principles of deduplication and compression.

**Figure 8-1** Deduplication and compression processes



1. When a user writes data, variable-length chunking is first performed. The system uses the multi-layer intelligent variable-length chunking algorithm to split the data into variable-length blocks.

2. Then, the system uses the strong hash algorithm to generate fingerprints for these variable-length blocks.
3. The system compares the new fingerprints with those in the fingerprint database. If the same fingerprint is found, the written data block is considered duplicate. In this case, the system updates the data count and returns the existing fingerprint index.
4. If the hash fingerprint is unique, the written data block is considered new data. In this case, the system compresses the data and then writes the data to disks.

### 8.1 Data Preprocessing

### 8.2 Deduplication

### 8.3 Deduplicated Replication

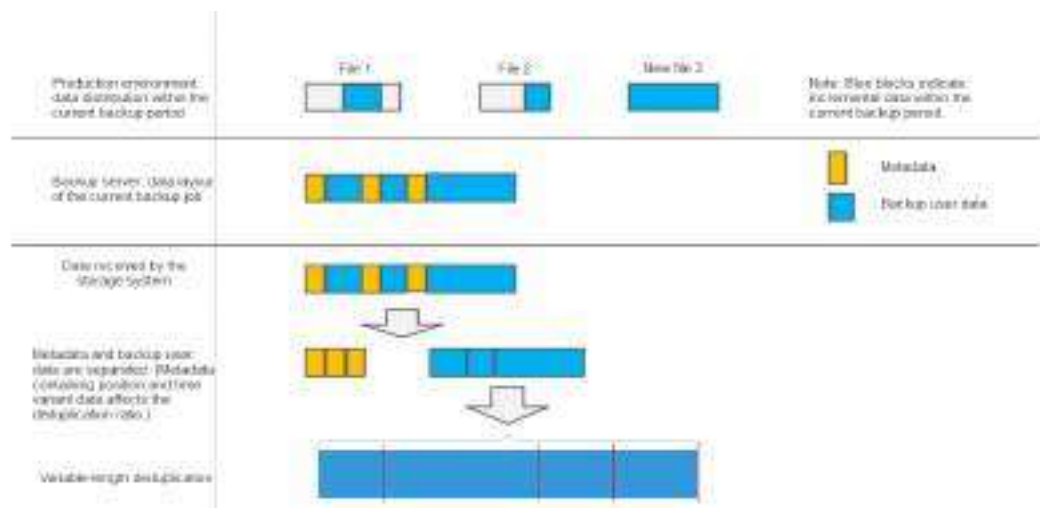
### 8.4 Compression

## 8.1 Data Preprocessing

Generally, data written by backup software includes metadata and backup user data. Some backup software combines metadata and backup data into one file and distributes the metadata to the entire file. The deduplication ratio of metadata is low because metadata contains random data such as timestamp and verification data. If metadata and backup user data are deduplicated together, the overall data deduplication effect will be compromised.

OceanProtect separates metadata and backup data based on feature identification. User data is first aggregated and then variable-length chunking and deduplication are performed. Metadata is directly compressed. In this way, the overall data reduction ratio is improved. See [Figure 8-2](#).

**Figure 8-2** Preprocessing of data written by backup software



## 8.2 Deduplication

## 8.2.1 Source Deduplication

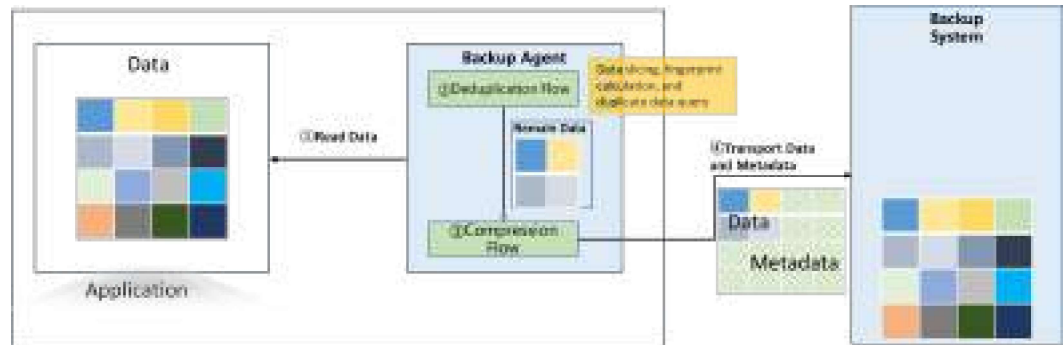
The OceanProtect source deduplication technology can be further divided into SourceDedupe and application-aware SourceDedupe (virtual synthetic full backup technology).

For SourceDedupe, all data (regardless of applications or backup types) is chunked into data blocks. The data block fingerprints are calculated to find the duplicate data (the data that already exists in the backup system). Only non-duplicate data and metadata of the duplicate data is transmitted, reducing the amount of data to be transmitted.

Application-aware SourceDedupe is developed based on backup service characteristics. Backup services include full backup, incremental backup, and differential backup. Similar to SourceDedupe, application-aware SourceDedupe is mainly applicable to full backup scenarios to prevent duplicate data transmission by integrating changed data and existing data on the storage device to synthesize full copies.

### 8.2.1.1 SourceDedupe Client — DataTurbo

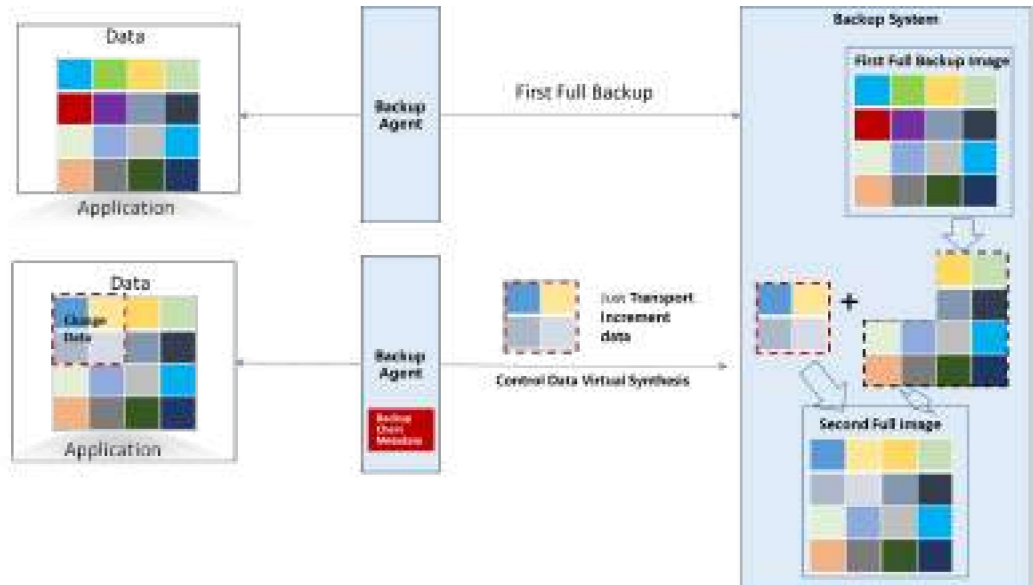
**Figure 8-3** SourceDedupe client — DataTurbo



For SourceDedupe, all data (regardless of applications or backup types) is chunked into data blocks. The data block fingerprints are calculated to find the duplicate data (the data that already exists in the backup system). Only metadata of the duplicate data is transmitted and index counting is added at the back end. SourceDedupe reduces the amount of data to be transmitted and improves backup performance. However, all data is chunked to check for the duplicate data, increasing the resource overhead of the host. To solve this problem, OceanProtect provides an application-aware SourceDedupe technology.

### 8.2.1.2 Application-Aware SourceDedupe

Figure 8-4 Application-aware SourceDedupe



Application-aware SourceDedupe is designed for feature applications. Backup services include full backup, incremental backup, and differential backup. Similar to SourceDedupe, application-aware SourceDedupe is applicable to full backup scenarios to prevent duplicate data transmission. The backup software records the original data generated by full backup or incremental backup of some applications, which is used to associate metadata (such as backup chain metadata) of subsequent backup. In this way, only incremental backup is required for the subsequent full backup because the unchanged data in the existing backup copies can be used to synthesize a new full copy based on the incremental data. Application-aware SourceDedupe achieves less host resource consumption and is widely used in virtualization backup, big data backup, Oracle backup, and host backup.

#### NOTE

For more technical description, see the *OceanProtect Technical White Paper for SourceDedupe*.

### 8.2.2 Variable-Length Deduplication

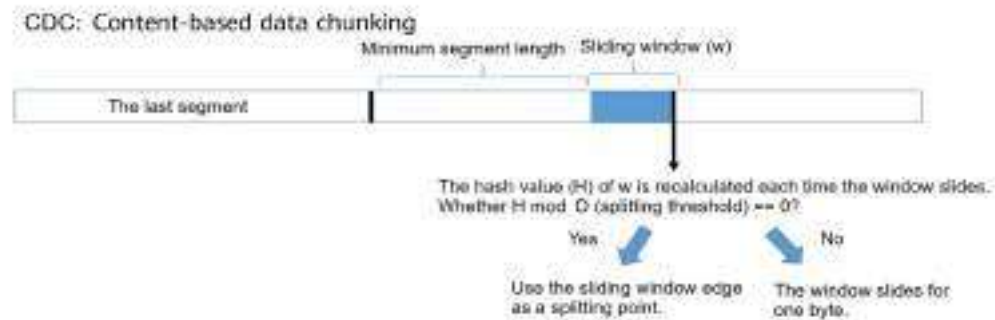
- 

Conventionally, a storage system divides a user file in to fixed-length segments before storing it. In backup scenarios, this will cause serious problems. Once some data is inserted into or deleted from a user file, fixed-length segments that are slightly different from the original ones will be generated. As a result, a large proportion of data before and after the modification cannot be deduplicated. OceanProtect uses variable-length chunking to solve this problem. The system calculates the features of data flows and slices data based on the features, thereby obtaining variable-length data blocks. After some data is inserted or deleted, the data flow features change only in a small range near the insertion or deletion position. In this way, the feature difference between the data before and after the

modification is related only to the amount of inserted or deleted data. For data that does not change, the same variable-length segments are generated and deduplicated by the storage system, delivering a higher deduplication ratio.

On OceanProtect backup storage, data blocks are deduplicated using the content-based chunking algorithm, which is mainly used to deal with low deduplication ratio caused by data offset. The following figure shows the basic principle of content-based chunking. The sliding window content is used to determine whether slicing is performed. In this way, the chunking point is defined based on the content.

