Huawei Cloud Adoption Framework

Issue	01
Date	2025-06-19

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions

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.

Security Declaration

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

Contents

1 Introduction to Cloud Adoption Framework

1.1 Purpose

Cloud computing fundamentally changes the way IT infrastructure and application systems are built, operated, and managed. In the traditional mode, organizations need to purchase, install, and operate their own hardware and software, including servers, storage devices, network devices, virtualization software, operating systems, database management software, and middleware. This approach involves lengthy deployment cycles, high operational burdens, and significant upfront costs.

In the cloud computing mode, service providers handle the construction and maintenance of IT infrastructure, allowing organizations to focus solely on developing and deploying application systems. Resources are available on demand, enabling rapid deployment, adjustments, and scaling. This significantly reduces operational burdens and lowers upfront investments. Cloud computing offers exceptional flexibility, reliability, and scalability. However, transforming an entire organization to the cloud is a comprehensive undertaking, involving every facet of processes, structures, and technologies. A mature and consistent approach is essential to ensure a successful transformation and maximize business value.

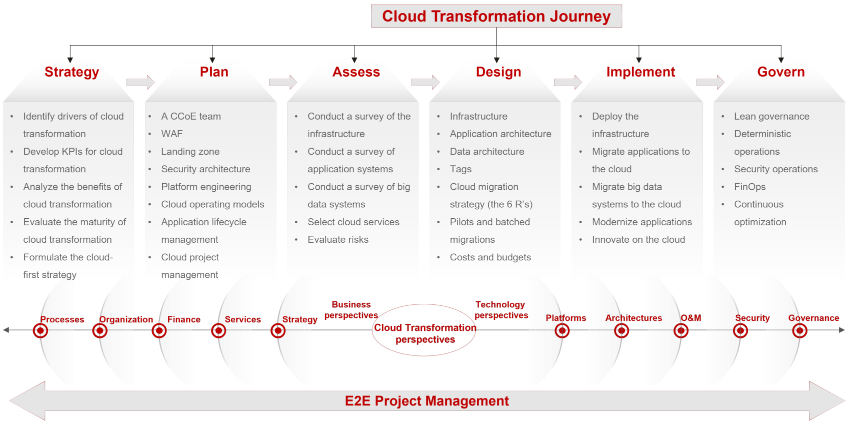
Huawei Cloud Adoption Framework (CAF) is an end-to-end lifecycle framework for cloud transformation. It covers all phases of the cloud transformation journey, including strategy development, top-level planning, survey and evaluation, solution design, adoption, implementation, and O&M governance. CAF provides methodologies, best practices, tools, and templates for each phase of the cloud transformation journey. It helps stakeholders such as business decision makers, IT decision makers, financial experts, O&M experts, and security experts make correct decisions in each phase of the cloud transformation journey and maximize the value of cloud computing. By following the best practices of CAF, your organization can better align with business and technology strategies and ensure the success of cloud transformation.

The methods, best practices, tools, and templates offered by CAF are drawn from the expertise of Huawei Cloud, partners, and customers in cloud migration, utilization, and management. They are constantly updated based on cloud transformation experiences.

1.2 Overall Framework

Huawei Cloud CAF provides comprehensive and systematic cloud transformation methodologies and best practices, covering the entire cloud transformation journey and the perspectives of all stakeholders involved in cloud transformation. The following figure shows the overall framework of CAF.

Figure 1-1 Overall framework of CAF


Cloud Transformation Journey
Strategy
Identify drivers of cloud transformation
Develop KPIs for cloud transformation
Analyze the benefits of cloud transformation
Evaluate the maturity of cloud transformation
Formulate the cloud-first strategy
Plan
A CCoE team
WAF
Landing zone
Security architecture
Platform engineering
Cloud operating models
Application lifecycle management
Cloud project management
Assess
Conduct a survey of the infrastructure
Conduct a survey of application systems
Conduct a survey of big data systems
Select cloud services
Evaluate risks
Design
Infrastructure
Application architecture
Data architecture
Tags
Cloud migration strategy (the 6 R's)
Pilots and batched migrations
Costs and budgets
Implement
Deploy the infrastructure
Migrate applications to the cloud
Migrate big data systems to the cloud
Modernize applications
Innovate on the cloud
Govern
Lean governance
Deterministic operations
Security operations
FinOps
Continuous optimization
Processes
Organization
Finance
Services
Strategy
Business perspectives
Cloud Transformation perspectives
Technology perspectives
Platforms
Architectures
O&M
Security
Governance
E2E Project Management

The cloud transformation journey consists of the following six phases in chronological order.

1. **Strategy formulation**: Formulate a cloud transformation strategy that aligns with the organization's business and technology strategies. In this phase, you need to analyze stakeholder interests, identify key drivers for cloud transformation, assess cloud maturity, set cloud transformation objectives, evaluate potential benefits, and develop a cloud transformation strategy that aligns with the organization's business strategy.
2. **Top-level planning**: Cloud transformation is a systematic project. It is not simply migrating application systems to the cloud. It necessitates a top-level framework for organizational structures, processes, and technologies. At the organizational level, you need to set up a Cloud Center of Excellence (CCoE) to lead, coordinate, and promote the cloud transformation project. At the technical level, you need to design the landing zone, security architecture, and platform engineering based on the Well-Architected Framework (WAF). At the process level, you need to design an optimized cloud operating model tailored to the organization's IT operations framework. Additionally, you need to formulate an application lifecycle management process aligned with the cloud operating model to fully realize the business value of cloud computing.
3. **Survey and evaluation**: Conduct a survey on the organization's IT infrastructure, application systems, and big data platforms, analyze static configurations and dynamic runtime data, and evaluate the current statuses to select appropriate cloud services and provide actionable insights for the subsequent solution design.
4. **Solution design**: In the top-level design framework of cloud transformation, design the cloud technical architecture, cloud application architecture, and cloud data architecture in detail based on the survey and evaluation results,

the business architecture of the organization, and the WAF framework. Ensure the security, reliability, and high performance of cloud infrastructure and application systems through WAF design. In this phase, you need to formulate POC and batch migration plans for application systems based on their importance, select appropriate migration policies based on their characteristics, formulate a cloud cost budget plan, and finally output the low-level design.

5. **Adoption and implementation**: Set up a landing zone based on the low-level design, deploy a scalable network infrastructure, and configure security and O&M baselines. Then, migrate application systems and big data platforms to or deploy them on the cloud, modernize applications on the cloud platform, or innovate applications and services on the cloud based on its innovative technologies.
6. **O&M governance**: After application systems are migrated to or deployed on the cloud, the O&M governance phase begins. During this phase, lean governance, deterministic O&M, security operations, and cost operations are essential for cloud infrastructure, application systems, and big data platforms. Continuous optimization is performed following the WAF framework.

Cloud transformation projects involve many departments and stakeholders in the organization. These stakeholders participate in the decision-making of cloud transformation projects or influence each phase of the cloud transformation journey. As shown in **Table 1-1**, different stakeholders have different concerns and different business or technical perspectives.

Huawei Cloud CAF provides appropriate suggestions across all perspectives. These suggestions serve as a foundation for decision-making and actions, enabling your organization to develop target strategies tailored to your business characteristics and preferences.

Table 1-1 Cloud transformation perspectives

Category	Perspective	Focuses	Stakeholders
Business	Strategy	● Support the business strategy and digital strategy of the organization, and fully leverage the advantages of cloud computing to build the core competitiveness of the organization.	Senior management personnel (CXOs)
Business	Services	● Improve business continuity and support sustainable development. ● Accelerate the launch of new services to quickly meet changing market requirements. ● Innovate businesses, products, or models based on innovative technologies on the cloud to bring incremental benefits to the organization.	Business directors and CIO

Category	Perspective	Focuses	Stakeholders
Business	Finance	● Compare and analyze the TCO before and after cloud adoption and reduce the IT TCO. ● Continuously optimize the performance efficiency and cost-effectiveness of cloud resources. ● Bring new revenue through user experience improvement and business innovation.	CFO and financial experts
Business	Organization	● Build the organizational structure for cloud transformation and define the roles and responsibilities of cloud transformation talents. ● Develop cloud transformation performance appraisal indicators for selecting, using, cultivating, and retaining cloud transformation talents.	CIO and HR experts
Business	Processes	● Optimize IT service processes and O&M processes based on the characteristics of the cloud platform and cloud services to support the fast iteration and secure and stable running of upper-layer application systems.	CIO and IT directors
Technology	Platforms	● Build enterprise-level, highly secure, reliable, high-performance, and easy-to-expand IT infrastructure or technical platforms based on the cloud platform and its cloud services to provide compute, storage, network, security, database, and middleware services for upper-layer application systems. This helps application teams quickly develop, test, deploy, and efficiently operate application systems to support the secure and stable running of application systems.	CIO, CTO, IT directors, IT O&M experts, application development and test experts, and application O&M experts
Technology	Architectures	● Build a highly secure, reliable, high-performance, and easy-to-expand technical architecture, application architecture, and data architecture based on the cloud platform and its cloud services.	CTO and cloud architects
Technology	O&M	● Build a comprehensive cloud-based IT O&M system based on the characteristics of the cloud platform and its cloud services to monitor, generate alarms for, and locate and rectify faults for IT infrastructure and application systems, ensuring the long-term stable running of IT infrastructure and application systems.	CTO, IT O&M experts, and application O&M experts

Category	Perspective	Focuses	Stakeholders
Technology	Security	● Build a comprehensive security protection system and continuous security operations mechanism based on the characteristics of the cloud platform and its cloud services to ensure the confidentiality, integrity, and availability of IT infrastructure and application systems.	CISO and security experts
Technology	Governance	● Build a comprehensive IT governance system based on the characteristics of the cloud platform and its cloud services to enable centralized and lean governance of personnel, finances, resources, permissions, and compliance on the cloud. By doing so, organizations can effectively mitigate cloud transformation risks, optimize business value, and ensure sustained growth and success.	CIO and IT governance experts

The cloud transformation journey is a long, multi-faceted process involving a large number of personnel and tasks. To ensure success, enterprises should assign dedicated project managers to oversee the entire project. Employing scientific project management methods and actionable plans is crucial for enhancing the efficiency and quality of cloud transformation and ultimately driving the achievement of strategic transformation goals.

The directory structure of Huawei Cloud CAF aligns with the six phases of the cloud transformation journey. Each section outlines focus areas and practical suggestions from both business and technology perspectives, providing comprehensive guidance for cloud transformation.

1.3 Intended Audience

CAF covers the entire cloud transformation journey and engages various roles across organizational departments. These roles can find their responsibilities within CAF and use its guidance as a foundation for effective decision-making and actions.

- CXOs (including CEO, CIO, CTO, COO, and CFO)
- Business directors
- IT directors and technical directors
- Financial experts
- HR directors
- Cloud architects, application architects, data architects, and network architects
- IT governance experts

- O&M directors and IT O&M experts
- CISO, security experts, and compliance audit experts
- Application development experts and application test experts
- Application O&M experts
- Migration implementation engineers
- Project managers

1.4 Glossary

Huawei Cloud CAF includes a wide range of terms related to IT and cloud computing. As interpretations may vary among readers, a glossary is provided to ensure clarity and prevent misunderstandings.

Table 1-2 Glossary

Term	Description
CAF	Cloud Adoption Framework is an end-to-end lifecycle framework for cloud transformation. It covers all phases of the cloud transformation journey, including strategy development, top-level planning, survey and evaluation, solution design, adoption and implementation, and O&M governance. CAF provides methodologies, best practices, tools, and templates for each phase of the cloud transformation journey.
WAF	The Well-Architected Framework is Huawei Cloud's technical framework, providing design guidance and best practices to address critical challenges faced after migrating customer services to the cloud. Building on Huawei and industry best practices, WAF addresses five core architectural principles: resilience, security, performance efficiency, cost optimization, and operational excellence. It empowers customers to create exceptional technical and application architectures on Huawei Cloud. WAF is also short for web application firewall. You need to determine the specific meaning of WAF based on the context.