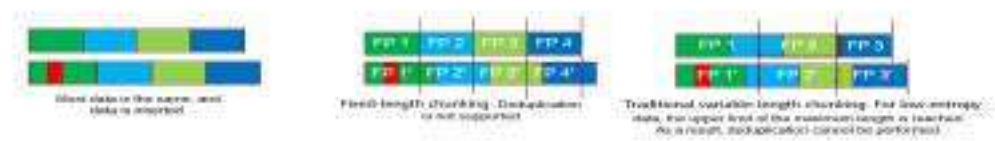
**Figure 8-5** Sliding window calculation



### 8.2.2.1 Intelligent Multi-Layer Variable-Length Chunking

Traditional variable-length chunking does not deliver a satisfactory effect in slicing low-entropy data. It is likely that no proper chunking points can be found even when the sliding window slides to the maximum extent. Consequently, the data block deduplication ratio is low because the maximum segment length is used for chunking, as shown in the following figure [Figure 8-6](#):

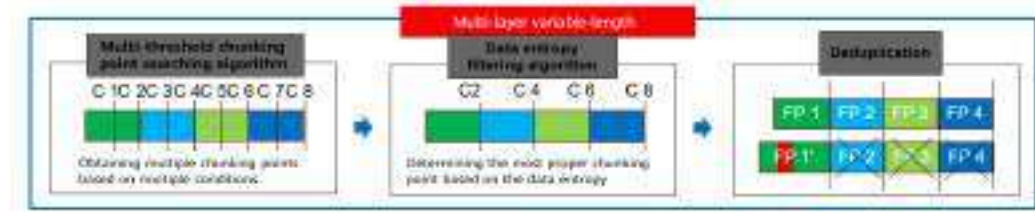
**Figure 8-6** Deduplication effects of traditional fixed-length and variable-length chunking modes in case of data insertion



To overcome the disadvantages of traditional variable-length chunking in slicing low-entropy data, Huawei develops the multi-layer variable-length chunking algorithm for OceanProtect backup storage. As shown in [Figure 8-7](#), a data segment is sliced based on multiple thresholds for obtaining different chunking points. Then, the data entropy filtering algorithm is used for all the chunking points to select proper chunking points that meet the requirements. Variable-length chunking is performed based on the selected chunking points.

Through data simulation, Huawei-developed multi-layer variable-length chunking increases the deduplication ratio by more than 10% compared with the traditional chunking algorithm.

**Figure 8-7 Multi-layer variable-length chunking**



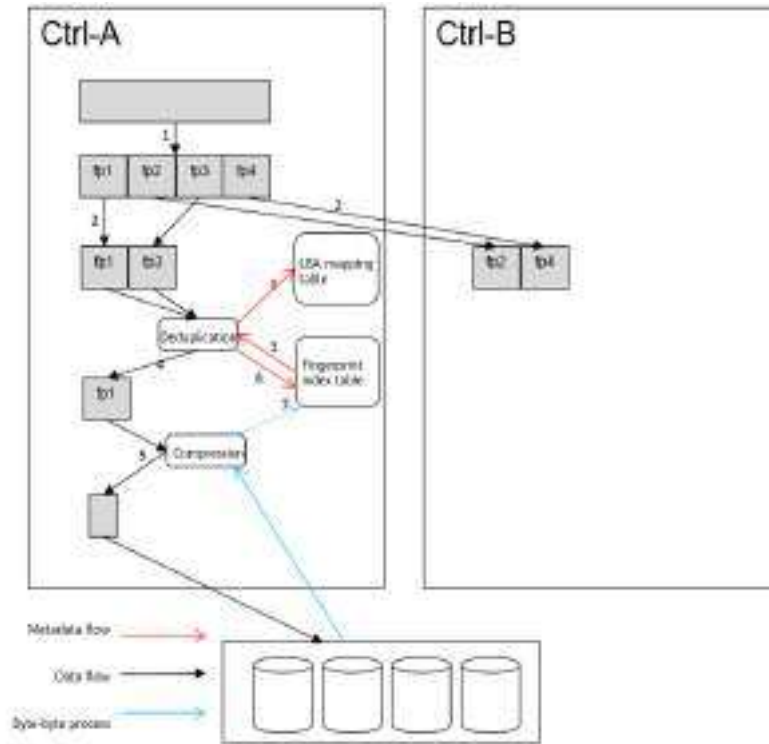
### 8.2.2.2 Deduplication Principles

The deduplication ratio of backup data is high in application scenarios where backup storage systems are used. The traditional mode in which a storage system reads duplicate data from disks, decompresses the data, and then compares the data byte by byte consumes a large amount of CPU computing resources. OceanProtect uses weak-hash fingerprint indexes and strong-hash fingerprint comparison to eliminate data inconsistency caused by a single hash conflict. This saves a large amount of CPU resources of controllers and relieves back-end disk pressure, improving system performance.

The deduplication process on OceanProtect is as follows:

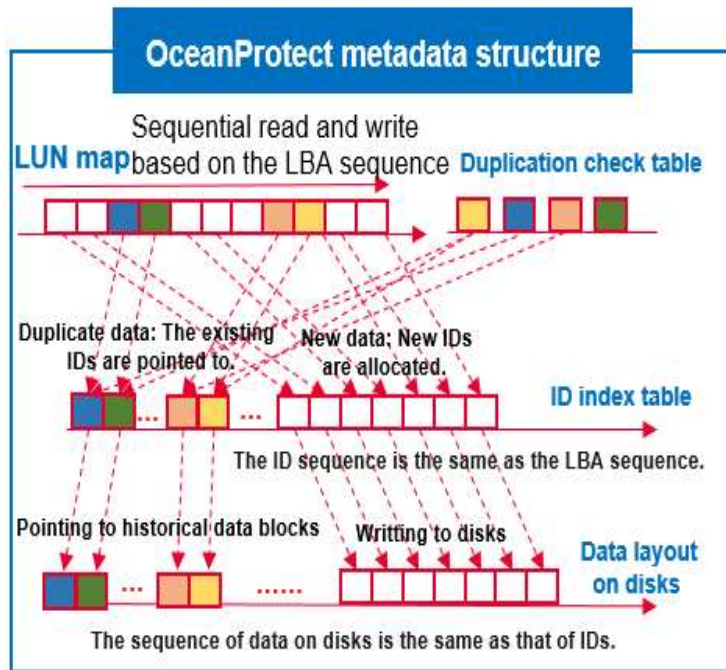
1. After user data is sliced into variable-length segments, the system calculates weak and strong hash fingerprints for each data block.
2. The system sends each data block and these fingerprints to the corresponding controller and compares the weak-hash fingerprint with those in the fingerprint database.
3. If the system finds an identical weak hash fingerprint, it reads the corresponding strong hash fingerprint according to location saved in the fingerprint table and compares it with the strong hash fingerprint of the new data block. If the two fingerprints are the same, the system performs step 7. If the two fingerprints are different, a hash conflict occurs and the system writes the data to disks instead of performing deduplication. If compression is enabled, the data is compressed before being written.
4. If no same fingerprint is found in the fingerprint table, the data to be deduplicated is new data, and new weak hash fingerprint is recorded in the fingerprint index table.
5. If compression is enabled, the system compresses the new data block. If compression is disabled, the system directly applies for storage space for the data block. The storage address will be recorded.
6. The mapping relationship between the weak hash fingerprint and the data storage address is stored in the fingerprint table for future search.
7. If the strong hash fingerprints are the same, the data block is duplicate and will not be saved. The system only increases the reference count of the fingerprint and uses an existing index as that for the data block.

**Figure 8-8** Deduplication process



Inline deduplication on OceanProtect deletes duplicate data before the data is written to storage media. The inline deduplication process is as follows:

The storage system performs variable-length chunking for newly written data, calculates fingerprints for the data blocks, and compares the fingerprints with existing fingerprints in the system. If the same fingerprint is found, the logical address of the file system directly points to the ID of the existing data block without writing any duplicate data block. If no same fingerprint is found, new IDs are allocated to the new data blocks, and mappings between the file system LBA and new IDs, fingerprints and new IDs, as well as the new IDs and write addresses of the storage media are recorded.

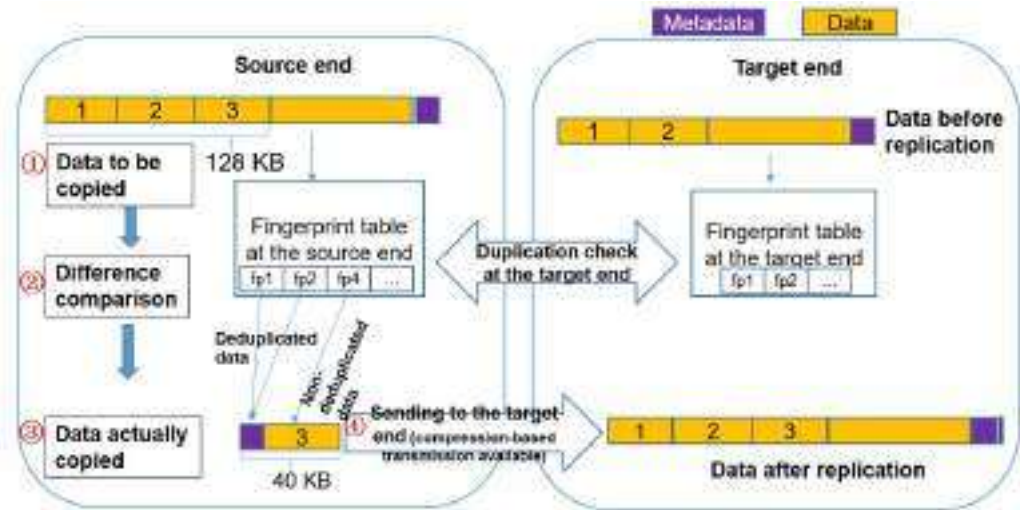


The ID table is used as the conversion center of LBA, fingerprint table, and storage medium address metadata. This design is oriented to the read and write service model of backup software and read and write characteristics of HDDs.

As shown in the preceding figure, after variable-length chunking is performed for file-level sequential write data, fingerprint calculation duplication check is performed. For duplicate data, the LBA directly points to an existing ID. New data blocks are sorted based on the LBA address and then new IDs are allocated. The new IDs are in ascending order and never reused. New data is compressed and then written to disks. New data is arranged on storage media based on the sequence of the new data IDs. According to the end-to-end result, the new data blocks that store the deduplicated sequential data written by the backup software are also saved in sequence on the storage media.

## 8.3 Deduplicated Replication

In backup service scenarios, local data needs to be fully backed up to the backup system and then securely stored remotely. This backup mode features high data duplication, which highlights the importance of data deduplication. Highly redundant backup data needs to be backed up within a shorter period of time to accelerate for efficiency and protection. Therefore, requirements on link replication are as follows: Deduplication must be provided to greatly save storage space, and lower bandwidth requirements are essential to accelerate backup and reduce WAN usage costs. OceanProtect implements deduplicated replication to greatly reduce the amount of data to be copied and save replication bandwidth resources.



Deduplicated replication process:

1. Before asynchronous replication, new backup data after the previous replication time point is confirmed at the source end to determine the data to be copied.
2. Fingerprints of the data to be copied are searched for and sent to the target end for duplication check, determining whether deduplication can be performed on the target end.
3. For data that can be deduplicated, only corresponding fingerprints are sent. For data that cannot be deduplicated, all data is copied. The two types of data form the data that is finally copied through the replication link.
4. Compression can be performed during transmission to further reduce the amount of data to be transmitted.

## 8.4 Compression

Data compression involves two processes. Input data blocks are first compressed to smaller ones using a compression algorithm and then compacted before being written to disks.

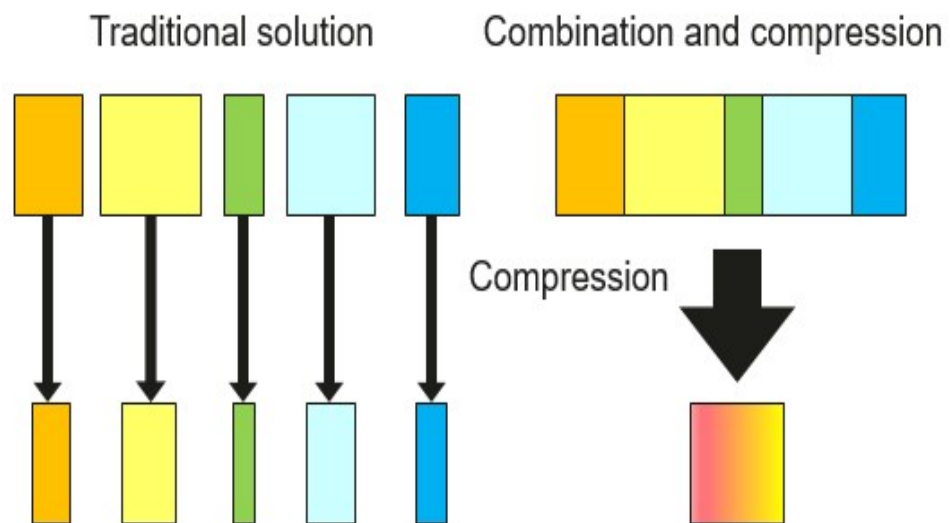
### 8.4.1 Compression After Combination

According to the test data, the larger the data block involved in compression at a time, the higher the compression ratio.



OceanProtect performs compression after combination. As shown in [Figure 8-9](#), the system combines data blocks with consecutive IDs into a large data block (128 KB) for compression. As backup software involves sequential writes and recovery involves sequential reads, data blocks with consecutive IDs are combined before being compressed. During backup recovery, data blocks with consecutive IDs obtained after one read are hit in the read cache by subsequent read requests initiated by backup software, fastening disk read and improving the overall performance.

**Figure 8-9** Compression after combination

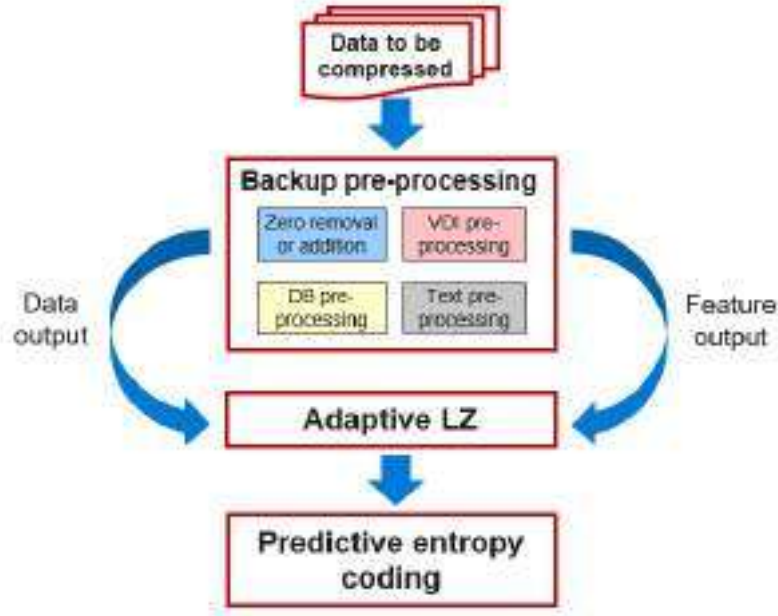


## 8.4.2 Compression (SmartCompression) Process

### Preprocessing in Backup Scenarios

Before data compression, OceanProtect uses multiple Huawei-developed preprocessing algorithms for feature-based processing in main backup scenarios. The preprocessing is to identify data scenarios based on data flow features and then clean (removing identifiers such as tags inserted by software), rearrange (reversible data rearrangement based on data type features for improved

compression ratio), and deduplicate (deleting redundant data based on data features) data to improve the compression ratio. The pre-processed data and identified data features are transmitted to the adaptive LZ module for deduplication and compression.



The pre-processed data and identified data features are output.

## Adaptive LZ

Based on the data features identified by the preprocessing module, multiple preset deduplication policies are matched to deduplicate the pre-processed data.

## Predictive Entropy Encoding

The Huawei-developed high-performance predictive encoding algorithm trains predictive models based on backup data features and works with word frequency information and prediction technologies to improve the compression ratio.

With the preprocessing, adaptive LZ, and predictive entropy encoding technologies, OceanProtect backup storage improves the compression ratio by more than 10% in the VM + file sharing backup scenario and by more than 30% in the database backup scenario.

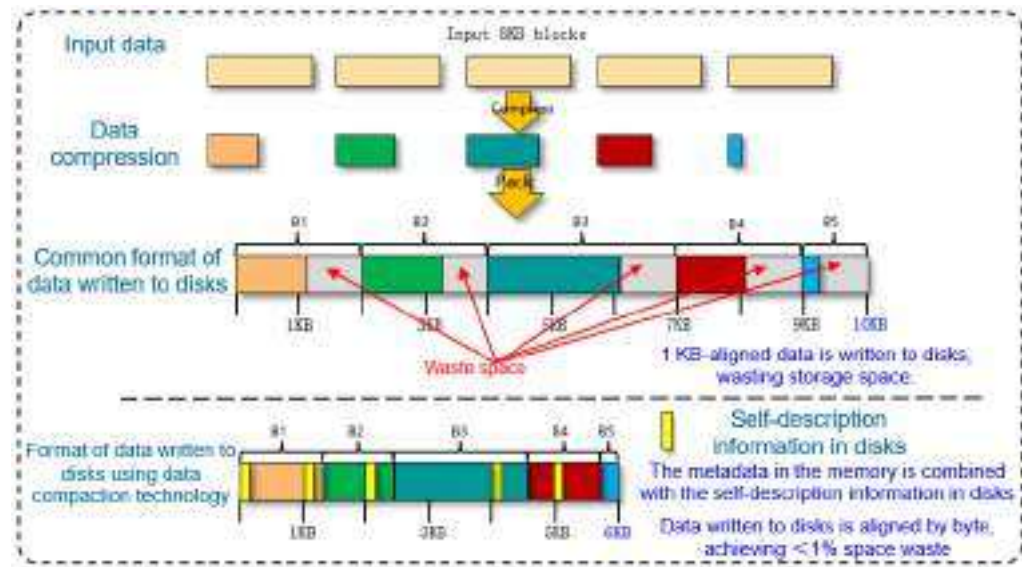
### 8.4.3 Data Compaction

This section describes how data compaction is implemented.

Compressed user data is aligned by byte and then compacted to reduce the waste of physical space. This provides a higher reduction ratio than the 1 KB-based alignment generally used in the industry.

As shown in [Figure 8-10](#), byte-level alignment saves 2 KB physical space compared with 1 KB-based alignment, improving the compression ratio.

**Figure 8-10** Alignment of compressed data by byte



## 8.4.4 Data Rearrangement

OceanProtect uses the data rearrangement technology. For each data block, the system inserts data verification information after every 512 bytes of data. Because random data exists in the verification information, the calculation of the compression algorithm may be disturbed by the random data, resulting in a low compression ratio. Therefore, the data rearrangement technology separates the user data from parity data, compresses the user data, and reduces the parity data in a customized way. In this way, a better compression ratio and higher compression and decompression speeds can be obtained.

**Figure 8-11** Data rearrangement



# 9 System Security Design

Storage security must be safeguarded by technical measures. Data integrity, confidentiality, and availability must be monitored. Secure boot and access permission control as well as security policies based on specific security threats to storage devices and networks further enhance system security. All these measures prevent unauthorized access to storage resources and data. Storage security includes device, network, service, and management security. This chapter describes software integrity protection, secure boot, and data encryption capabilities related to system security. The digital signature technology ensures that the product package (including the upgrade package) developed by Huawei is not tampered with during device installation and upgrade. The secure boot technology guarantees that startup components are verified during startup of storage devices to prevent startup files from being tampered with. The disk encryption feature is used to protect data stored on disks and prevent data loss caused by disk loss.

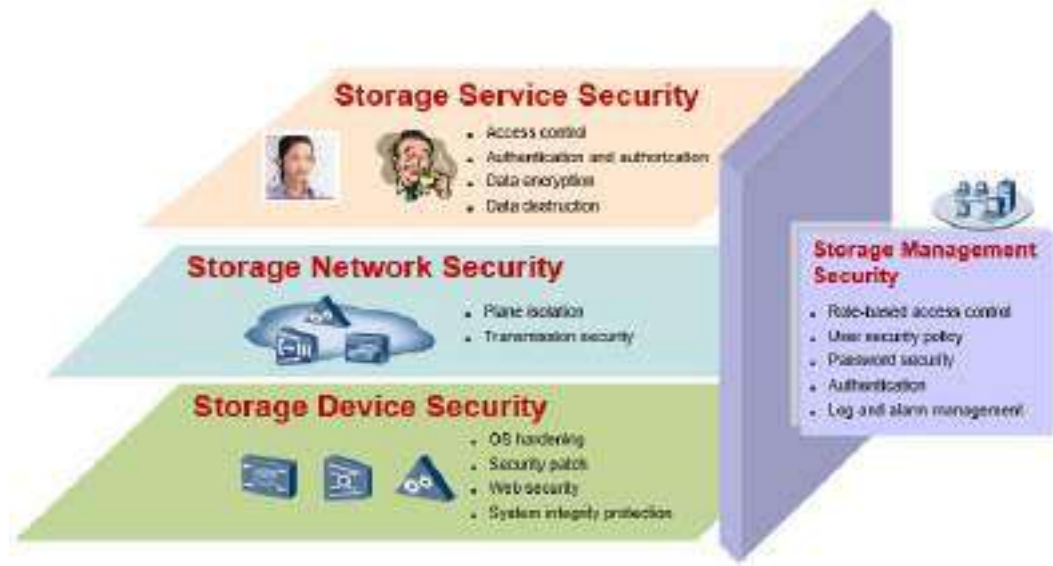
## 9.1 Overall Security Architecture

## 9.2 Security Capabilities

## 9.1 Overall Security Architecture

Based on the ITU-T X.805 telecom network security architecture model, the overall security architecture employs security measures to cope with security threats to storage devices and networks and meet the security requirements of enterprise operations, as shown in [Figure 9-1](#).