Term	Description
IT infrastructure	An IT infrastructure is a platform-based environment for ensuring the secure and stable running of all application systems within an organization. Underlying IT resources (such as data centers, hardware, networks, and virtualization) are abstracted, managed, and optimized to provide necessary compute, storage, network, database, middleware, and other IT services and a stable, reliable, and efficient runtime environment for application systems. This accelerates the development, testing, and deployment of application systems. Cloud computing can greatly accelerate the construction and expansion of your organization's IT infrastructure and simplify the O&M management of the IT infrastructure, enabling your organization to focus on high-value fields such as application system development and O&M. IT infrastructure is also called technical platform or technical middle end. The IT infrastructure built based on cloud computing is also called a cloud infrastructure.
Application system	An application system is a software system designed to complete specific tasks or solve specific problems. It supports specific business processes and scenarios in an organization. It usually consists of a series of interrelated applications, databases, middleware, configuration files, and documents, and runs on the IT infrastructure. An application system can be independent or part of a larger application system. Application systems are also called business systems, information systems, business application systems, business information systems, and workloads.
IT management system	IT management systems, such as the security operations center, Identity and Access Management (IAM), and monitoring and O&M system, are IT support and management systems established to support the long-term secure and stable running of application systems.
Cloud services	Cloud services are various IT services provided by cloud service providers through the Internet or dedicated networks, including compute, storage, network, security, O&M management, database, middleware, big data, and AI services. Users can access these services on demand without purchasing and maintaining physical hardware and software infrastructure. They only need to pay for the resources they actually use. The main types of cloud services include Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).
IaaS	IaaS provides compute, storage, network, and other infrastructure resources as cloud services. Users can deploy and run any software based on these cloud services, including operating systems, databases, middleware, and applications. Users do not control the underlying cloud infrastructure, but they can control the operating system, storage, applications, and possibly limited network components (such as host firewalls).

Term	Description
PaaS	PaaS provides platform resources such as DevOps tool chains, middleware, databases, microservice engines, and big data platforms as cloud services. Users can develop, test, deploy, and maintain applications based on these cloud services. Users do not manage or control the underlying cloud infrastructure and platform resources such as middleware, databases, and microservice engines, but they can control the deployed applications and their related data.
SaaS	SaaS provides applications as cloud services. Applications can be accessed through various client devices, such as web browsers, mobile apps, or APIs. Users do not manage or control the underlying cloud infrastructure, platform resources, or applications, but they can control the data generated during application running.
Cloud resources	Cloud resources are IT resource instances created by users using cloud services. They include compute, storage, network, security, O&M management, database, middleware, big data processing, and AI resources. Users can use cloud resources to quickly build the IT infrastructure required by upper-layer application systems.
Cloud service provider	A cloud service provider (CSP) is a company that provides cloud services. CSPs design, build, and manage large-scale cloud data centers, and provide on-demand cloud services to customers through the Internet or private networks. CSPs are also called cloud vendors.
Landing zone	In aviation, a landing zone refers to the area where an aircraft (such as a helicopter) can land safely. This term is used to name the solution for securely and smoothly migrating application systems to and running them on the public cloud. A landing zone is designed to create a multi-account runtime environment on the cloud that features robust architecture, security, compliance, and scalability. Built on this foundation, a streamlined IT governance system is established to enable structured and centralized management of personnel, resources, finances, permissions, and compliance. The ultimate goal is to systematically address the IT governance and security compliance challenges arising from enterprises' large-scale adoption of cloud services.

Term	Description
Account	Huawei Cloud accounts function as resource containers where users can deploy any resources and application systems. They are isolated from each other. Faults and security risks in one account will not affect or spread to other accounts. Huawei Cloud accounts define the security management boundaries. Each account has an independent identity and permissions management system. Without authorization, users in an account cannot access resources, data, and applications in other accounts. From the perspective of IT governance, accounts are classified into management accounts and member accounts. Management accounts are used to create and manage organizations, member accounts, and SCP policies. Member accounts are used to carry specific application systems (such as ERP) or IT management responsibilities. From the perspective of financial governance, accounts are classified into master accounts and member accounts. They can form centralized or decentralized financial management relationships. The management account and the master account are the same entity. The member accounts are also the same entity.
Cloud organizational structure	The cloud organizational structure organizes cloud resources in a hierarchical structure. It consists of multiple levels of organizational units and accounts. An organizational unit can contain multiple lower-level organizational units and accounts. According to **Conway's law**, the cloud organizational structure is usually consistent with the enterprise's business structure.
Cloud Center of Excellence (CCoE)	CCoE is a centralized team established by an enterprise for cloud transformation. It is responsible for the entire cloud transformation journey, including strategy formulation, top-level planning, survey and evaluation, solution design, adoption and implementation, and O&M governance. Its goal is to help enterprises maximize the value of cloud computing and ensure the successful implementation of cloud transformation projects by providing best practices, guidance, and resources.
Cloud operating model (COM)	COM is a collection of processes and regulations designed for enterprises to adopt, manage, and operate cloud computing technologies to maximize the business value of cloud computing. The cloud operating model must align with the organization's business operating model and define the collaboration between the CCoE and the application team. An effective cloud operating model enables the CCoE to centrally manage the entire cloud platform, enhancing efficiency and minimizing technical risks. Meanwhile, the application team gains the flexibility to manage and utilize cloud resources, fostering rapid application innovation.

Term	Description
Digital transformation	Digital transformation refers to the comprehensive reshaping and innovation of an organization's business models, operating processes, products, and services using digital technologies (such as cloud computing, big data, IoT, AI, and blockchain) to adapt to the rapidly changing market environment and meet customers' increasing requirements. Through digital transformation, organizations can not only improve efficiency and competitiveness, but also create new value and growth opportunities.
Cloud transformation	Cloud transformation refers to the process of migrating an organization's IT infrastructure, application systems, and business processes to a cloud computing platform, or using cloud computing technologies to reconstruct and optimize its business models and operating processes. It is not just a simple "migration to the cloud", but a comprehensive transformation involving strategy, technology, organization, and processes. The goal is to leverage the advantages of cloud computing to improve business agility and continuity, reduce costs, and drive business innovation. Cloud transformation plays a crucial role in enabling digital transformation and significantly accelerates an organization's journey toward achieving it.
Business unit	A business unit is an independent operational and managerial unit within an enterprise. It is structured based on factors such as products, services, markets, customer groups, or functional domains. Each business unit usually has its own strategic objectives, responsibilities, resources, and performance indicators, and is responsible for specific business activities and market areas. A business unit can be a subsidiary, business group, product line, department, or project team.

1.5 Acronyms and Abbreviations

Table 1-3 Acronyms and abbreviations (in alphabetical order)

Acronym or Abbreviation	Full Name
AIOps	Artificial Intelligence for IT Operations
AOM	Application Operations Management
ALM	Application Lifecycle Management
CAF	Cloud Adoption Framework
Capex	Capital expenditure
CBH	Cloud Bastion Host

Acronym or Abbreviation	Full Name
CC	Cloud Connect
CCE	Cloud Container Engine
CCI	Cloud Container Instance
CCM	Cloud Certificate Manager
CCoE	Cloud Center of Excellence
CFW	Cloud Firewall
CMDB	Configuration Management Database
CMM	Cloud Maturity Model
CNCF	Cloud Native Computing Foundation
COC	Cloud Operations Center
CSMS	Cloud Secret Management Service
CSP	Cloud Service Provider
CSR	Corporate social responsibility
DBSS	Database Security Service
DC	Direct Connect
DCMM	Data Management Capability Maturity Assessment Model
DDoS	Distributed Denial of Service
DevOps	Development and Operations
DevSecOps	Development, Security, and Operations
DEW	Data Encryption Workshop
DSC	Data Security Center
ECS	Elastic Cloud Server
EIP	Elastic IP Address
ELB	Elastic Load Balancing
ER	Enterprise Router
ESW	Enterprise Switch
EVS	Elastic Volume Service
FinOps	Finance Operations
GRC	Governance, Risk & Compliance

Acronym or Abbreviation	Full Name
HSM	Hardware Security Module
HSS	Host Security Service
IaaS	Infrastructure as a service
IaC	Infrastructure as Code
IAM	Identity and Access Management
IDC	Internet Data Center
IDP	Internal Developer Platform
IoT	Internet of Things
ITSM	IT Service Management
ITSS	Information Technology Service Standards
KMS	Key Management Service
KPS	Key Pair Service
LLM	Large Language Model
MFA	Multi-Factor Authentication
MSP	Managed Service Provider
MTBF	Mean Time Between Failure
MTTR	Mean Time To Repair
NAT	Network Address Translation
OACA	Open Alliance for Cloud Adoption
OBS	Object Storage Service
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
Opex	Operational expenditure
PaC	Policy as Code
PaaS	Platform as a service
PUE	Power Usage Effectiveness
QPS	Query Per Second
ROI	Return on Investment
RPO	Recovery Point Objective

Acronym or Abbreviation	Full Name
RTO	Recovery Time Objective
SaaS	Software as a Service
SCIM	System for Cross-domain Identity Management
SCP	Service Control Policy
SecMaster	Security Master
SFS	Scalable File Service
SLA	Service Level Agreement
SLO	Service Level Objective
SNAT	Source Network Address Translation
SOC	Security Operation Center
SRE	Site Reliability Engineering
SSO	Single Sign-On
TOGAF	The Open Group Architecture Framework
TPS	Transactions Per Second
VPC	Virtual Private Cloud
VPN	Virtual Private Network
WAF	Well-Architected Framework
WAF	Web Application Firewall

2 Strategy Development

2.1 Overview

In the new era of globalization and rapid technological development, the market environment is undergoing changes. Consumer demands are constantly upgrading, market competition is becoming increasingly fierce, and traditional business models are facing huge challenges. Faced with these pressures, digital transformation has become the only way for enterprises to achieve business innovation and sustainable development.

Digital transformation refers to the comprehensive reshaping and innovation of an organization's business models, operations processes, products, and services using digital technologies (such as IT infrastructure, databases, big data, IoT, AI, and blockchain) to adapt to the rapidly changing market environment and meet customers' ever-increasing requirements. Mainstream cloud computing platforms now integrate a wide array of secure, highly available digital technologies. These technologies are delivered globally as on-demand cloud services in the fastest and most efficient manner. Cloud transformation plays a pivotal role in driving digital transformation. Looking ahead, cloud services will serve as the foundation for most digital services. Electricity was the hallmark of the industrial era, with consumption levels serving as indicators of a society's industrial progress. In the digital era, cloud computing has become the hallmark, where cloud usage reflects the level of digital advancement within a society.

Cloud transformation is the process of migrating the IT infrastructure, application systems, and business processes of an organization to the cloud computing platform or using cloud computing technologies to reconstruct and optimize the business models and operating processes. It is not only a simple migration of data and applications to the cloud, but also a comprehensive transformation involving strategies, technologies, organizations, and processes.

The first step of cloud transformation is to formulate a cloud transformation strategy, which requires comprehensive planning and careful preparation. Developing a cloud transformation strategy goes beyond technical considerations. It requires alignment with the organization's business and digital strategies. Strategic objectives also need to be aligned with senior leaders and all stakeholders across the company. It typically involves several key steps: analyzing stakeholder interests, identifying the drivers of cloud transformation, assessing the organization's cloud maturity, setting clear transformation objectives, and evaluating the potential benefits. The following sections provide a detailed breakdown of each aspect.

2.2 Analyzing Stakeholder Interests

The first step in creating a cloud transformation strategy is identifying the key stakeholders. This includes identifying everyone involved in the decision-making process, analyzing their interests in detail, and working collaboratively to determine the drivers of cloud transformation, evaluate its benefits, and design a strategy that aligns with the organization's objectives. The following table lists some common stakeholders and their interests, as well as their ways of participating in cloud transformation strategy decision-making and project execution.