Figure 9-1 Security architecture



## 9.2 Security Capabilities

The security layers are described as follows:

- **Storage device security:** Covers OS hardening, patch management, web security, system integrity protection, software integrity protection, and secure boot.
- **Storage network security:** Covers plane isolation, network isolation, and transmission security.
- **Data storage and backup service security:** Covers access control, authentication and authorization, data encryption, data destruction, data retention, and data audit.
- **Management security:** Covers role-based access control (RBAC), user security policy, password security, authorization and authentication (LDAP, local user, and two-factor authentication), certificate management, as well as log and alarm management.

### NOTICE

For more information about security design, see the *Huawei OceanProtect Security Technical White Paper*.

# 10 System Serviceability Design

This chapter describes how to manage the OceanProtect storage through various interfaces (including DeviceManager, CLI, RESTful API, SNMP, and SMIS) and describes the upgrade mode transparent to hosts.

[10.1 Storage System Management](#)

[10.2 Backup System Management](#)

[10.3 Intelligent Cloud Management](#)

[10.4 OceanProtect Appliance Upgrade](#)

## 10.1 Storage System Management

OceanProtect provides device management interfaces and integrated northbound management interfaces. Device management interfaces include a graphic management interface DeviceManager and a command-line interface (CLI). Northbound interfaces are RESTful interfaces, supporting SNMP, evaluation tools, and third-party network management plug-ins. For details, see <https://info.support.huawei.com/storage/comp/#/home>.

### 10.1.1 DeviceManager

DeviceManager is a built-in HTML5-based management system for OceanProtect. It provides wizard-based GUI for efficient management. Users can enter `https://Storage management IP address:8088/` on the browser to use DeviceManager. On DeviceManager, you can perform almost all required configuration and management operations.

You can use the following functions on DeviceManager:

- Storage space management: includes management of the storage pool, LUNs, and mappings between LUNs and hosts.
- Data protection management: uses snapshots and replication to protect data.
- Configuration task: provides background tasks for complex configuration operations to trace the procedure of the configuration process.

- **Fault management:** monitors the status of storage devices and management units on storage devices. If faults occur, alarms will be generated and troubleshooting suggestions and guidance will be provided.
- **Performance and capacity management:** supports real-time performance and capacity monitoring, historical performance and capacity data collection and query, and performance data association analysis.
- **Security management:** supports the management of users, roles, permissions, certificates, and keys.

DeviceManager uses a new UI design to provide a simple interactive interface. Users can complete configuration tasks only in a few operations, improving user experience.

### 10.1.1.1 Storage Space Management

#### 10.1.1.1.1 Flexible Storage Pool Management

OceanProtect manages storage space by using storage pools. A storage pool consists of multiple SSDs and can be divided into multiple LUNs or file system for hosts to use. You can use one storage pool to manage all of the space or create multiple storage pools.

- The entire storage system uses only one storage pool.  
This is the simplest method. You only need to create one storage pool with all disks during system initialization.
- Multiple storage pools are divided to isolate different applications.  
If you want to use multiple storage pools to manage space and isolate fault domains of different applications, you can manually create storage pools at any time in either of the following ways:
  - Specify number of disks used to create a storage pool. The system automatically selects the disks that meet the requirements.
  - Manually select specific disks to create a storage pool.

#### 10.1.1.2 Configuration Task

DeviceManager automatically creates a background configuration task after a user submits a complex configuration operation. The task is running on the storage background. In this way, the user can perform other operations while the task is still being executed.

For example, the background configuration task mechanism applies to protection group-based protection, and cross-device protection. Suppose that a user needs to configure protection for a protection group that has hundreds of LUNs. It takes a long time and hundreds of operations to complete the configuration. However, the background configuration task mechanism helps the user complete the configuration on the background, freeing the user from long-term waiting.

- **Task steps and progress**  
A complex task usually contains multiple executable steps. You can check which task step is being executed and view the overall task execution progress (%) on DeviceManager.

- Tasks can be executed only in the background  
After a configuration task is submitted, the task is automatically executed in the background. You can close the DeviceManager page without waiting for the task to complete. You can also submit multiple configuration tasks. If the resources on which the tasks depend do not conflict, the tasks are automatically executed in sequence in the background.
- Resuming tasks from the breakpoint  
If the system is powered off unexpectedly during the task execution, the task will be executed at the breakpoint and all the steps will be performed after the system is restarted.
- Retrying failed tasks manually  
If an exception occurs during the execution of a step, the task is automatically interrupted and the cause is displayed. For example, a task cannot be executed because the storage space is insufficient during files system creation. If you manually rectify the fault that causes task interruption, you can manually restart the task after the fault is rectified and the task continues from the break point.

### 10.1.1.3 Fault Management

#### 10.1.1.3.1 Monitoring Status of Hardware Devices

This function provides hardware views in a what-you-see-is-what-you-get manner and uses colored digits to make status of different hardware more distinguishable. **Figure 10-1** shows the status statistics of hardware components.

**Figure 10-1** Inventory management



You can further navigate through a specific device frame and its hardware components, and view the device frame in the device hardware view. In the hardware view, you can monitor the real-time status of each hardware component and learn the physical locations of specific hardware components (such as ports and disks), facilitating hardware maintenance.

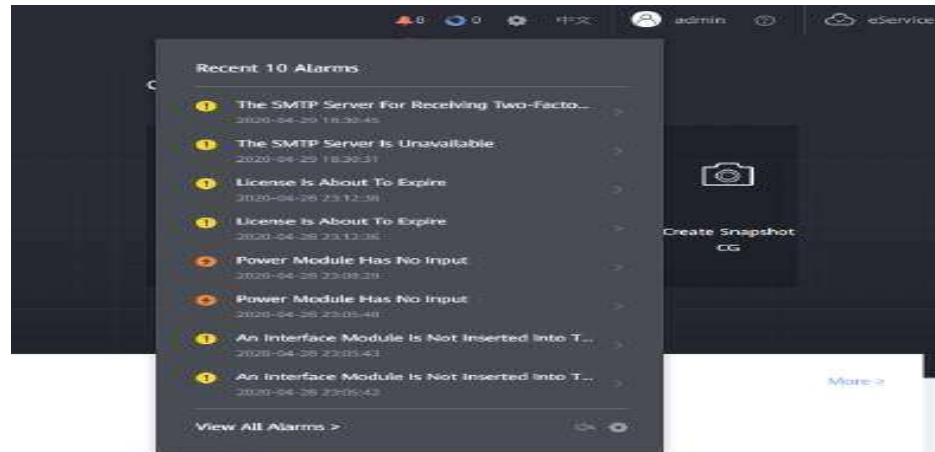
You can query the real-time health status of the disks, ports, interface modules, fans, power modules, BBUs, controllers, and disk enclosures. In addition, disks and

ports support performance monitoring. You can refer to [10.1.1.4 Performance and Capacity Management](#) for more details.

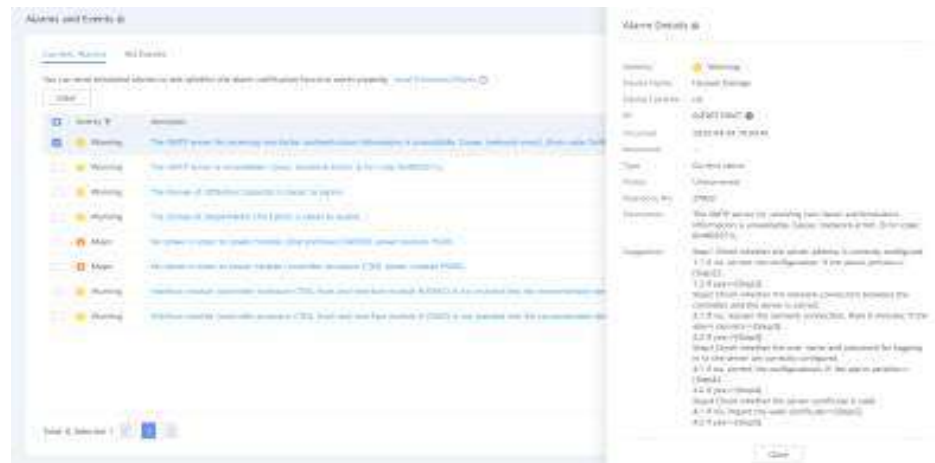
### 10.1.1.3.2 Alarm and Event Monitoring

This function provides you with real-time fault monitoring information. If a system fault occurs, the fault is pushed to the home page of DeviceManager in real time. [Figure 10-2](#) shows an example.

**Figure 10-2** Alarm notification



A dedicated page is available for you to view information about all alarms and events and also provides you with troubleshooting suggestions.



You can also receive alarm and event notifications through syslog, email, and SMS (a dedicated SMS modem is required). You can configure multiple email addresses or mobile phone numbers to receive notifications.

### 10.1.1.4 Performance and Capacity Management

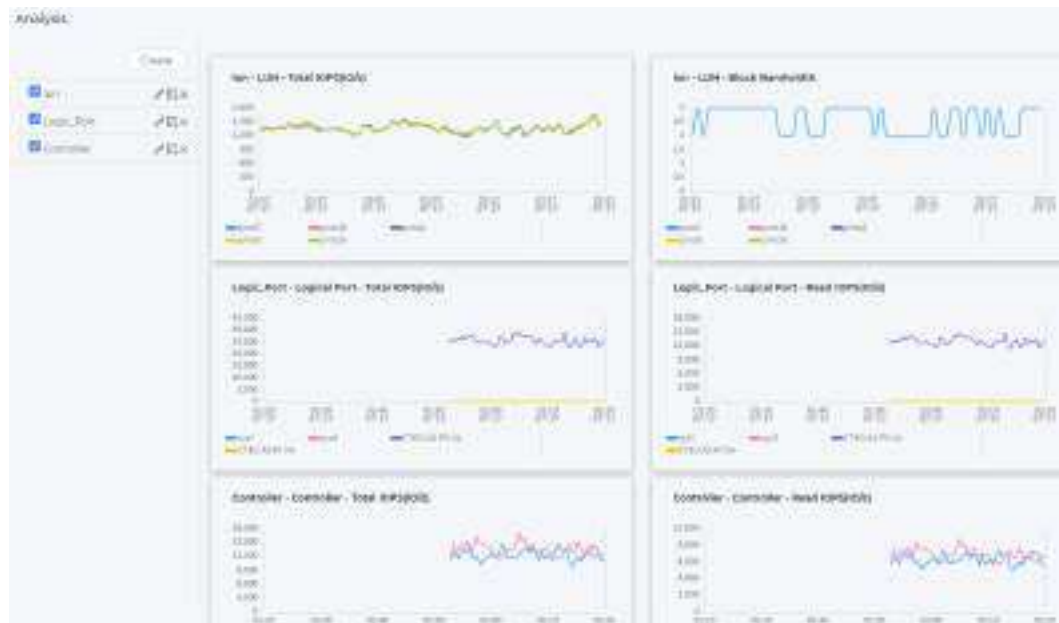
Performance data collection and analysis are essential to daily device maintenance. Because the performance data volume is large and analyzing the data consumes many system resources, an extra server is often required for installing dedicated performance data collection and analysis software, making performance management complex.