Table 2-1 Cloud transformation stakeholders and interests

Stakeholder	Interests	Ways of Participation
CEO	● Promote the achievement of enterprise strategic goals and improve business agility and market competitiveness. ● Promote revenue and profit growth and ensure sustainable development of the enterprise. ● Reduce operational risks and ensure business continuity. ● Accelerate business innovation and explore new markets and new business models. ● Enhance the brand awareness and social responsibility and focus on sustainable development.	● Comprehensively lead the formulation and implementation of the cloud transformation strategy, act as the final decision maker of the project, and ensure that the cloud transformation strategy is aligned with the company's business strategy. ● Coordinate resources of various departments to ensure cross-departmental cooperation. ● Regularly review project progress and provide strategic guidance and support. ● Identify and evaluate the driving forces and expected benefits of cloud transformation with the executive team.

Stakeholder	Interests	Ways of Participation
CIO	● Improve the IT department's service capabilities to rapidly respond to business requirements. ● Promote technological innovation to make the technical architecture more advanced and flexible. ● Optimize the IT cost structure and improve the resource utilization. ● Enhance information security and ensure the reliability of data and application systems.	● Lead the technical planning and roadmap formulation for the cloud transformation strategy to ensure its alignment with the company's business strategy. ● Facilitate collaboration between the IT team and other business departments to ensure that the technical solution meets business requirements. ● Manage the selection of and cooperation with cloud service providers. ● Oversee the execution of cloud transformation projects to ensure they are implemented according to the established plans.
COO	● Optimize business processes to improve operational efficiency and quality. ● Ensure business continuity and reduce operational risks. ● Support business expansion and innovation to meet market demands.	● Participate in the formulation of cloud transformation strategies and provide operational requirements and suggestions. ● Coordinate the resources of the operations department to support project implementation. ● Monitor the impact of cloud transformation on business operations and ensure smooth transition.
CTO	● Introduce advanced technologies to improve the technological competitiveness of the enterprise. ● Ensure the scalability and flexibility of the technical architecture to meet future business requirements. ● Promote technological innovation and support the development of new products and services.	● Lead the design and evaluation of technical solutions to ensure the technical feasibility of cloud transformation. ● Guide the work of the technical team to ensure that the technical implementation is consistent with the strategic objectives. ● Work with the CIO to formulate technical standards and specifications.

Stakeholder	Interests	Ways of Participation
Chief information security officer (CISO)	● Ensure information security and prevent against data leakage and network attacks. ● Ensure compliance with industry and legal requirements. ● Maintain the enterprise reputation and avoid negative impacts of security incidents.	● Evaluate security risks brought by cloud transformation and formulate corresponding security policies. ● Guide the security team to implement security control measures to ensure the security of the cloud environment. ● Work with compliance audit experts to ensure adherence to security and compliance requirements.
CFO	● Optimize financial performance, reduce IT costs, and improve return on investment (ROI). ● Manage capital expenditure and operating expenditure to improve cash flow. ● Evaluate the financial risks and benefits of cloud transformation to support strategic decision-making. ● Innovate products and services based on cloud services to increase revenue.	● Participate in the cost-benefit analysis of cloud transformation and provide financial suggestions. ● Review and approve project budgets and expenditures to ensure effective use of funds. ● Develop financial KPIs for cloud transformation and monitor the achievement of financial goals.
Business directors	● Improve the performance of business departments to meet market and customer needs. ● Accelerate product and service innovation and explore new business opportunities. ● Ensure the stability and reliability of business systems to support daily operations.	● Provide business requirements and expectations and participate in the formulation of cloud transformation solutions. ● Work with the IT team to ensure that technical solutions meet business requirements. ● Coordinate resources of business teams to support project implementation and change management.

Stakeholder	Interests	Ways of Participation
IT directors	● Improve resource utilization, implement elastic scaling of IT systems, and support rapid business growth. ● Reduce IT costs through cloud transformation. ● Use the high-availability architecture and security protection measures of cloud service providers to improve the stability and security of IT systems and reduce faults and security incidents. ● Increase the value of the IT department through cloud transformation.	● Assist the CIO in formulating the cloud transformation strategy and specific cloud transformation objectives. ● Select a cloud service model suitable for the organization, evaluate solutions of different cloud service providers, and formulate technical specifications. ● Establish a dedicated cloud transformation team and cultivate and introduce cloud computing talent. ● Act as the general owner of cloud transformation projects to promote the construction of cloud infrastructure and cloudification of business systems.
HR directors	● Plan and manage talent requirements and support skill improvement required for cloud transformation. ● Promote organizational and cultural transformation to help employees adapt to new working methods. ● Design incentive mechanisms to motivate employees to participate in and support cloud transformation.	● Develop training and development plans to improve employees' cloud computing skills. ● Participate in organizational structure adjustments to ensure that the team configuration meets cloud transformation requirements. ● Participate in the formulation of KPIs for cloud transformation teams and monitor the achievement of KPIs.
O&M directors	● Improve O&M efficiency and reduce faults and downtime. ● Implement O&M automation to reduce labor costs. ● Improve system availability and reliability to support business continuity.	● Develop cloud O&M processes and standards based on the characteristics of the cloud platform. ● Promote the use of cloud O&M tools to achieve automation and intelligence. ● Train the O&M team to improve cloud O&M skills.

Stakeholder	Interests	Ways of Participation
Application architects	● Optimize application architecture to improve system performance, scalability, and reliability. ● Support application modernization and make full use of cloud service advantages. ● Ensure that applications meet business requirements and are agile and flexible.	● Design the cloud architecture of applications and guide the development team to implement the architecture. ● Evaluate and select cloud services to ensure that they meet application requirements. ● Solve technical challenges encountered during cloud transformation and provide professional support.
Data architects	● Design efficient data architectures to support data analysis and business decision-making. ● Keep your data secure and compliant. ● Implement data integration and sharing to improve the value of data.	● Plan the data storage and management in the cloud environment. ● Select suitable cloud databases and big data services. ● Implement data migration and governance, maintain data quality, and ensure data security.
Network architects	● Design flexible and reliable network architectures to support connections between application systems. ● Ensure network security and performance to meet data transmission requirements. ● Implement network elasticity and scalability to adapt to service changes.	● Plan cloud network architectures and configure virtual networks, subnets, and security groups. ● Work with the security team to implement network security policies. ● Monitor network performance and optimize network configurations.
Compliance audit experts	● Ensure that cloud transformation complies with relevant laws, regulations, and industry standards. ● Reduce compliance risks and avoid legal disputes and fines. ● Maintain the reputation of the enterprise and improve the trust of customers and partners.	● Identify compliance requirements in cloud transformation and provide professional suggestions. ● Participate in the formulation of compliance policies to ensure that cloud service providers meet the requirements. ● Regularly audit and evaluate compliance and propose improvement measures.

Stakeholder	Interests	Ways of Participation
IT governance experts	● Establish an effective IT governance framework to standardize the use and management of IT resources. ● Ensure the alignment between IT strategies and enterprise strategies to improve IT value. ● Control IT risks, improve decision-making transparency, and clarify responsibilities.	● Develop governance strategies and policies for cloud transformation and clarify responsibilities and processes. ● Monitor the progress and risks of cloud transformation and provide governance reports. ● Coordinate communication between departments to ensure information sharing and collaboration.
Product managers	● Accelerate product development and rollout to meet market demands. ● Introduce new technologies to improve product competitiveness. ● Collect customer feedback and continuously improve products.	● Formulate product requirements and cooperate with development and operations teams. ● Use cloud services to quickly verify and iterate products. ● Analyze product data to guide product optimization.

By identifying and analyzing the interests and demands of these stakeholders, you can better formulate and implement the cloud transformation strategy, ensuring that the interests of all parties are balanced and met.

2.3 Identifying Cloud Transformation Drivers

2.3.1 Overview of Cloud Transformation Drivers

Identifying the drivers of cloud transformation is the prerequisite for formulating strategies. You need to analyze internal and external factors.

- Internal factors include business growth requirements, cost optimization requirements, operational efficiency improvement, and business innovation requirements.
- External factors include market competition pressure, customer requirement changes, technology development trends, and regulatory compliance requirements.

Taking both internal and external factors into account, cloud transformation drivers can be categorized into business drivers, technical drivers, and financial drivers. These categories align with the perspectives of the CEO, CIO, and CFO, respectively.

2.3.2 Business Drivers

Business drivers are the fundamental reasons prompting CEOs and business directors to adopt cloud computing. Their focus lies in harnessing cloud computing to enhance service agility, accelerate service innovation, improve service continuity, expand market reach, ensure regulatory compliance, foster sustainability, and ultimately improve the enterprise's core competitiveness and business outcomes.

- **Enhance service agility**

  Service agility refers to the ability to quickly respond to market changes and customer needs. In a rapidly changing market environment, enterprises need to have agile service capabilities to maintain a competitive edge.

  - **Fast deployment of service systems**: Cloud platforms provide a highly flexible and scalable infrastructure. Enterprises can quickly deploy systems and services to shorten the time to market.
  - **Auto scaling**: Cloud computing's elasticity allows enterprises to dynamically adjust resources to meet service requirements, especially unexpected spikes, during high-demand periods.
  - **Agile development and iteration**: Cloud platforms support DevOps practices, accelerating the software development cycle and allowing quick iteration and update of service systems.

- **Accelerate service innovation**

  Service innovation is the key for enterprises to obtain new growth and maintain competitiveness. Cloud computing provides enterprises with innovative platforms, technologies, and tools, greatly lowering the innovation threshold and accelerating the innovation of products, services, business models, service processes, and operational modes.

  - **Easier access to advanced technologies**: Cloud service providers provide advanced technologies such as AI, large models, big data, IoT, and virtual humans. Enterprises can quickly use these technologies for innovation without building them by themselves.
  - **Lower technical thresholds**: Cloud services simplify the application of complex technologies. Enterprises can focus on service innovation without worrying about the complexity of underlying technologies.
  - **Global cooperation**: Cloud service providers have built a global ecosystem partner network, enabling enterprises to innovate with global partners and developers and expand global businesses.

- **Improve service continuity**

  Service continuity refers to the ability to continuously and stably provide products and services in the event of various faults, external attacks, or emergencies. The high availability and security of cloud platforms and cloud services can ensure the stable running of service systems and reduce operational risks.

  - **High-availability architecture**: Cloud service providers provide multi-region and multi-AZ deployment modes, which support cross-region DR and backup to improve the reliability of service systems.
  - **Automatic failover**: Cloud platforms have an automatic detection and failover mechanism. When a hardware or software fault occurs, they can quickly restore services with the minimal downtime.

  - **Security protection capability**: Cloud service providers possess extensive expertise in security protection. They offer end-to-end security technology systems, comprehensive management processes, and well-defined specifications, supported by a robust team of security experts. This ensures the continuous protection of cloud platforms. As a result, public clouds provide stronger information security assurance compared to most in-house IT teams within organizations.

- **Expand market reach**

  Expanding businesses to global markets is an important way for enterprises to increase revenue. With the global layout of cloud service providers, enterprises can effectively enter new markets, expand their business scope, and reach more customers.

  - **Global deployment**: Cloud service providers have cloud data centers around the world. Enterprises can quickly deploy services in target markets. The technological barriers and the time required to enter these new markets have been dramatically lowered.
  - **Local services**: Cloud platforms offer local language support and services that comply with local regulations, enabling enterprises to swiftly adapt to the demands of local markets.
  - **Lower entry costs**: Enterprises do not need to build data centers or purchase hardware devices locally, alleviating initial investment pressures and lowering market entry costs.
  - **Better customer experience**: Low-latency and high-performance services are provided through nearby deployment and optimized network architecture, improving customer satisfaction.

- **Ensure regulatory compliance**

  In today's rapidly changing business environment, compliance has become a key factor for enterprises to survive and develop. As regulations and standards around the world become increasingly strict and complex, enterprises need to ensure that their operations comply with local laws, regulations, and industry standards to avoid legal risks, financial losses, and reputation damage. With the compliance support services provided by cloud service providers, enterprises can effectively reduce compliance risks and focus on core business development.

  - **Global compliance support**: Cloud service providers' data centers and services around the world comply with local regulations and industry standards, such as Compliance General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), and Payment Card Industry Data Security Standard (PCI-DSS). Enterprises can use the compliance certification and services of the cloud platform to quickly meet compliance requirements in different markets.
  - **Local compliance services**: Cloud platforms provide local compliance services, such as data residency, data sovereignty, and data encryption, to help enterprises meet compliance requirements in specific regions. For example, some countries require that data be stored within the country. Cloud service providers can provide local data centers and services that meet the requirements.
  - **Higher compliance efficiency**: Cloud platforms provide automated compliance tools, such as security audit, vulnerability scanning, and

access control, to help enterprises manage compliance automatically. This helps improve the efficiency and reduce the risk of human errors.