OceanProtect has a built-in performance and capacity data collection and analysis component that is ready for use. The component is specially designed to consume minimal system resources.

#### 10.1.1.4.1 Built-In Performance Data Collection and Analysis Capabilities

OceanProtect has built-in performance collection and analysis software. You do not need to install the software separately. You can collect, store, and query historical performance and capacity data of a maximum of the past three years. Alternatively, you can specify the period as required. Performance data of multiple specific objects can be collected and analyzed, such as controllers, ports, disks, file systems, remote replication, and replication links. Multiple performance indicators can be monitored for each object, such as bandwidth, IOPS, average I/O response time, and usage.

Different objects and performance indicators can be displayed in the same view to help you analyze performance issues. You can specify the desired performance indicators to analyze the top or bottom objects, so that you can locate overloaded objects more efficiently and tune performance more precisely. The following figure shows performance monitoring. For details about monitoring objects, monitoring indicators, and performance analysis functions, see the online help.



#### 10.1.1.4.2 Independent Data Storage Space

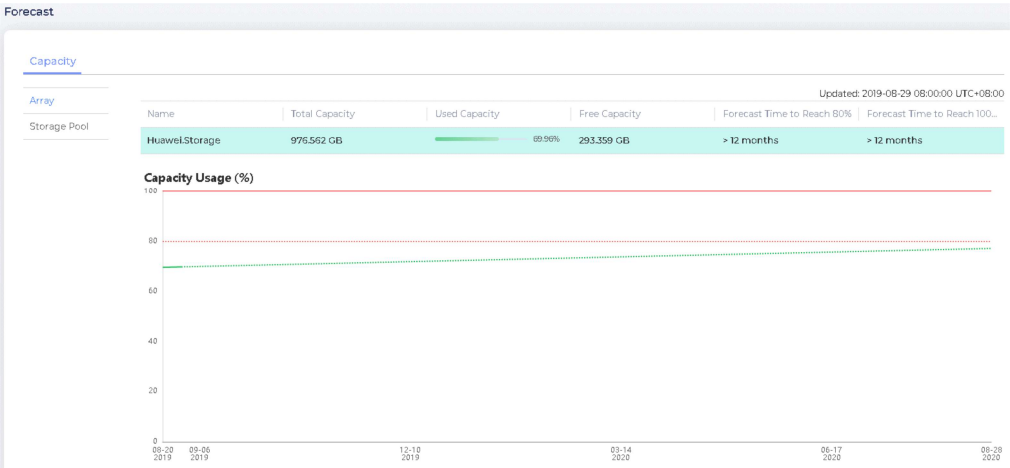
To store collected performance and capacity data, a dedicated storage space is required. DeviceManager provides a dedicated configuration page for users to select a storage pool for storing data.



You can also customize the data retention period. The maximum retention period is three years. DeviceManager automatically calculates the required storage space based on your selection and allocates the required storage space in the storage pool.

10.1.1.4.3 Capacity Prediction

DeviceManager supports the prediction of system and storage pool capacity usage in the next year. Users can formulate a better resource usage plan in advance and schedule resources based on the prediction results.



10.1.1.4.4 Performance Threshold Alarm

This function allows you to configure threshold alarms for objects such as controllers, ports, file systems, and replication. The alarm threshold, flapping period, and alarm severity can be customized.

Different storage resources carry different types of services. Therefore, common threshold alarms may not meet requirements. Performance management allows you to set threshold rules for specified objects to ensure high accuracy for threshold alarms.



10.1.1.4.5 Scheduled Report

Performance and capacity reports for specific objects can be generated periodically. Users can learn about the performance and capacity usage of storage devices periodically.



Reports can be generated by day, week, or month. You can set the time when the reports are generated, the time when the reports take effect, and the retention duration of the reports. You can select a report file format. Currently, the \*.pdf and \*.csv formats are supported.

You can select the objects for which you want to generate a performance report. All the objects for which you want to collect performance statistics can be included in the report. You can also select the performance indicators to be displayed in the report. The capacity report collects statistics on the capacity of the entire system and storage pools.

You can create multiple report tasks. Each report task can be configured with its own parameters. The system automatically generates reports according to the task requirements.

## 10.1.2 CLI

The Command Line Interface (CLI) allows administrators and other system users to manage and maintain the storage system. It is based on the secure shell protocol (SSH) and supports key-based SSH access.

## 10.1.3 RESTful APIs

RESTful APIs of OceanProtect allow system automation, development, query, and allocation based on HTTPS interfaces. With RESTful APIs, you can use third-party applications to control and manage arrays and develop flexible management solutions for OceanProtect.

## 10.1.4 SNMP

The storage system reports alarms and events through SNMP traps and allows you to use SNMP interfaces to query storage alarms and events, collect data performance, and query information of objects managed on the storage system.

## 10.1.5 SMI-S

The Storage Management Initiative Specification (SMI-S) is a storage standard management interface developed and maintained by the Storage Networking Industry Association (SNIA). Many storage vendors participate in defining and implementing SMI-S. The SMI-S interface is used to configure storage hardware and services. Storage management software can use this interface to manage storage devices and perform standard management tasks, such as viewing storage hardware, storage resources, and alarm information. OceanProtect supports SMI-S 1.6.1, 1.6.0, and 1.5.0.

## 10.1.6 Tools

OceanProtect provides diversified tools for pre-sales assessment (eDesigner) and post-sales delivery (SmartKit). These tools effectively and quickly help deploy, monitor, analyze, and maintain OceanProtect.

## 10.2 Backup System Management

Huawei OceanProtect data backup feature provides an independent backup service management interface and integrated northbound management interfaces. Northbound interfaces are mainly RESTful APIs, supporting SNMP alarms and event reporting.

The following roles are supported:

User Management		
User Role	System administrator	Has all system permissions.
	Data protection administrator	Has the data protection permission. Data protection administrators of multiple roles can be created to meet domain-based management requirements.
	Remote device administrator	Machine-machine account for remote replication, which has the remote replication permission
	Auditor	Has the permission to audit system events and operation logs.

### 10.2.1 OceanProtect GUI

There is a built-in HTML5 management system in OceanProtect data protection system. It provides wizard-based GUI for efficient management. Users can enter <https://management IP address:25080/> on a browser to use. On the management system, you can configure and manage all backup services. The following figure shows the OceanProtect login page.

**Figure 10-3** OceanProtect login page



You can use the following functions on OceanProtect:

- Management protected resources: includes resource list, status management, and resource scanning.
- Data protection management: includes SLA management, management of resource and SLA binding, and data restoration.
- Data repurposing: includes live mount, data anonymization, and global search.
- System monitoring: includes traffic monitoring, job monitoring, and alarm monitoring of nodes.
- System management: includes configuration management, including network, storage, log, and application settings.
- Security management: includes user management, role management, security policy management, certificate management, and user data isolation.

## 10.2.2 RESTful APIs

RESTful APIs allow system automation, development, query, and allocation based on HTTPS interfaces. With the API, you can use third-party applications to control and manage the data protection services. RESTful APIs enable users to develop flexible management solutions for Huawei OceanProtect data protection system.

## 10.2.3 Centralized Management of Multiple Devices

With various applications in customers' production environments and large-scale services, a single backup device cannot meet the backup requirements of the entire data center, which makes O&M more difficult. OceanProtect enables O&M of multiple devices on one platform through a unified management console.

**Figure 10-4** Unified management and monitoring page for multiple devices



Unified management of multiple devices provides the following capabilities:

- **Login-free redirection for devices in a cluster**  
You can switch to any device in the cluster through the unified cluster management page without login.
- **Unified job status monitoring**  
You can view the job execution status of all clusters on the global dashboard, collect statistics on the job status of all clusters, and select a specified cluster on the same page to view the job monitoring list.



- **Unified resource status monitoring**  
You can view the resource protection status of all clusters on the global dashboard, collect statistics on the resource protection of all clusters, and select a specified cluster on the same page to view the resource protection status.
- **Unified alarm status monitoring**  
You can view the alarm statistics of all clusters on the global dashboard and select a specified cluster on the same page to view the alarm information.
- **Unified system capacity statistics and monitoring**  
You can view the total capacity of all devices on the global dashboard to learn the overall capacity usage of multiple clusters. You can also select a cluster on the same page to view the capacity information.
- **Unified monitoring of data reduction ratios**  
You can collect statistics on data reduction ratios of multiple clusters and monitor them in real time in the global dashboard. During deduplication, a single device is used as the deduplication domain.

- **Device status and performance monitoring**

You can monitor the online and offline status of devices on the global dashboard and select a specified cluster to view performance monitoring data.

- **Unified global search**

You can carry out unified global search of resources. After the search is triggered, requests are distributed to all clusters and information about resources and files that meet the search criteria is returned.

- **Redirection-free function operations**

You can perform all service configuration and O&M functions of the backup function on a specified cluster on the same page without logging in to the specified cluster.

#### NOTICE

For details, see the *OceanProtect Administrator Guide*.

## 10.2.4 Client push installation in batches

The out-of-the-box data function server software is installed in the device before delivery. The backup client software installation package is also provided in the device. During deployment on the live network, you can install the backup client software on ProtectManager in push installation mode. During installation, you do not need to log in to each protected resource or external proxy node.

**Figure 10-5** Client push installation in batches



After the push job starts, an installation job is generated for each client. The installation failure of a single client does not affect the installation of other clients. A client installation job that fails is automatically rolled back and can be pushed again.

- **Supports IPv4/IPv6 and Windows/Linux.**

Supports push installation of **IPv4/IPv6**, Windows, and Linux clients.

- **Batch push**

Batch push is supported. You can enter multiple network segments and IP addresses. An independent push job is generated for a single client. A push job that fails is automatically rolled back without affecting other push jobs.

## 10.3 Intelligent Cloud Management

In traditional service support mode, technical support personnel provide services manually. Faults may not be detected in a timely manner and information may not be delivered correctly. To resolve the preceding problems, Huawei provides the eService cloud intelligent management system (eService for short) based on the cloud native. With the customer's authorization, device alarms and logs are sent to eService at a scheduled time every day. Based on artificial intelligence (AI) technologies, eService implements intelligent fault reporting, real-time health analysis, and intelligent fault prevention to identify potential risks, automatically locate faults, and provide troubleshooting solutions, minimizing device running risks and reducing operation costs.

**Figure 10-6** Typical eService networking



eService enables the client to work with the cloud system.