- **Foster sustainability**

  Sustainability is an important aspect of enterprises' social responsibility and long-term development. Cloud computing helps reduce energy consumption and carbon emissions, enabling enterprises to meet their social responsibility commitments and promote sustainable, green development.

  - **Lower energy consumption and carbon emissions**: Cloud data centers often leverage advanced energy management and cooling technologies, achieving a Power Usage Effectiveness (PUE) of less than 1.2. This results in significantly lower energy consumption and carbon emissions compared to self-built data centers.
  - **Optimized resource utilization**: Cloud computing improves the utilization of servers and storage devices through virtualization and resource pooling, reducing the demand for physical devices.
  - **Environmental-friendly innovation**: The digital technologies of cloud platforms facilitate the creation of green solutions, such as smart cities, intelligent transportation systems, and online collaboration tools. These innovations contribute to environmental protection and energy conservation.
  - **Environmental compliance**: Cloud service providers comply with strict environmental standards and regulations. Enterprises can use cloud services to meet environmental compliance requirements.

The business drivers of cloud transformation cover the requirements of key business areas such as agility, innovation, continuity, market expansion, and sustainable development. By understanding the business drivers of cloud transformation, enterprises can:

- Formulate clear cloud transformation strategic objectives to ensure that the cloud transformation strategy is closely aligned with the business strategy.
- Gain the support of management and employees, align their perspectives, and collaboratively drive cloud transformation efforts.
- Maximize business value, improve market competitiveness, and promote sustainable development.

2.3.3 Technical Drivers

In the digital era, cloud computing has become the core of enterprise technology strategies. For CIOs, CTOs, and technical directors, cloud transformation is not only a business development requirement but also an inevitable choice for technological innovation and transformation. Cloud computing provides technical advantages in terms of resource elasticity, system resilience, scalability, security, and O&M efficiency. These technical drivers are the underlying technical support for business drivers and financial drivers.

- **Improve resource elasticity**

  Resource elasticity is one of the core features of cloud computing. It means that the cloud platform can quickly expand and reduce the compute, storage, and network resources required by the service system based on service requirements. Improving resource elasticity can effectively improve service agility and continuity.

  - **Dynamic resource allocation**: Cloud computing supports on-demand resource allocation. Enterprises can quickly expand the resource scale during service peak hours to cope with traffic surges. During service off-peak hours, enterprises can release idle resources to reduce costs.
  - **Automatic scaling**: Cloud platforms can automatically scale resources based on preset policies and real-time load conditions through automatic monitoring and scheduling.
  - **Quick deployment and reclamation**: Compared with the traditional hardware procurement and deployment, the creation and destruction of cloud resources can be completed within minutes or even seconds. This greatly improves the elasticity speed of resources.

- **Improve system resilience**

  System resilience refers to the ability of a system to maintain normal operation or quickly recover when facing various external disasters and internal software and hardware faults. Cloud platforms and cloud services can greatly improve the resilience of application systems, thereby effectively improving service continuity.

  - **High-availability architecture**: Cloud service providers provide multi-region and multi-AZ deployment modes and allow application systems to adopt cross-data center, cross-region disaster recovery (DR) and active-active solutions, or even cross-region, multi-active solution. These solutions can greatly improve the availability and DR capabilities of application systems.
  - **Backup and DR**: Cloud platforms provide built-in functions such as data backup and DR switchover to help enterprises build a comprehensive disaster recovery solution and ensure system availability in emergencies.
  - **Service level agreement (SLA)**: Cloud services usually provide SLAs of more than 99.9%. This effectively ensures the reliability SLO of application systems built based on these cloud services.
  - **Automatic fault handling**: Cloud platforms can automatically detect and rectify faults, reducing manual intervention and shortening the fault handling time.

- **Improve scalability**

  Scalability refers to the ability of a system to maintain stable performance and efficiency by adding or adjusting resources (such as servers, storage, and bandwidth) without changing the system architecture or making minimum modifications to the system when the system is faced with increasing workloads or requests. Cloud platforms and cloud services can greatly improve the scalability of application systems and smoothly handle the increasing number of users, data, or transactions without causing performance degradation or system breakdown. Improving system scalability can effectively improve service agility and continuity.

  - **Distributed architecture**: Cloud platforms support distributed system architecture design, allowing applications to run on multiple servers or nodes and distributing loads to avoid single points of failure. This improves system scalability and availability.
  - **Auto scaling**: Cloud platforms provide the auto scaling function. Systems can automatically add resource instances based on preset policies to cope with traffic peaks and ensure stable performance.

  - **Microservice architecture**: Cloud platforms are suitable for microservices and containerized deployment. They support application splitting and separate expansion to improve flexibility and maintainability.
  - **Serverless computing**: Cloud platforms provide serverless compute services (such as **FunctionGraph**) that allow users to deploy code on the cloud without managing servers. Cloud platforms automatically allocate compute resources based on requests and release resources after requests are processed. This greatly simplifies scalability management.

- **Improve security**

  Security refers to the ability to protect data and application systems from unauthorized access, use, disclosure, tampering, destruction, or loss. Cloud service providers invest heavily in information security to provide enterprises with multi-layer security protection. Improving security can effectively improve service continuity.

  - **Cloud platform security**: Cloud platforms comply with strict security standards and certifications, such as ISO 27001, CSA, SOC 1, SOC 2, SOC 3, DJCP, PCI-DSS, and NIST CSF.
  - **Abundant cloud native security services**: Cloud service providers provide many kinds of cloud native security services, covering the host, data, application, network, identity, and O&M security, to help enterprises quickly build a comprehensive security defense line for application systems on the cloud.

- **Improve O&M efficiency**

  O&M efficiency refers to the ability of the IT O&M team to manage as many IT resources as possible with the minimum resource investment (labor, time, and cost) and maintain high service quality and stability. It reflects resource utilization and personnel productivity. After adopting cloud computing, enterprises can greatly improve O&M efficiency and effectively reduce O&M costs.

  - **No need to manage IT infrastructures**: Cloud service providers are responsible for the O&M of IT infrastructures, such as cloud data centers, hardware, networks, and virtualization. Enterprises only need to focus on the O&M of application systems.
  - **Intelligent monitoring systems**: Cloud service providers provide full-stack, intelligent monitoring systems to collect and analyze cloud resource and application performance metrics in real time, automatically identify exceptions, predict potential risks, and provide alarms and visualized reports, helping O&M personnel quickly locate faults.
  - **Automated O&M**: Cloud service providers provide tools for automated deployment, configuration management, monitoring, alarm reporting, and O&M, reducing the workload of O&M personnel and improving O&M efficiency. Automated O&M also reduces the risk of human errors, thereby reducing unnecessary troubleshooting.
  - **Serverless architecture**: If enterprises use serverless services such as function computing, they only need to write service logic code and do not need to manage any server, which further reduces the O&M burden.

- **Improve performance efficiency**

  The goal of improving performance efficiency is to process more service requests with fewer IT resources, which is ultimately reflected in higher throughput, response time, or number of concurrent users. With excellent architecture design principles and performance detection and optimization tools provided by cloud service providers, enterprises can effectively improve their system performance metrics.

  - **Selecting appropriate resources**: Select the most appropriate specifications for compute, storage, middleware, and database resources based on service requirements, and avoid over-provisioning to reduce resource waste.
  - **Performance testing and planning**: Use performance test tools provided by cloud platforms to evaluate the current performance metrics of application systems, and plan the capacity based on the service growth trend in advance.
  - **Performance tuning**: Explore the performance potential of existing resources, including database query optimization, code optimization, and using cache and CDN acceleration, to improve system throughput and response speed.
  - **Architecture optimization**: Use more efficient architectures. For example, use asynchronous processing and message queues to decouple system components and improve the concurrency processing capability.

Technical drivers of cloud transformation bring profound changes to the IT strategy and technical architecture of enterprises. For technical leaders, a deep understanding and good use of these technical drivers will help:

- Formulate forward-looking technical strategies to lead the digital development of enterprises.
- Optimize the IT architecture and resource configuration to improve the contributions made by technical departments.
- Promote technological innovation and service convergence to support enterprises in gaining competitive advantages.

2.3.4 Financial Drivers

Cloud transformation brings many financial benefits, including pay-per-use billing, cost reduction, and increased revenues, all of which are attractive to CFOs and financial directors.

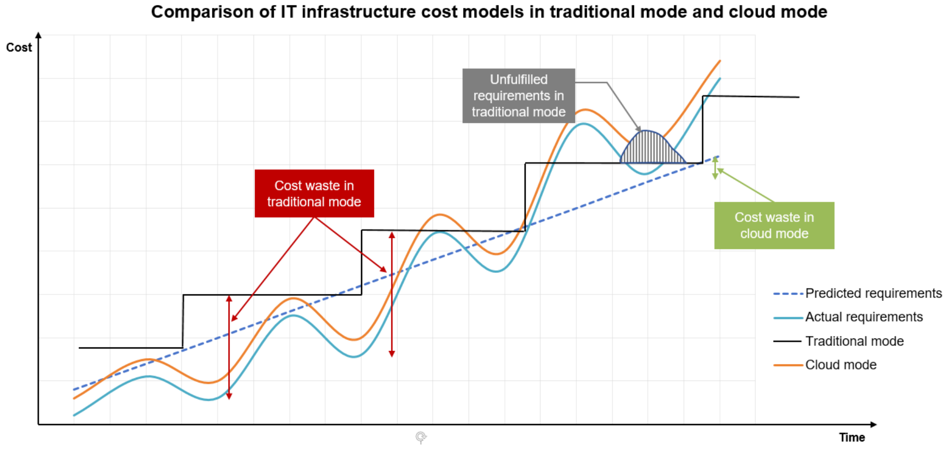
- **Cost-effectiveness**

  In the traditional mode (on-premises data center), enterprises purchase IT resources such as hardware and software in advance based on the predicted service peaks. To avoid performance bottlenecks or service interruption during peak hours, enterprises usually over-purchase resources. Actual service loads fluctuate and run at the average level or below in most of the times. As a result, IT resources are idle or underutilized for a long time, causing huge waste of costs.

  In the cloud mode, enterprises flexibly purchase cloud resources required based on actual service loads and only need to pay for the resources used. They do not need to purchase a great number of hardware and software resources in advance. During service peak hours, cloud platforms can quickly add cloud resources to meet requirements. During service off-peak hours, they can release redundant cloud resources to reduce resource waste and costs.

The following figure shows the comparison between the cost models of the traditional mode and cloud mode.

Figure 2-1 Comparison of IT infrastructure cost models in traditional mode and cloud mode



Why the cloud mode saves IT infrastructure costs can be considered from the following:

- **Pay-per-use billing**: In the cloud mode, the cloud resource quantity can be automatically adjusted based on service peak and off-peak hours. Enterprises only need to pay for the cloud resources that are actually used. In the traditional mode, enterprises need to purchase and maintain a large number of hardware and software resources in advance. The pay-per-use mode avoids resource idleness and waste.
- **Economies of scale**: Cloud computing is essentially an economy of scale. Cloud service providers build and operate ultra-large data centers, which have ultra-large compute and storage capabilities. Due to the large scale, cloud service providers can significantly reduce operating costs by purchasing hardware and software in batches, optimizing resource utilization, improving energy efficiency, and implementing automatic management. This way, they can provide IT resources at lower costs.
- **Reduce O&M costs**: Cloud service providers are responsible for the maintenance and management of IT infrastructures. Enterprises do not need to invest a lot of labor and money in the routine O&M of IT infrastructures. In addition, cloud platforms provide intelligent monitoring systems and automatic O&M systems, which greatly improve the O&M efficiency of application systems. Enterprises can reduce the labor investment in application system O&M, further reducing the O&M labor cost. Automatic O&M also reduces the risk of human errors, thereby reducing the cost of error correction.

Note that cost waste may also occur in the cloud mode. The part between the orange curve and the blue curve in the above figure shows the cost waste in the cloud mode. Many reasons lead to the waste. For example, cloud resources applied for during service peak hours are not released in a timely manner. In the subsequent O&M governance section, we will describe in detail how to optimize costs and perform continuously cost operations.

- **Increase revenues**

Cloud computing not only reduces costs but also helps enterprises create new revenue sources:

- **Shorten time to market**: Cloud platforms support quick deployment and expansion of applications to help enterprises launch new products and services faster, seize market opportunities, and gain more market share and revenue.
- **Expand new markets**: The global coverage capability of cloud platforms makes it easier for enterprises to enter new markets, expand business scope, and reach a wider customer base, thereby increasing revenues.
- **Improve customer experience**: Cloud platforms can provide more stable, reliable, and high-performance services. These improve customer satisfaction and loyalty, thereby increasing revenues and the customer retention rate.
- **Business innovation**: The flexibility and integrated new technologies of cloud platforms support enterprises in innovating products and services, exploring new operating and business models, and exploring new market opportunities. All these create new revenue sources.

- **Capital expenditure to operating expenditure**

  In the traditional mode (self-built data center), IT infrastructure construction requires huge capital expenditure (Capex). Enterprises need to purchase hardware such as servers, storage devices, and network devices, and bear maintenance and update costs. This requires enterprises to invest a large amount of money at a time. Cloud services use pay-per-use billing to convert capital expenditure to operating expenditure (Opex). Enterprises only need to pay for the compute resources, storage space, and network bandwidth actually used, just like how they pay for water and electricity. This significantly lowers the initial investment threshold of enterprises, improves the capital utilization efficiency, and flexibly adjusts resource usages based on service requirements to avoid resource idleness and waste. This is crucial for budget planning and cash flow management, and also allows enterprises to respond more flexibly to market changes.

In a word, the financial advantages brought by cloud transformation are the core reasons why CFOs and financial directors embrace cloud computing. By reducing costs, increasing revenues, and converting capital expenditure to operating expenditure, cloud computing can help enterprises improve financial performance.

2.3.5 Multi-Cloud Strategy Drivers

The multi-cloud strategy is becoming a mainstream trend. More and more organizations are deploying service systems on cloud platforms of multiple cloud service providers instead of relying on a single cloud service provider. This trend is driven by multiple factors. The following are some of the main drivers:

- **Avoid single-cloud failures**: Deploying services on a single cloud platform poses risks of SPOFs. If the cloud platform fails, for example, due to a large-scale breakdown or regional disaster, enterprise services will be severely affected. By adopting the multi-cloud strategy, you deploy service systems on multiple independent cloud platforms for cross-cloud DR. Even if one cloud platform fails, services on other cloud platforms can still run properly to ensure service continuity.

- **Avoid vendor lock-in**: Deploying all services on the cloud platform of a single cloud service provider results in vendor lock-in. It puts enterprises in a weak position when negotiating with the vendor and makes it difficult for enterprises to migrate to other cloud platforms. The multi-cloud strategy can avoid this situation and keep flexibility for enterprises in selecting cloud service providers.
- **Reduce costs and improve efficiency**: The multi-cloud strategy introduces competition. By cooperating with multiple cloud service providers, enterprises can select the most suitable cloud services based on their requirements and reduce costs through the competition among cloud service providers. In addition, different cloud service providers may adopt different pricing policies on services or in different regions. The multi-cloud strategy can help enterprises optimize resource allocation and improve cost-effectiveness.
- **Fully utilize the advantages of different vendors**: Different cloud service providers have their own advantages in terms of technology, services, and functions. For example, a cloud service provider may have stronger technical strength in artificial intelligence and machine learning, while another may have advantages in database services. The multi-cloud strategy allows enterprises to select the most suitable cloud service providers and their advantageous services based on their service requirements, thereby maximizing the value of cloud computing.
- **Regulatory compliance**: Some countries and regions have specific regulations on data storage and processing. The global layout and compliance level of each cloud service provider are different. The multi-cloud strategy can help enterprises select the most suitable cloud service providers to meet these regulatory requirements, for example, storing sensitive data on cloud platforms in specific regions.

Adopting the multi-cloud strategy improves service continuity, cost-effectiveness, and security. Though it also brings challenges like management complexity, these challenges are gradually being resolved with the continuous development of cloud management tools and technologies.

2.3.6 Identifying Drivers

Identifying drivers is the prerequisite for cloud transformation and determines whether an organization has a justified reason to start cloud transformation. Identifying drivers is a complex process. After comprehensively considering the business strategy, service requirements, financial requirements, and technical requirements, an agreement needs to be reached among executives and stakeholders. The recommended steps are as follows:

- **Respond to key business events**

  Enterprise executives usually make the cloud transformation decision based on actual service requirements. Key business events are often opportunities that drive cloud transformation. Key business events that enterprises may face now and in the future must be fully considered. The following lists some common key business events:

  - **Digital transformation**: Enterprises need more advanced IT technologies and platforms to support digital transformation. Cloud platforms can provide a variety of digital tools and services.

  - **Data center retirement**: Existing data centers are about to expire or devices are aged, calling for upgrade and replacement. So, cloud migration becomes an attractive option.
  - **Mergers and splits**: Mergers, acquisitions, or splits of enterprises have a significant impact on the IT infrastructure. The flexibility of cloud services can help enterprises quickly adjust IT resources to adapt to new organizational structures.
  - **Tight cash flow**: Enterprises have tight cash flow and hope to reduce capital expenditure, including investment in the IT infrastructure, and convert Capex to Opex. Pay-per-use billing that cloud services offer can meet this requirement.
  - **Termination of key technologies**: The existing provider is about to stop support on existing key technologies. Upgrade or migration is required. Cloud migration can provide more modernized, reliable, and secure technical solutions.
  - **Changes in regulatory compliance**: New regulations or compliance requirements may require enterprises to adjust IT systems. Cloud platforms can better meet these requirements.
  - **Key service system interruption**: Enterprises have experienced interruption of key service systems, causing losses and reputational damage. They hope to improve the reliability of service systems. Cloud platforms can provide higher reliability and DR capabilities.
  - **Failing to meet carbon emission goals**: Enterprises hope to reduce energy consumption and carbon emissions and enhance the brand awareness and social responsibility. Cloud data centers usually adopt advanced energy management and cooling technologies to improve energy efficiency.
  - **Fast market changes**: The market environment and customer requirements change rapidly. Enterprises need to speed up product launch. Cloud platforms provide more flexible and elastic IT infrastructure to support quick product and feature launch.
  - **Security attacks**: Enterprises have recently suffered from hacker attacks and hope to improve the service system and data security to defend against attacks and data breaches. Cloud platforms can provide more comprehensive and powerful security protection.
- **Map key business events to drivers**

  Associate the key business events identified with cloud transformation drivers to better understand how cloud transformation can cope with the challenges brought by these key business events.

Table 2-2 Mapping between key business events and drivers

Key Business Event	Business Drivers	Technical Drivers	Financial Drivers
Digital transformation	Enhance service agility Accelerate service innovation Improve service continuity Expand market reach	Improve resource elasticity Improve system resilience Improve scalability Improve security	Increase revenues
Data center retirement	-	Improve resource elasticity Improve system resilience Improve scalability Improve security Improve O&M efficiency Improve performance efficiency	Pay-per-use billing Cost-effectiveness
Mergers and splits	Enhance service agility	Improve resource elasticity Improve scalability	-
Tight cash flow	-	Improve resource elasticity Improve O&M efficiency Improve performance efficiency	Pay-per-use billing Cost-effectiveness
Termination of key technologies	Improve service continuity	Improve resource elasticity Improve system resilience Improve scalability Improve security	-

Key Business Event	Business Drivers	Technical Drivers	Financial Drivers
Changes in regulatory compliance	Ensure regulatory compliance	Improve security	-
Key service system interruption	Improve service continuity	Improve system resilience Improve performance efficiency	-
Failing to meet carbon emission goals	Foster sustainability	-	-
Fast market changes	Enhance service agility	Improve resource elasticity Improve scalability	Increase revenues
Security attacks	Improve service continuity	Improve security	-

- **Prioritize the drivers**

  Business events have different urgency and importance. You need to prioritize the identified drivers based on the business strategy and condition of the enterprise. For example, for an enterprise that is undergoing digital transformation, "enhance service agility" and "accelerating service innovation" have higher priorities. For an enterprise with tight cash flow, "pay-per-use billing" and "cost-effectiveness" have higher priorities. These priorities determine which aspects should be considered first during the future solution design. For example, if resilience, security, and cost conflict, enterprises with tight cash flow should prefer low-cost design solutions and make some compromises in terms of security and resilience.

- **Align with executives and stakeholders**

  After the drivers and priorities of cloud transformation are determined, record the drivers, priorities, and expected benefits of cloud transformation, communicate and align with executives and stakeholders, listen to their opinions and suggestions, and obtain their understanding and support.

2.4 Assessing Cloud Maturity

2.4.1 Cloud Maturity Model

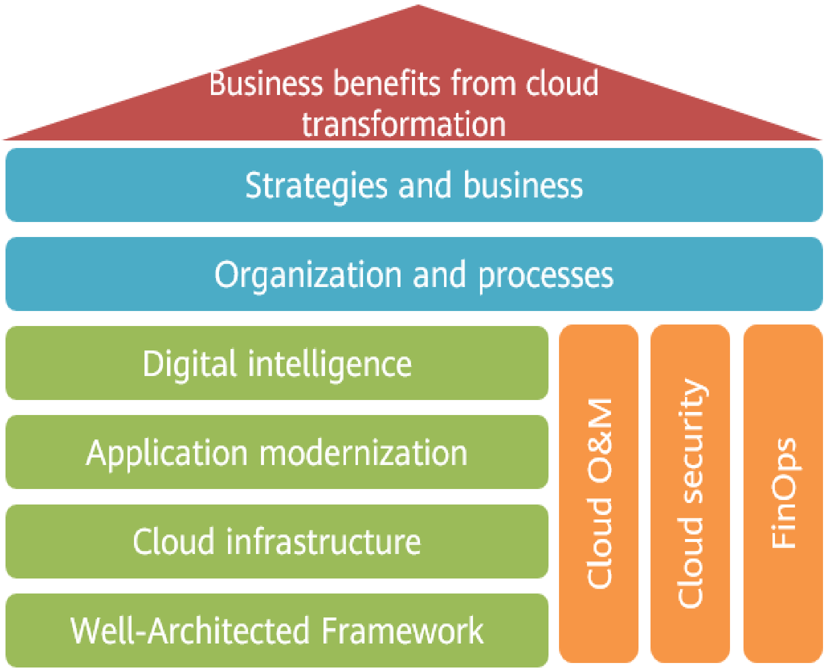
The purpose of assessing cloud maturity is to help your organization fully understand your capabilities during cloud transformation, identify gaps, and develop targeted improvement plans. This helps ensure that your organization's cloud transformation objectives are realistic and feasible, avoiding overly high or low expectations.

Huawei Cloud has designed the cloud maturity model based on the Cloud Maturity Model from the Open Alliance for Cloud Transformation (OACA), the Cloud Native Maturity Model from the Cloud Native Computing Foundation (CNCF), the Information Technology Service Standards (ITSS) of the China National Technical Committee 28 on Information Technology, and the Data Management Capability Maturity Model (DCMM), along with the best practices from a large number of government and enterprise customers of Huawei Cloud. This model is based on the following five principles:

- **Business-driven**: The assessment model must be centered on driving business growth. The cloud transformation objectives must align with your company's business strategies and objectives. Cloud transformation needs to go beyond technical benefits to bring about business benefits.
- **All-element**: The assessment model must go beyond technical capabilities to cover all elements, including people, technology, and processes.
- **Full-stack**: The assessment model must cover the full technology stack, including the Well-Architected Framework, cloud infrastructure, application modernization, big data and AI, O&M, and security protection.
- **E2E**: The assessment model must cover the whole journey of cloud transformation, including strategy formulation, top-level planning, solution design, adoption and implementation, and O&M.
- **Integrated**: The assessment model must consider integrated management capabilities between on-premises and cloud, between clouds, between regions in a cloud, and between accounts in a cloud.