- eService Client: deployed on the customer side  
eService Client is used or eService is enabled on DeviceManager to connect to the eService cloud system. Alarm information about customer devices is collected and sent to the Huawei cloud system in a timely manner.
- eService cloud system: deployed in Huawei technical support center  
eService receives device alarms from customer clients all day long, automatically reports problems to Huawei technical support center, and creates service requests (SRs). Huawei service engineers will assist customers in resolving the problems in time.

eService has the following advantages:

- eService provides a self-service O&M system for customers, aiming for precise customized information services.
- Customers can use a web to access eService to view device information anytime anywhere.
- High data security and reliability are ensured. Secure information transmission is provided and eService can access customers' systems only after being authorized by the customers.
- eService provides 24/7 secure, reliable, and proactive O&M services. SRs can be automatically created.
- Based on Huawei Cloud, the eService cloud system drives IT O&M activities through big data analytics and artificial intelligence (AI) technologies to

identify faults in advance, reduce O&M difficulties, and improve O&M efficiency.

### 10.3.1 Scope of Information to Be Collected

With the authorization of customers, Huawei storage systems can be connected to eService through a network to periodically collect their O&M data, helping fully understand storage O&M activities. The O&M data includes performance data, configuration information, alarm information, system logs, and disk information.

**Table 10-1** Scope of information to be collected

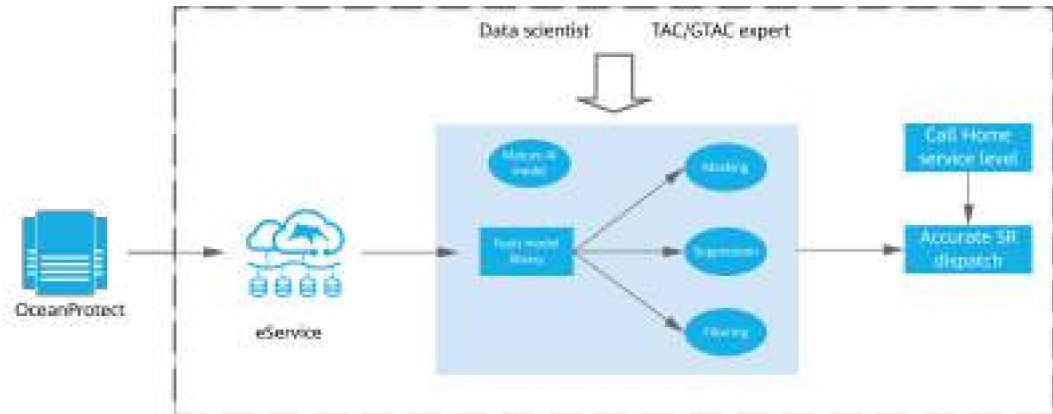
Data Type	Data Description	Interval of Data Upload
Performance data	A <b>.txt</b> file in the JSON format	Uploaded automatically. The new performance data is uploaded to the eService cloud system every 5 minutes.
Configuration information	A <b>.txt</b> file	Uploaded automatically. The configuration is uploaded to the eService cloud system once a day.
Alarm information	HTTPS message	Uploaded automatically. The new device alarm messages are uploaded to the eService cloud system every 30 seconds.
System logs	Email	Uploaded automatically. The new alarm messages are uploaded to the eService cloud system every 5 minutes.
	A <b>.tgz</b> file	Manually uploaded by Huawei technical support personnel on the eService cloud system. In the current version, all system logs and system logs in the latest one hour, latest two hours, latest 24 hours, or a specific time period can be uploaded.
Disk information	A <b>.txt</b> file	Uploaded automatically. The disk information is uploaded to the eService cloud system once a day.

### 10.3.2 Intelligent Fault Reporting

eService provides 24/7 health reporting. If a device fails, eService is automatically notified. Traditional fault reporting mechanisms require the configuration of rules to mask, suppress, and filter alarms, therefore, traditional mechanisms have difficulties in covering all scenarios and have problems such as false alarm reporting and alarm missing. eService provides 24/7 active monitoring for customer device alarms. The alarms generated by the devices are reported to eService. Based on the fault feature model library of global devices, eService performs automatic alarm masking, suppression, and filtering, improving the

accuracy and efficiency of alarm handling. According to the Call Home service level, eService can automatically create an SR and send it to the corresponding Huawei engineer for problem processing. In addition, eService notifies customers of the problems according to the agreed contact details (email by default) to accelerate fault locating and resolution.

**Figure 10-7** Implementation of intelligent fault reporting



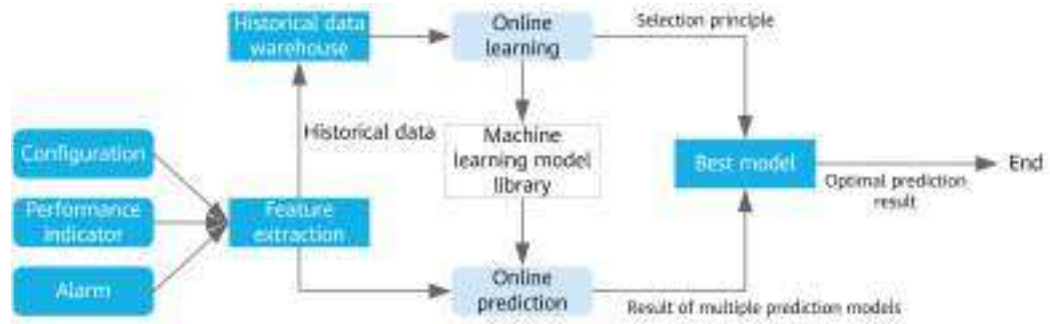
### 10.3.3 Capacity Prediction

The system capacity changes are affected by multiple factors. The traditional single prediction algorithm cannot ensure the accuracy of prediction results. eService ensures reasonable and accuracy of prediction results from the following aspects:

- eService uses multiple prediction model clusters for online prediction, outputs the prediction results of multiple models, and then selects the best prediction results based on the selection rules recommended by online prediction. At the same time, eService trains and verifies itself based on the historical capacity data periodically, identifies the trend, period, and partial changes of capacity based on linear prediction, and optimizes the model parameters, ensuring that the optimal prediction algorithm is selected.
- The eService prediction algorithm model can accurately identify various factors that affect capacity changes, for example, sudden capacity increase and decrease caused by major events, irregular trend caused by capacity reclamation of existing services, and capacity hops caused by new service rollout. In this way, the system capacity consumption can be predicted more accurately.

eService selects the best prediction model using the AI algorithm, and predicts the capacity consumption in the next 12 months. Based on the capacity prediction algorithm, eService provides the overloaded resource warning, capacity expansion suggestions for existing services, and annual capacity planning functions for customers.

**Figure 10-8** Working principles of capacity prediction



Responsibilities of each component:

- Data source collection: collects configurations, performance indicators, and alarm information to reduce the interference of multiple factors on machine learning and training results.
- Feature extraction: uses algorithms to transform and extract features automatically.
- Historical database warehouse: stores historical capacity data of the latest year.
- Online training
  - Uses a large number of samples for training to obtain the measurement indicator statistics predicted by each model and output the model selection rules.
  - For the current historical data, performs iteration for a limited number of times to optimize the model algorithm.
- Machine learning model library: includes ARIMA, fbprophet, and linear prediction models.
- Online prediction: performs prediction online using the optimized models and outputs the prediction results of multiple models and the mean absolute percentage error (MAPE) values of measurement indicators.

$$M = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

In the preceding formula,  $A_t$  indicates the actual capacity and  $F_t$  indicates the predicted capacity.

- Best model selection: weights the model statistics of online training and results of online prediction, and selects the optimal prediction results.

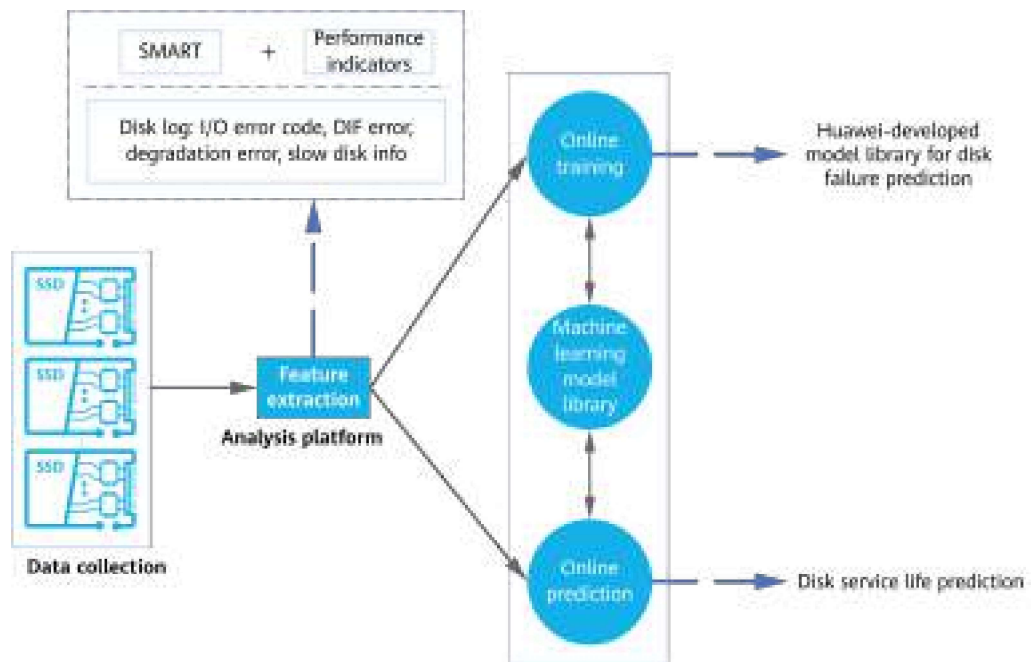
### 10.3.4 Disk Health Prediction

Disks are the basis of a storage system. Although various redundancy technologies are used in storage systems, they only tolerate the failure of a limited number of disks. For example, RAID 5 tolerates the failure of only one disk. If two disks fail, the storage system stops providing services to ensure data reliability. Disks are the largest consumables in a storage system. Therefore, disk service life is the most concerned topic for many users. SSDs are electronic components and their service

life prediction indicators are limited. In addition, the number of read and write requests varies every day, which further complicates disk service life prediction.

eService collects the Self-Monitoring, Analysis and Reporting Technology (S.M.A.R.T.) information, I/O link information, as well as reliability indicators of HSSDs, and enters such information to hundreds of HSSD failure prediction models, implementing accurate SSD service life prediction. eService uses intelligent AI algorithms to predict disk risks to detect failed disks and replace risky disks in advance, preventing faults and improving system reliability.

**Figure 10-9** Working principles of disk health prediction



- Data source collection

Disk vendors provide the S.M.A.R.T. data of disks. The S.M.A.R.T. data can indicate the running status of the disks and help predict risky disks in a certain range. However, it is difficult to ensure the accuracy of prediction results. eService uses AI to dynamically analyze S.M.A.R.T. changes of disks, performance indicator fluctuations, and disk logs, ensuring more accurate prediction results.