Based on these principles, Huawei Cloud has designed the following 10 assessment dimensions:

Figure 2-2 10 assessment dimensions of the cloud maturity model


Business benefits from cloud transformation
Strategies and business
Organization and processes
Digital intelligence
Application modernization
Cloud infrastructure
Well-Architected Framework
Cloud O&M
Cloud security
FinOps

- **Business benefits from cloud transformation**

  This dimension assesses the business and financial benefits that an organization can achieve through cloud transformation, including improved business agility, improved business continuity, reduced TCO, accelerated business innovation, and improved sustainability. This dimension is placed first because business benefits are the most important dimension. If business benefits are not achieved, it is futile to do well in other dimensions.

- **Strategies and business**

  This dimension assesses the strategic planning capability of an organization during cloud transformation, including whether the cloud strategy is consistent with the overall business strategy and objectives, whether the key business drivers are identified, and whether clear cloud strategies, cloud transformation objectives, and migration policies are formulated. In addition, this dimension assesses the foresight, comprehensiveness, and feasibility of strategy formulation, as well as the grasp of industry trends and cloud technology trends.

- **Organization and processes**

  This dimension assesses the adaptability and transformation capability of an organization in terms of organizational structure, personnel skills, and work processes during cloud transformation. It measures whether the organization has an organizational structure and talent team that can support cloud transformation and whether work processes suitable for the cloud environment have been established.

- **Digital intelligence**

  This dimension assesses the organization's capabilities in big data and AI, and whether it can use cloud platform data intelligence services to achieve data-driven business innovation and intelligent transformation. It also measures the organization's data lifecycle management, data governance capabilities, and practice level in AI (such as AI development and large models).

- **Application modernization**

  This dimension assesses whether the organization's application systems use modern design and development models, such as microservice architecture, event-driven architecture, containerization, serverless architecture, and DevSecOps practices, as well as whether it has the capability to develop and deploy cloud native applications.

- **Cloud infrastructure**

  This dimension assesses the organization's design, deployment, and management capabilities of cloud infrastructure, including landing zone design and implementation, integrated management of networks and IAM, infrastructure automation deployment based on IaC, data backup, and auto scaling policies.

- **Well-Architected Framework**

  This dimension assesses whether the organization follows the design principles and best practices of the Well-Architected Framework, which covers five aspects: resilience, security, performance, cost optimization, and operational excellence.

- **Cloud O&M**

  This dimension assesses the organization's O&M capabilities in the cloud environment, including observability, CMDB, automated O&M, chaos engineering, ITSM, and AIOps. It also assesses whether the organization has established the most suitable cloud operations model and O&M processes for the current business situation that support agile delivery and stable operations of cloud-based business systems.

- **Cloud security**

  This dimension assesses the organization's security measures and operations capabilities in the cloud environment, including identity security, network security, data security, host security, application security, O&M security, security management regulations, and integrated security operations.

- **FinOps**

  This dimension assesses the organization's capabilities in managing and optimizing cloud resource costs, including cost budgeting, visualization, optimization, operations, and integrated financial management.

These assessment dimensions have taken into account the elements of people, technology, and process, spanning the entire journey and technology stack of cloud transformation, with integrated management involved. There are more than 70 assessment questions in total. Each of these questions involves five levels: initiating, emerging, performing, advancing, and leading.

Table 2-3 Five levels of cloud maturity

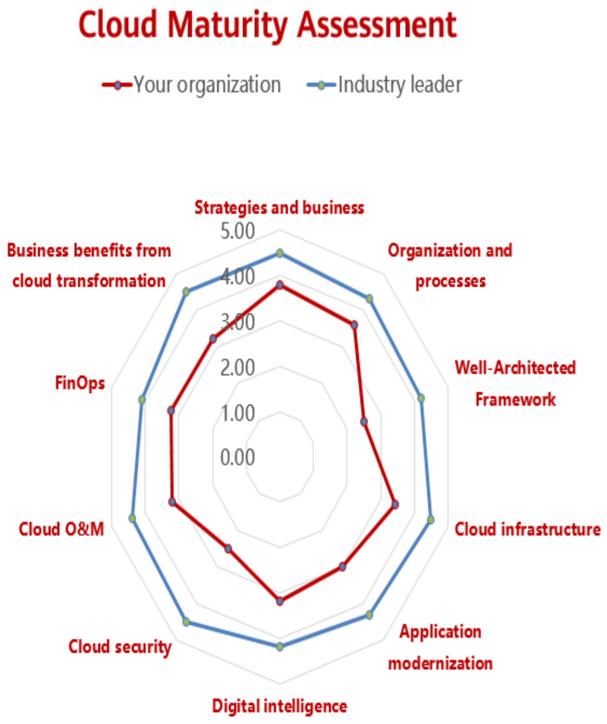
Level	Score	Maturity Level

Initiating	1 point	● The understanding and application of cloud computing are in the initial stage. The application of cloud-native technologies and best practices is limited, and there are security and cost risks. ● The cloud transformation has not started yet. There is no overall planning and strategy. There is no organization or processes to support cloud transformation.
Emerging	2 points	● Cloud computing technologies are partially applied, and some preliminary results have been achieved. However, the overall system and integrity are still inadequate. ● Cloud-native technologies and best practices are gaining attention, but they are not yet seeing sufficiently widespread adoption. The automation level is low, and security and cost management need to be further strengthened. ● A transition from traditional IT to cloud-native IT is underway.
Performing	3 points	● The cloud transformation has achieved results. Technical competitiveness has been established by applying cloud computing technologies, but business advantages have not been established and business benefits are not obvious. ● Various cloud services can be used skillfully, and certain automatic management capabilities are available. ● Cloud-native technologies and best practices, such as DevOps and microservices, are systematically applied. Security and cost management have also been given appropriate attention and implemented. ● A complete process and organization are available, and the cloud transformation objectives have been established within the IT department.
Advancing	4 points	● The cloud transformation has achieved remarkable results. Business competitive advantages have been established by applying cloud computing technologies, and remarkable business benefits, such as improved business agility and continuity and increased revenue, have been achieved. ● Cloud computing has become a key factor in driving business innovation and improving competitiveness. ● Cloud-native technologies and best practices can be fully used to achieve high agility and scalability. The security and cost management systems are mature and efficient. ● Processes are effectively executed, and cloud transformation objectives are established across the organization.

Leading	5 points	● In the cloud computing field, the organization is in the leading position and can lead the innovation of technologies and business models. ● The advantages of cloud-native technologies are fully exerted to achieve highly automated and intelligent operations. ● Security and cost management reach the industry-leading level. ● Cloud transformation not only promotes the rapid development of corporate business, but also sets a benchmark for the industry. ● Close attention is paid to the changes in business requirements and cloud computing technologies to continuously iterate and optimize the solutions.

The detailed assessment and analysis from these 10 dimensions can help your organization fully assess your capabilities during the cloud transformation process. By drawing a radar chart to quickly find the gap between your organization and the industry-leading enterprises, you can efficiently formulate targeted plans for improvement.

Figure 2-3 Cloud maturity radar chart



Note that the cloud maturity model is for relatively general and coarse-grained assessment. The main purpose is to help your organization quickly identify your capability gaps and determine cloud transformation objectives. It cannot replace a detailed assessment in the survey and assessment phase, which aims to help your organization design a detailed technical implementation solution.

2.4.2 Assessment Procedure

With a relatively complete cloud maturity model in place, it takes about one week to complete an assessment. The procedure for cloud maturity assessment is as follows:

Step 1 Define the scope of assessment.

This is the basis of the entire assessment process. In this phase, you need to determine the specific scope to be assessed based on the current situation and business requirements of your organization. Cloud transformation involves multiple assessment dimensions and many assessment questions. You may not be able to cover all of them in one assessment. You can focus on the aspects most relevant to the current development stage and business objectives of the organization and select some key dimensions for assessment. By communicating with relevant business departments and technical teams, you need to identify the areas that need to be improved the most and ensure that the assessment focuses on the aspects most valuable to your organization. The goal of this step is to develop a clear and executable assessment scope, laying a solid foundation for subsequent assessment work.

Step 2 Identify and coordinate assessors.

This step is critical to the accuracy and effectiveness of the assessment. You need to determine the most suitable personnel within your organization to answer the selected assessment questions. These personnel should have solid knowledge of the areas they are responsible for and be able to provide objective and detailed answers. The following table describes the recommended assessors from the 10 dimensions of cloud maturity assessment:

Table 2-4 Recommended assessors

Dimension	Recommended Assessor
Strategies and business	CEO or CIO
Organization and processes	CIO or HR director
Well-Architected Framework	CIO or enterprise application architecture leader
Cloud infrastructure	CIO or IT director
Application modernization	Application architect
Digital intelligence	Business director or data architect

Cloud security	CISO
Cloud O&M	O&M director
FinOps	Cloud cost management leader
Cloud transformation benefits	Business director or IT director

You need to coordinate the time of assessors and try to concentrate assessment activities within a period of time to improve efficiency. In addition, to ensure that assessors can fully understand the purpose and requirements of the assessment, you can provide instructions and training before the assessment to explain the purpose, process, and requirements of the assessment in detail.

Step 3 Perform the assessment.

This step is the core of the entire process. Assessors need to answer questions carefully and objectively based on the assessment scope and questions determined earlier. In answering each question, the actual situation of the organization should be fully considered, and specific data and examples should be provided for support. If assessors have any questions about some assessment questions, they should contact the assessment expert in a timely manner and attend Q&A meetings to ensure that the questions are accurately understood. When performing the assessment, avoid subjective assumptions. Assessment must be based on facts. In addition, the objectivity and confidentiality of the assessment need to be emphasized, and assessors are encouraged to report the truth and assured that the reported content will not cause any negative impacts. The quality of this step directly affects the reliability of the assessment results and the effectiveness of subsequent improvement measures. Therefore, assessors need to pay enough attention.

Step 4 Produce an assessment report.

The assessment expert needs to summarize and analyze all the answers provided by the assessors to identify the organization's strengths and gaps in cloud maturity. For the identified capability gaps, the assessment expert should conduct an in-depth analysis of their causes and impacts and propose targeted improvement suggestions. These suggestions should be specific and actionable, including the priority of improvement measures, implementation methods, and expected results. The assessment report should be logically clear and well-organized, covering the assessment results comprehensively and providing strong support for the organization to develop the subsequent action plan. Through the assessment report, the organization's management and departments can clearly understand their current capabilities and the areas that need improvement, thereby formulating and adjusting cloud strategies, cloud objectives, and implementation plans in a targeted manner. A quality assessment report is an important decision-making basis for an organization's cloud transformation.

----End

2.5 Setting Cloud Transformation Objectives

Cloud transformation objectives must align with the business strategy and objectives of your organization. The objectives must comply with the SMART principle, that is, they must be specific, measurable, achievable, relevant, and time-bound. For example, if your organization experiences revenue loss and brand damage due to service system interruptions, you can set a specific objective such as improving the availability of service systems from 99% to 99.9% within the next year. When setting cloud transformation objectives, you need to fully consider the current resources and capabilities of your organization to ensure that the objectives are achievable. Through the cloud maturity assessment, you can learn the current capabilities and gaps of your organization and set cloud transformation objectives accordingly to avoid excessively high or low objectives.

To set measurable cloud transformation objectives, you need to design reasonable quantitative metrics based on the cloud transformation drivers, so that the management of your organization can track and evaluate the actual effects of cloud transformation. The following table lists the quantitative metrics designed based on the cloud transformation drivers:

Table 2-5 Metrics for cloud transformation drivers

Category	Driver	Metric
Business drivers	Enhancing service agility	• TTM of new products (new service systems), including the design, development, testing, and market launch • Iteration time of new versions or features of existing products (existing service systems), including the design, development, testing, and market launch
	Accelerating service innovation	• Number of new users brought by new products, new services, and new business models • Revenue brought by new products, new services, and new business models
	Ensuring service continuity	• Availability SLO of service systems • Economic loss caused by service interruptions
	Expanding market reach	• Number of new users brought by entering new markets • Revenue brought by entering new markets
	Ensuring regulatory compliance	• Economic loss caused by non-compliance, including compensation and fines
	Fostering sustainability	• Reduction in carbon emissions

Technical drivers	Improving resource elasticity	● Resource utilization ● Resource expansion and deployment speed
	Improving system resilience	● RPO and RTO ● Number of P1, P2, P3, and P4 incidents
	Improving scalability	● System expansion speed
	Improving security	● Number of security incidents ● Economic loss caused by security incidents, including compensation and fines
	Improving O&M efficiency	● O&M time required by a unit of resources ● Number of resources (such as VMs and storage) that can be maintained by each O&M engineer ● Mean time to repair (MTTR)
	Improving performance	● Throughput metrics such as TPS and QPS ● System response time ● Concurrent users ● Resource utilization
Financial drivers	Pay-per-use	● Capital expenditure on IT infrastructure
	Improving cost-effectiveness	● Business unit cost, such as the cost of each order and each user ● TCO of IT infrastructure
	Increasing revenue	● Revenue brought by business innovation and market expansion

As it is time-consuming to track, evaluate, and manage all of the above metrics, you are advised to select some of them to evaluate the effects of cloud transformation. Some metrics overlap with each other. For example, the availability SLO of the service system encompasses metrics such as RPO, RTO, and MTTR. The ultimate goal of cloud transformation is to maximize business benefits, not technical benefits. Therefore, you are advised to categorize and sort these metrics based on the business objectives and priorities of your organization. Choose metrics first based on your business and financial drivers, and then based on your technical drivers. The following metrics are recommended for determining cloud transformation objectives:

Table 2-6 Cloud transformation metrics and objectives

Metric	Objective
Availability SLO of the service system	Increase the annual availability SLO of the service system from 99% to 99.95%.
TTM of new products or versions	Shorten the TTM of new products from six months to three months, and shorten the iteration time of existing products from three months to one month.
Economic loss caused by non-compliance and security incidents	Reduce the number of non-compliance and security incidents by 75% and reduce the economic loss by 75%.
TCO of IT infrastructure	Reduce the TCO of IT infrastructure by 15%.
Revenue brought by business innovation and market expansion	Innovate cloud business and expand the global market, bringing *X*0 thousand active users and revenue of USD*Y*00 million within the next three years.
Carbon emission reduction	Reduce carbon emissions by 50% following cloud transformation.
Number of resources that can be maintained by each O&M engineer	Double the number of servers that can be managed by each O&M engineer from 100 to 200.

2.6 Analyzing the Benefits of Cloud Transformation

You need to further analyze the financial revenues of cloud transformation based on the previously determined objectives and calculate the ROI of the project. This can help managers make strategic decisions. The following evaluates the financial revenues based on the seven recommended objectives. You can put them together to obtain the total revenue of the entire cloud transformation project.

- **Improved service system SLO**

  An improved SLO can reduce system downtime, thereby reducing revenue losses. Therefore, you can calculate the financial revenue of this metric based on the system's downtime loss per hour. Assume that a service system suffers a loss of USD0.1 million per hour due to downtime. Since the SLO before and after cloud transformation is 99% and 99.95%, the annual downtime loss is USD8.76 million (USD0.1 million per hour × 87.6 hours) and USD0.438 million (USD0.1 million per hour × 4.38 hours), respectively. Then, the annual financial revenue is USD8.322 million.

- **Short time to market (TTM) of new products or versions**

  After cloud transformation, enterprises can use DevOps tool chains, application modernization, and elastic compute resources provided by the cloud platform to greatly shorten the TTM of new products and accelerate iteration of existing products. If the TTM of a new product is shortened from six months to three months, and the new product generates USD2 million in revenue per month, launching the product three months earlier will bring an additional revenue of USD6 million. For an existing product, if the version iteration cycle is shortened from three months to one month, and a new version can bring USD1 million in revenue per month, launching the new version two months ahead of schedule will bring an additional revenue of USD2 million.

- **Low economic losses caused by non-compliance and security incidents**

  After cloud transformation, enterprises can leverage the all-round security protection measures, compliance audit tools, and automated compliance reports provided by cloud service providers to significantly enhance security and compliance. If there are five security and non-compliance incidents, rather than the previous 20, each year, and each incident causes an average loss of USD0.2 million, the total economic loss can be reduced by USD3 million per year.

- **Reduced TCO of IT infrastructure**

  In traditional self-built data centers, enterprises provision IT resources such as hardware and software based on forecasted peak service demand. However, the actual workload typically stays at an average level over the long term, which results in low resource utilization and a huge waste of costs. In the cloud, enterprises provision resources on demand and pay only for what they use, which significantly reduces IT infrastructure costs. There are many different cost items involved in calculating the TCO for traditional data centers and the cloud. In traditional data centers, you must account for the cost of building and maintaining data centers, whereas in the cloud, these costs are absent because they are already included in the price of cloud services. The following table lists the costs for traditional data centers, which can be divided into capital expenditure (CapEx) and operating expense (OpEx).

Table 2-7 Costs for traditional data centers

Category	Item	Description
CapEx	Hardware cost	Costs for purchasing servers, storage devices, and network devices (such as routers, switches, and firewalls)
	Software cost	License costs for software such as operating systems, virtualization software, databases, and middleware
	Data center cost	Costs for building self-built data centers and for installing the security system (including video surveillance devices and access control devices), power supply system, and cooling system in data centers
	Implementation cost	One-off costs for system integration, testing, deployment, and others

Category	Item	Description
OpEx	Labor cost	Labor costs for data center and IT maintenance and security, including expenses for salaries, benefits, and training
	Hardware maintenance cost	Costs for maintaining, repairing, and replacing hardware devices
	Software maintenance cost	Costs for software updates, patches, technical support, and others
	Data center maintenance cost	Costs for maintaining systems (including security system, power supply system, and cooling system) that ensure the normal running of data centers
	Energy cost	Costs for energy and electricity that are used to run the entire data center
	Rack leasing cost	Costs for leasing IDC racks
	Bandwidth cost	Internet access bandwidth costs

After cloud transformation, enterprises mainly incur OpEx, as listed in the table below.

Table 2-8 Costs for the cloud

Type	Item	Description
OpEx	Compute resource cost	Costs for VMs, containers, serverless computing, and others. Generally, these services are billed by use time, CPU, and memory.
	Storage resource cost	Costs for services such as object storage, block storage, and file storage. Generally, these services are billed by storage space, number of requests, and data transmitted.
	Network resource cost	Costs for network services such as Internet bandwidth, public IP addresses, NAT gateways, load balancers, and VPNs
	Database cost	Costs for services such as relational databases and NoSQL databases. Generally, these services are billed by instance specifications, storage space, and number of requests.

Type	Item	Description
	Security operation cost	Costs for using security services, such as network firewalls, application firewalls, and data security protection
	Other service cost	Costs for other cloud services such as middleware, big data, AI, and IoT
	Cloud management cost	Costs for cloud management services such as monitoring, logging, O&M, auditing, and governance
	Cloud migration cost	One-off costs for cloud migration, resource deployment, and integration testing
	Technical support cost	Costs for support plans
	Labor cost	Salaries, benefits, and training costs for IT O&M (mainly application O&M) personnel

For traditional data centers, depreciation of IT devices also needs to be considered. Generally, IT devices need to be upgraded or replaced every three to five years. Therefore, the TCOs for traditional data centers and the cloud should be compared based on the depreciation period (three to five years) of IT devices.

Calculating costs is complex. First, the prices of IT devices and cloud services change dynamically. For example, cloud service providers frequently adjust the prices of cloud services. You can refer to the **price calculator** and commercial discounts provided by Huawei Cloud when calculating the cost of cloud services. Second, the prices of the same IT devices and cloud services vary depending on the countries and regions. Also, the electricity costs and labor wages differ significantly between different locations. Third, you need to obtain the quantity and configuration specifications of current IT resources, and then map them to cloud resources with different specifications one by one. You can then accurately estimate the TCOs. Huawei Cloud provides an Excel template for you to easily compare the TCOs for traditional data centers and the cloud. You can contact your sales personnel to obtain the template. In addition, when calculating the labor cost, you need to consider the IT O&M cost saved by improving the IT O&M efficiency. For details, see the last point "Improved IT O&M efficiency".

According to the reports and cloud migration practices of many enterprises, cloud transformation can reduce the IT infrastructure TCO by 10% to 30%. If the 3-year IT infrastructure TCO of a traditional data center amounts to USD80 million, the TCO is expected to decrease by USD8 million to USD24 million after cloud transformation.

- **New revenues from business innovations and market expansion**

  After cloud transformation, enterprises can innovate products, services, and business models with the advanced technologies of cloud platforms and quickly enter the global market based on the global layout of cloud service providers. If business innovations and market expansion can bring 500,000 active users to an enterprise, and each user contributes an average of USD60 per year, an additional revenue of USD30 million can be generated for the enterprise per year.

- **Reduced carbon emissions**

  Thanks to the economies of scale and more efficient energy utilization of large cloud data centers, as well as the extensive use of renewable energy by cloud service providers, enterprises can greatly reduce energy consumption and carbon emissions after cloud transformation. Calculating carbon emissions involves many factors. You can use the carbon emission calculator provided by cloud service providers. If the annual carbon emissions are reduced by 50% (from 10,000 tons to 5,000 tons) after cloud transformation and the carbon transaction price is USD100 per ton, the annual revenue from carbon emission reduction is USD500,000.

- **Improved IT O&M efficiency**

  After cloud transformation, enterprises do not need to manage IT infrastructure. Besides, they can use the intelligent monitoring system and automatic O&M tools provided by cloud service providers to greatly improve IT O&M efficiency. If the number of servers that each O&M engineer can manage doubles from 100 to 200, the annual salary of each O&M engineer is USD0.2 million, and an enterprise has a total of 2,000 servers, then the O&M cost can be reduced by USD2 million each year. You should note that this revenue is included in the revenue generated by reducing the TCO of IT infrastructure.

  The preceding benefits are just estimates. These benefits cover cost saving, loss reduction, and revenue increase, but do not include indirect benefits such as brand value improvement from improved system availability, improved security and compliance, and reduced carbon emissions. By taking the aforementioned benefits into consideration, you can get the total benefits of a cloud transformation project. You can then calculate ROI based on the total benefits and the total investment to help managers make strategic decisions.

2.7 Formulating a Cloud Transformation Strategy

After identifying drivers, assessing cloud maturity, setting cloud transformation objectives, and analyzing potential benefits, you need to formulate a cloud transformation strategy. To be specific, follow the following steps:

Step 1 Craft a vision.

A vision represents an organization's expectations and blueprint for the future. It is the direction toward which all employees collectively strive. You need to design the vision based on the organization's mission, values, and insights into the industry. A vision should be forward-looking, inspiring, and instructive. For example, a powerful vision can be: "Through cloud transformation, we will build a flexible, efficient, and secure digital platform to empower business innovations and improve customer experience, and finally become an industry-leading intelligent enterprise." This vision clearly states that the goal of cloud transformation is to build a digital platform, highlights the key features of the platform, and finally points to the direction of becoming an industry leader. Such a clear, inspiring vision can motivate employees and provide them with clear goals to strive toward. When designing a vision, the leadership team must deeply think about what the core competencies of the organization are and how cloud transformation will enhance these competencies. In addition, industry trends, technology development, and customer requirements need to be considered to ensure that the vision is realistic and feasible.

Step 2 Set goals.

Setting specific cloud transformation goals is a key step to turn the vision into executable actions. Cloud transformation goals must be SMART and be aligned with the organization's business strategy. For example, a goal can be "reducing the IT infrastructure OpEx by 15% within two years" or "building an elastic, scalable service system and improving resource utilization by 30% within one year". These two example goals have clear measurement criteria, making it easier to track and evaluate the progress in the future. For details about how to set a goal, see **Setting Cloud Transformation Objectives**.

Step 3 Align goals with stakeholders.