- S.M.A.R.T

For SSDs, the interfaces provide SCSI log page information that records the current disk status and performance indicators, such as the grown defect list, non-medium error, and read/write/verify uncorrected errors.

- Performance indicators

Workload information such as the average I/O size distribution per minute, IOPS, bandwidth, and number of bytes processed per day, as well as performance indicators such as the latency and average service time are included.

- Disk logs

I/O error codes collected by Huawei storage systems, DIF errors, degradation errors, slow disk information, slow disk cycles, and disk service life information are included.

- Feature extraction

Based on massive historical big data, feature transformation and feature extraction are automatically performed using algorithms.

- Analysis platform

- Online training: Trainings is performed based on model algorithms, and these model algorithms are optimized through iteration of a limited number of times.
- Machine learning model library: Huawei-developed disk failure prediction models are contained.
- Online prediction: Optimized training models are used to predict disk failures.

- Prediction results

eService tests and verifies massive SSD data and can accurately predict the SSD service life based on Huawei-developed disk failure prediction models.

### 10.3.5 Device Health Evaluation

Generally, after a device goes online, users perform inspection to prevent device risks, which has two disadvantages:

- Inspection frequency

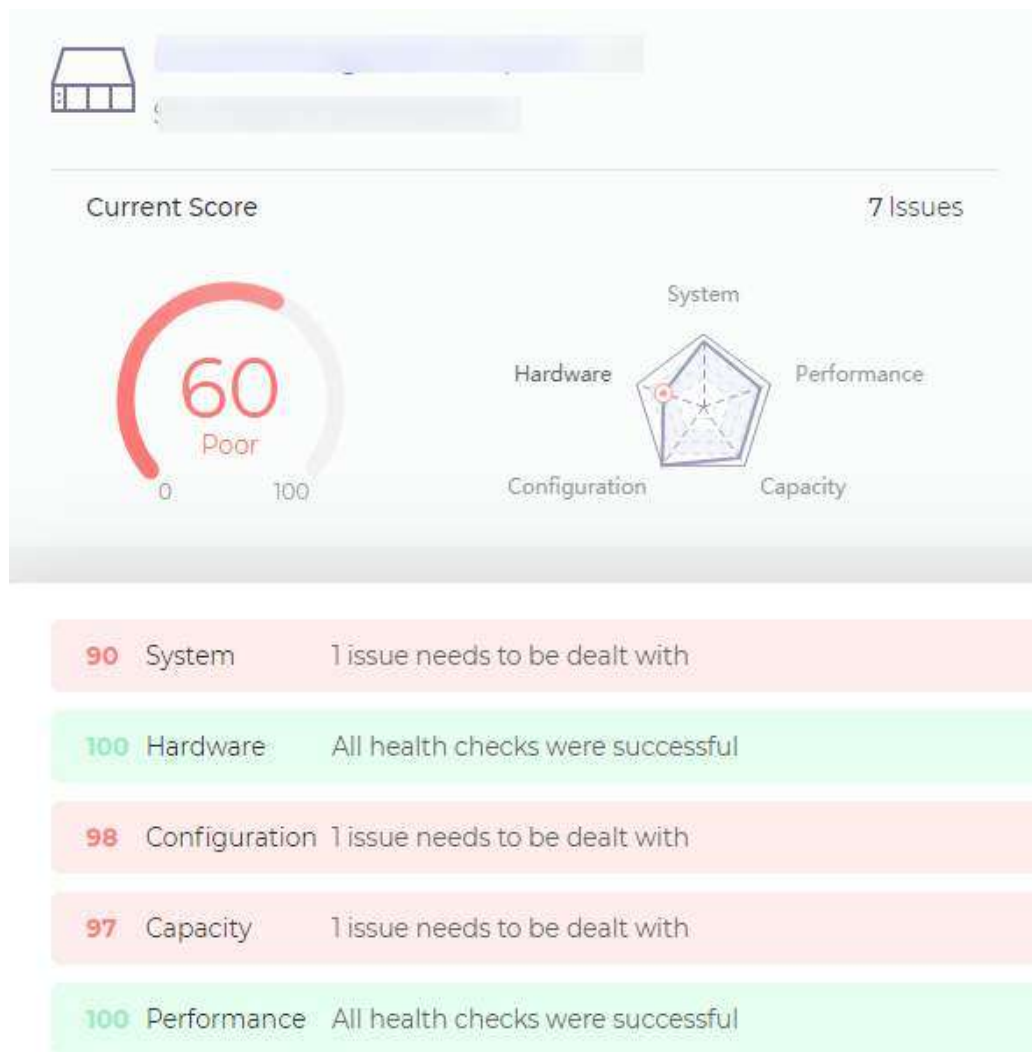
Inspection is generally performed monthly or quarterly. As a result, users cannot detect problems in a timely manner.

- Inspection depth

Users can only check whether the current device is faulty. System, hardware, configuration, capacity, and performance risks are not analyzed.

eService evaluates device health in real time from system, hardware, configuration, capacity, and performance dimensions, detects potential risks, and displays device running status based on health scores. In addition, eService provides solutions to prevent risks.

**Figure 10-10** Device health evaluation details

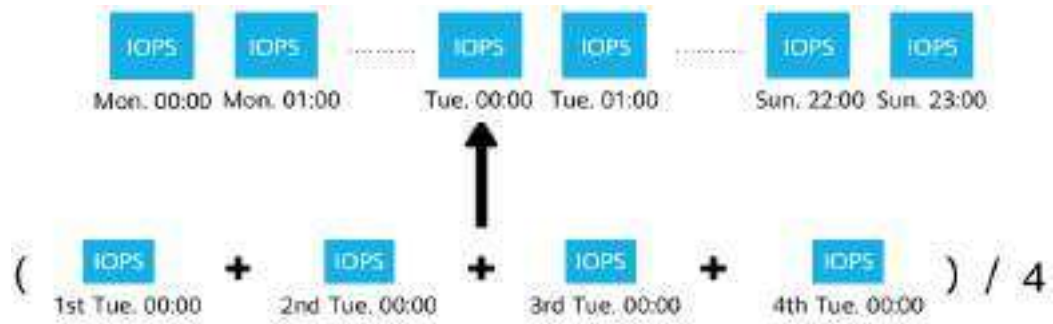


### 10.3.6 Performance Fluctuation Analysis

Periodic service operations (such as scheduled snapshot or SmartTier) or temporary changes (such as online upgrade, capacity expansion, and parts replacement) should be performed during off-peak hours to avoid affecting online services. Traditionally, O&M personnel estimate the proper time window according to experiences or through performance indicators of a past period of time.

Based on historical device performance data, eService analyzes performance fluctuations from the aspects of load, IOPS, bandwidth, and latency. Users can view the service period rules from the four dimensions and select a proper time window to perform periodic operations (such as scheduled snapshot) or temporary service changes (such as online upgrade, capacity expansion, and parts replacement) to prevent impact on services during peak hours.

**Figure 10-11** Working principles of weekly performance fluctuation analysis



eService calculates performance indicator values of each hour from Monday to Sunday based on performance data in the past four weeks. For example, the IOPS is calculated as follows: Calculate the sum of the IOPS on each hour in the past four weeks and then divide the sum by 4 to obtain the weekly performance fluctuation. For daily and monthly calculation, the methods are similar.

Users can view the service performance statistics by day, week, or month as required, as shown in the following figure.

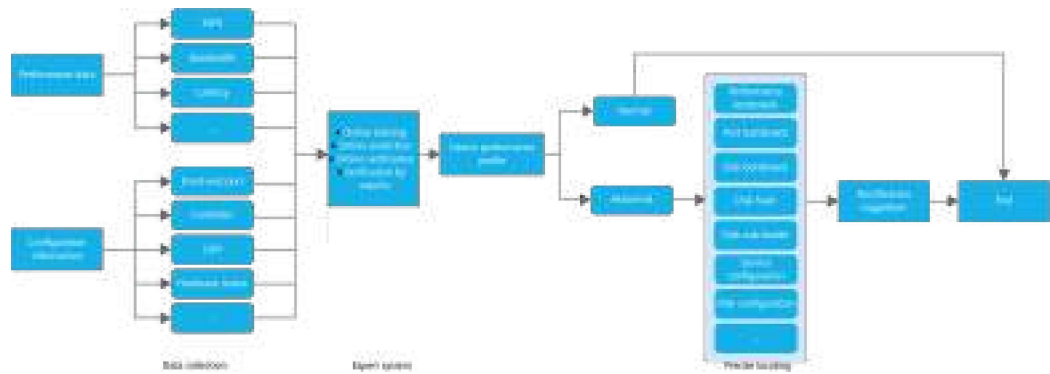
**Figure 10-12** Weekly performance fluctuation



### 10.3.7 Performance Exception Detection

Enterprises are concerned most about service running. However, because performance problems are complicated and difficult to identify and solve in advance, such problems get worse before being detected, affecting services and causing losses to enterprises. eService provides the performance exception detection function. For service latency, the deep learning algorithm is used to learn service characteristics based on historical performance data. Combining service characteristics with the industry and Huawei expertise, eService obtains device performance profiles which show real-time exceptions, precisely locates faults, and provides rectification suggestions.

**Figure 10-13** Working principles of performance exception detection

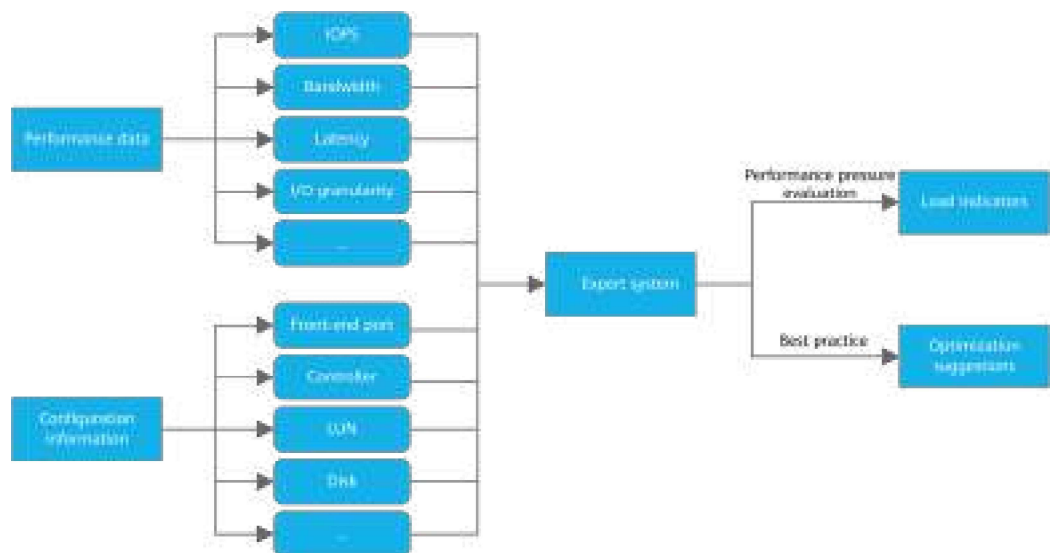


### 10.3.8 Performance Bottleneck Analysis

After services go online, O&M personnel are concerned about device performance pressure and stable service running. Due to complicated factors that affect device performance, such as hardware configuration, software configuration, service type, and performance data, multiple indicators need to be compared and analyzed concurrently and manually. Therefore, performance pressure evaluation and performance tuning are big challenges.

eService performance bottleneck analysis covers device configurations and performance data. eService automatically evaluates device performance pressure, provides clear overall device loads and loads on each component based on Huawei expertise, identifies performance bottlenecks, and provides optimization suggestions. Users can make adjustment based on these suggestions to ensure stable service running.

**Figure 10-14** Working principles of performance bottleneck analysis



Users can view the overall device loads and loads on each component, as shown in [Figure 10-15](#).

**Figure 10-15** Performance bottleneck analysis overview



## 10.4 OceanProtect Appliance Upgrade

The storage system and backup software of OceanProtect appliance need to be upgraded separately. This section describes how to upgrade the storage system, backup software, and backup client (ProtectAgent).

### 10.4.1 Upgrading Storage System

The storage system of the OceanProtect appliance supports non-disruptive upgrade (NDU).

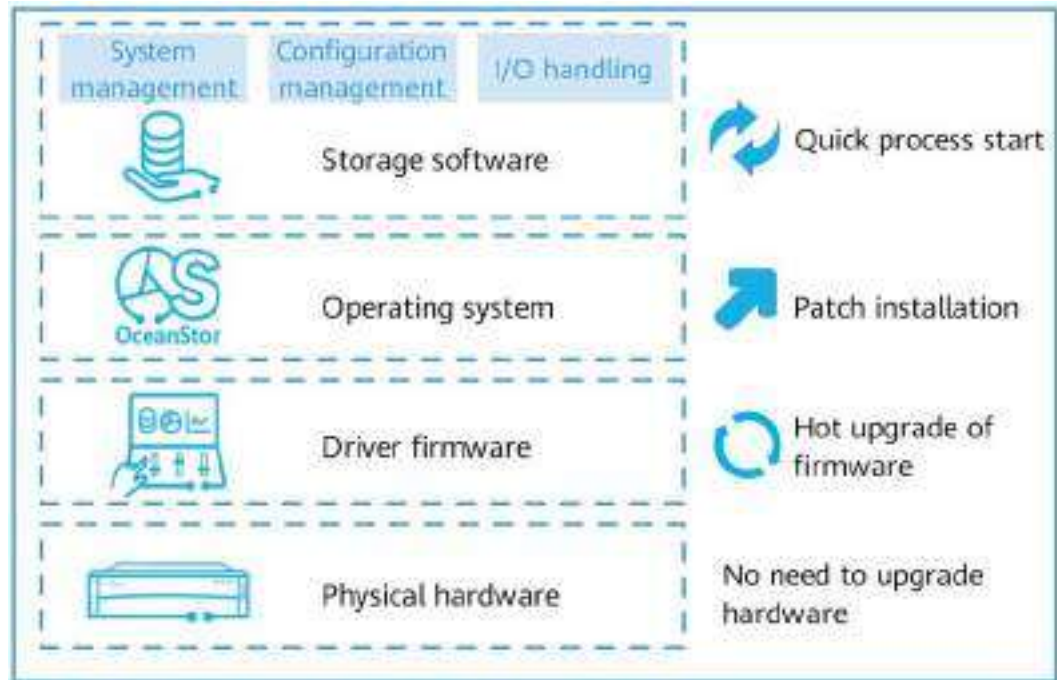
Storage system upgrade usually requires controller restart and service switchover between controllers to ensure service continuity, which greatly affects service performance.

OceanProtect provides an upgrade mode that does not require controller restart. During the upgrade, the link between the storage system and the backup service is not interrupted and link switchover is not involved. Services are not affected and performance can quickly recover after the upgrade is complete.

### Component-based Upgrade

A storage system can be divided into physical hardware, driver firmware, OS, and storage software. They are upgraded in different ways to complete the system upgrade of OceanProtect, as shown in [Figure 10-16](#).

**Figure 10-16** Component-based upgrade



- Physical hardware does not need to be upgraded.
- Driver firmware, including the BIOS, CPLD, and interface module firmware, supports hot upgrade without controller restart.
- The OS is upgraded by installing hot patches.
- Storage software is user-mode processes. Earlier processes are killed and new processes are started using upgraded codes, which are completed within seconds.

The component-based upgrade eliminates the need of controller restart, so the storage system NDU upgrade is transparent to the backup services.

## Zero Performance Loss and Short Upgrade Duration

OceanProtect does not involve service switchover. Therefore, an upgrade has nearly no impact on host performance, and performance can recover to 100% within seconds. As controller restart and link switchover are not involved, you do not need to collect host information or perform compatibility evaluation. The upgrade process (from package import to upgrade completion) takes less than 30 minutes, 10 minutes in general.

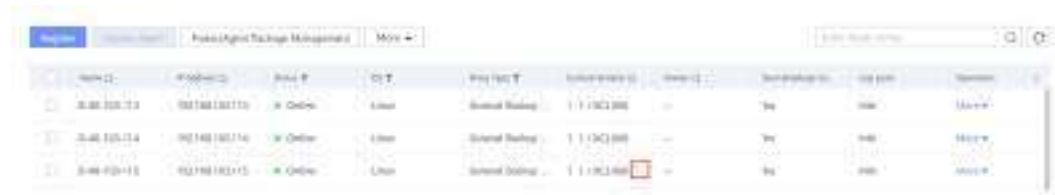
### 10.4.2 Upgrading Backup Software

The backup software is upgraded separately from the storage system. During the upgrade, the container image is completely replaced, and the service container POD is restarted to restart the application. This ensures that the upgrade does not affect the overall automation job.

### 10.4.3 Upgrading Backup Clients in Batches

After the backup software is upgraded, the client (ProtectAgent) software version must be updated. The OceanProtect supports batch client upgrade and visualized operations. The server proactively monitors the client version. If the client version is earlier than the current server version, the client monitoring page displays a message indicating that the client can be upgraded. Users can select the client to be upgraded and perform batch upgrade.

**Figure 10-17** Upgrading backup clients in batches



During the upgrade, an agent update job is generated for each client. If a single job fails, rollback is automatically performed, which does not affect the upgrade of other clients.

# 11 Acronyms and Abbreviations

Acronym/ Abbreviation	Full Name	Description
FRU	Field Replaceable Unit	Field replaceable unit
FlashLink®	FlashLink®	Disk-controller coordination technology
CK	Chunk	Data block
CKG	Chunk Group	Data block group
DIF	Data Integrity Field	Data integrity field
RDMA	Remote Direct Memory Access	Remote direct memory access
FC	Fiber Channel	Fibre channel
FS	File System	File system
FTL	FLASH Translation Layer	FLASH translation layer
GC	Garbage Collection	Garbage collection
SSD	Solid State Disk	Solid state disk
LUN	Logical Unit Number	Logical unit number
OLAP	On-Line Analytical Processing	Online analytical processing
OLTP	On-Line Transaction Processing	Online transaction processing system
OP	Over-Provisioning	Reserved space
RAID	Redundant Array of Independent Disks	Redundant array of independent disks

Acronym/ Abbreviation	Full Name	Description
RAID-TP	Redundant Array of Independent Disks-Triple Parity	Redundant array of independent disks - triple parity
SAS	Serial Attached SCSI	Serial attached SCSI
SCSI	Small Computer System Interface	Small computer system interface
SSD	Solid State Disk	Solid state disk
T10 PI	T10 Protection Information	T10 protection information
VDI	Virtual Desktop Infrastructure	Virtual desktop infrastructure
VSI	Virtual Server Infrastructure	Virtual server infrastructure
WA	Write amplification	Write amplification
Wear Leveling	Wear Leveling	Wear leveling
TCO	Total Cost of Ownership	Total cost of ownership
DC	Data Center	Data center
DCL	Data Change Log	Data change log
TP	Time Point	Time point
GUI	Graphical User Interface	Graphical user interface
CLI	Command Line Interface	Command line interface
FIM	front-end interconnect I/O module	Front-end interconnect I/O module
SCM	storage class memory	Storage class memory
FRU	field replaceable unit	Field replaceable unit
PI	Protection Information	Protection information
SFP	Similar Fingerprint	Similar fingerprint
DTOE	Direct TCP/IP Offloading Engine	TOE passthrough technology
NUMA	non-uniform memory access	Non-uniform memory access
ROW	redirect-on-write	Redirect-on-write

Acronym/ Abbreviation	Full Name	Description
PID	Proportional Integral Derivative	Proportional calculus algorithm
SMP	symmetrical multiprocessor system	Symmetrical multiprocessor system

## PROCURAÇÃO

A firma MICROWARE TECNOLOGIA DE INFORMAÇÃO LTDA, com sede na Rua Samuel Morse, 120 – 14º andar – Cidade Monções – São Paulo / SP, inscrita no CNPJ sob o nº 01.724.795/0001-43, e suas filiais em Niterói /RJ, CNPJ nº 01. 724.795/0004-96, Brasília / DF, CNPJ nº 01.724.795/0006-58 e Vila Velha / ES, CNPJ 01.724.795/0007-39 representadas pela Sra. Kátia Maria M. T. Valente dos Reis, portadora do RG nº 5.651.651-4 SSP/SP e do CPF/MF nº 817.130.967-49, nos termos de seu Estatuto Social, pela presente CREDENCIA o funcionário **Eduardo Porto Rangel**, portador do Registro nº 1797578, expedida pelo SSP/ES e do CPF/MF nº 100531957-01 para representá-la em licitações e pregões, pelo período de 06 (seis) meses a partir da presente data com poderes para formular ofertas e lances, assinar propostas, concordar, desistir, renunciar, transigir, assinar atas e outros documentos, acompanhar todo o processo licitatório até o seu final, tomar ciência de outras propostas da Comissão de Licitação, bem como efetivar depósitos em caução, caução em fiança bancária e seguro garantia, podendo para tanto, praticar todos os atos necessários para o bom e fiel cumprimento deste mandato.

Niterói, 08 de agosto de 2025.

KATIA MARIA MATTOS  
TAVARES VALENTE DOS  
REIS:81713096749

Assinado de forma digital por  
KATIA MARIA MATTOS TAVARES  
VALENTE DOS REIS:81713096749  
Dados: 2025.08.08 17:50:14 -03'00'

Kátia M. M. T. Valente dos Reis  
Diretora - Sócia

**MICROWARE TECNOLOGIA DE INFORMAÇÃO LTDA**