Cloud transformation requires the support and approval of senior executives and stakeholders. To achieve this, in-depth communication with them is required to ensure that the goals of cloud transformation align with those of stakeholders. At this phase, the organization can hold a high-level strategic meeting, inviting the CEO, CIO, heads of various business departments, and key stakeholders to participate. The meeting should focus on the vision and goals of cloud transformation, the impact of transformation on each department, and the expected benefits. For example, the IT department may be concerned about changes in the technical architecture, while the business department is concerned about how cloud transformation promotes business growth. Through such communication, all parties can understand the importance of cloud transformation, recognize its value, and provide their suggestions and opinions. During this process, the organization needs to listen to the concerns of stakeholders, respond to their questions in a timely manner, and adjust the strategy to adapt to the actual situation. This alignment process facilitates easy cloud transformation and ensures cross-department collaboration.

Step 4 Invite senior executives to officially announce the strategy.

After aligning the goals with stakeholders, the organization should officially announce the cloud transformation strategy. This not only reflects the importance that senior executives attach to cloud transformation but also fosters the understanding and recognition of the strategy among all employees. The organization can plan an all-hands meeting or a live streaming event for the CEO or other senior executives to officially announce the cloud transformation strategy. At the event, senior executives can elaborate on the vision and goals of cloud transformation, outline future development prospects, and raise expectations for employees. For example, the CEO can say, "Cloud transformation is an important step for us to move toward a digital future. It will empower our business innovation and improve customer experience. I believe that through our joint efforts, we will achieve this goal." Such a formal announcement can enhance the authority of the strategy and inspire employees' enthusiasm for participation.

Step 5 Promote the strategy.

Promoting the strategy within the organization is a key step to ensure that the cloud transformation strategy becomes a common understanding among all employees. Strategy promotion should be conducted in multiple ways for employees at different levels and in different departments. The organization can hold a series of presentations and training courses and create brochures and videos to explain the significance, goals, and specific measures of cloud transformation in detail. For example, the organization can set up a cloud transformation column on the internal website to regularly update related information and progress. Besides, department heads should actively convey the strategy to team members and explain how cloud transformation will affect their work and what contributions they can make, based on the actual situation of the department. These efforts can help employees better understand the strategy and better engage in cloud transformation.

----End

In conclusion, it is a systematic and comprehensive process for an organization to formulate a cloud transformation strategy. By crafting a clear and inspiring vision, setting specific goals, aligning goals with stakeholders, inviting senior executives to officially announce the strategy, and deeply promoting the strategy within the organization, a solid foundation can be laid for cloud transformation. To address new challenges and embrace new opportunities in the digital era, organizations should have a strategic vision and actively adopt the cloud computing technology to take the lead in the industry.

2.8 Anti-patterns in Strategy Formulation

Some common anti-patterns in strategy formulation may hinder the success of cloud transformation, and even result in waste of enterprise resources and service interruptions. Identifying and avoiding these anti-patterns is crucial to a successful cloud transformation. Below gives some common anti-patterns and corresponding optimization suggestions.

- **The cloud transformation strategy is not aligned with the business strategy.**

  This anti-pattern indicates that cloud transformation is not closely integrated with the company's overall business strategy, which turns cloud transformation into an isolated initiative within the IT department. The cloud transformation objectives are not aligned with business objectives, and senior executives lack understanding of the significance and value of cloud transformation. This results in insufficient support or low engagement in cloud transformation. All of these will further lead to insufficient resource allocation and transformation direction deviation, making it difficult to achieve the expected business value. For example, an enterprise blindly pursues cloud migration and adopts the latest cloud technology without considering whether it can truly address business pain points or improve efficiency. Instead, this will increase costs and complexity.

  The optimization suggestions for this anti-pattern are as follows:

  – **Closely integrate the cloud transformation strategy with the business strategy**:

    ■ Specify how cloud transformation supports the achievement of business objectives. For example, cloud transformation can help accelerate business innovations, reduce costs, improve customer experience, and explore new markets.

    - Use business language to explain the value of cloud transformation and avoid using pure technical terms.
  - **Obtain support and participation from senior executives**:
    - Report the value and expected business benefits of cloud transformation to senior executives to win their support and resource investment.
    - Invite senior executives to participate in the formulation and execution of the cloud transformation strategy to ensure that the transformation direction is aligned with the company's overall strategy.
- **The cloud transformation strategy focuses on technical benefits but ignores business benefits.**

  This anti-pattern is manifested in the excessive focus on improving technical metrics, such as resource elasticity, storage capacity, or service level objectives (SLOs), while ignoring the actual impact of cloud transformation on the business. Although it is important to improve technical metrics, the ultimate goal is to improve the business value through technical improvement. If only technical benefits are concerned, the ROI may be low, and there may even be negative effects on the business. For example, after cloud migration, an enterprise excessively pursues technical benefits such as performance, elasticity, and reliability, but ignores cost optimization and operations. As a result, the cost increases rapidly.

  The optimization suggestions for this anti-pattern are as follows:
  - **Set business-centered cloud transformation objectives**:
    - Determine the objectives and direction of cloud transformation based on business requirements.
    - Link the improvement of technical metrics to business benefits. For example, enhance service continuity by improving system resilience, scalability, and security. For details, see **Setting Cloud Transformation Objectives**.
  - **Quantify business benefits**: Analyze the benefits based on cloud transformation objectives and convert them into financial revenues to evaluate the project ROI, helping managers make strategic decisions. For details, see **Analyzing the Benefits of Cloud Transformation**.
  - **Continuously track and evaluate business benefits**: Regularly evaluate the business benefits of cloud transformation and adjust the cloud transformation objectives if needed.
- **The cloud transformation strategy is not aligned with stakeholders.**

  Cloud transformation involves multiple departments and teams inside the company, such as the IT, business, and finance departments, as well as external partners and customers. A lack of communication and alignment with all stakeholders can lead to obstacles or even failures during the transformation. For example, if the IT department starts migrating service systems to the cloud without sufficient communication with the business department, the service systems will be interrupted and business operations will be affected.

  The optimization suggestions for this anti-pattern are as follows:

- **Analyze stakeholders' interests**: Identify all departments, teams, and individuals that are involved in cloud transformation decision-making or affected by cloud transformation. Understand the interests of different stakeholders and develop corresponding strategies to meet their needs to reduce potential resistance. For details, see **Analyzing Stakeholder Interests**.
- **Conduct proactive communications**: Develop a detailed communication plan and specify the communication objectives, content, methods, and schedule. Communicate with stakeholders in multiple ways, such as meetings, trainings, emails, and internal websites, to ensure that all stakeholders can understand the progress and impact of cloud transformation and what support they need to provide.
- **Establish a feedback mechanism**: Encourage stakeholders to provide comments and suggestions, and actively adopt reasonable suggestions.

Cloud transformation is complex and full of challenges. A successful cloud transformation requires careful planning, sufficient communication, and continuous optimization. By identifying and avoiding the preceding anti-patterns, enterprises can better avoid cloud transformation risks, ensure the alignment between transformation strategies and business strategies, achieve expected business values, and lay a solid foundation for future development.

3 Top-Level Planning

3.1 Overview

Enterprise cloud transformation is a complex and systematic project. It involves multiple aspects such as organizations and processes, platforms and architectures, as well as O&M and management. Just like building a skyscraper, you need to design a blueprint before digging the foundation. Similarly, enterprises need to make comprehensive and clear top-level planning before building cloud infrastructure and migrating business systems to the cloud. To maximize the benefits of the cloud and boost business value, thorough planning and preparation are needed.

In terms of organizations and processes, you need to design a Cloud Center of Excellence (CCoE). The CCoE is a core organization for promoting enterprise cloud transformation. It is responsible for developing cloud standards, best practices, and governance frameworks, and coordinates cooperation between business units to ensure efficient cloud transformation. In addition, the application lifecycle management process needs to be transformed. Traditional development and deployment modes cannot meet the fast iteration requirements of the cloud environment. Advanced methods such as agile development and DevOps can help improve development efficiency, shorten the delivery period, and improve the response efficiency to market changes.

In terms of platforms and architectures, Well-Architected Framework (WAF) provides a set of best practices and architecture design principles to help enterprises build secure, highly available, high-performance, and cost-optimized cloud infrastructure and application systems on the cloud. Landing Zone provides a secure, compliant, and scalable multi-account environment on the cloud, which accelerates application deployment and improves security. In addition, the planning and design of platform engineering is also important. It provides standard tools, processes, and infrastructure support for development teams to improve development efficiency, reduce complexity, and accelerate software delivery.

In terms of O&M and project management, the design of the cloud operating mode is critical to efficient collaboration between CCoE and application teams. A suitable cloud operating mode that aligns with enterprise's internal collaboration mode and application systems' characteristics needs to be set up. This ensures that application systems are iterated agilely and run stably. In addition, a detailed management plan for the cloud transformation project, covering project planning, project appointment, progress management, and risk management, can ensure that all tasks are carried out in an orderly manner as planned. The plan can also improve project transparency and controllability, and reduce uncertainty during implementation.

In a word, to successfully implement cloud transformation, enterprises must make comprehensive top-level planning and design in the early stage. This includes building excellent organizational structures, optimized processes, efficient platforms and architectures, and sound cloud operating models and project management. Without these key top-level designs, there may be chaos and risks after a large number of application systems are migrated to the cloud. Post-event rectification is costly and may severely impact the stability of business systems. Top-level planning is critical to the smooth execution and long-term success of cloud transformation.

3.2 CCoE

3.2.1 What Is CCoE?

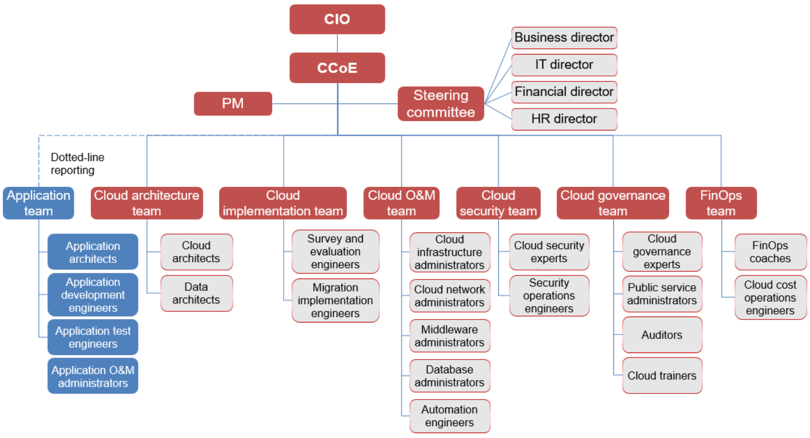
Enterprise cloud transformation is a complex and systematic project. A Cloud Center of Excellence (CCoE) is required to lead, coordinate, and promote an entire cloud transformation project. CCoE is a centralized team established by an enterprise for cloud transformation. It is responsible for the entire cloud transformation journey, including strategy development, top-level planning, survey and evaluation, solution design, adoption and implementation, and O&M governance. It aims to help enterprises maximize the benefits of cloud computing and ensure the successful implementation of cloud transformation projects by providing best practices, guidance, and resources. CCoE has the following main responsibilities:

- **Cloud strategy development**: Develop a cloud strategy that aligns with the enterprise's business objectives, assess the maturity of cloud transformation, specify the cloud transformation objectives and expected benefits, and plan the specific implementation roadmap.
- **Top-level planning for cloud transformation**: Plan and design the cloud transformation projects at the top level, including the optimization of the application cloud transformation process, landing zone design, platform engineering design, and cloud operating model design.
- **Cloud governance framework setup**: Develop the governance framework, governance policies, security standards, and compliance requirements for the cloud platform to ensure the security, stability, and compliance of cloud infrastructure and application systems.
- **Cloud technical support**: Provide training, consulting, and support on cloud technologies for enterprises to help departments better understand and use cloud technologies.
- **Cloud platform and resource management**: Manage the routine operations of the cloud platform, including resource allocation, cost control, and performance monitoring.

- **Cloud best practices promotion**: Promote cloud best practices, such as Cloud Adoption Framework (CAF) and Well-Architected Framework (WAF), to help enterprises build efficient, reliable, and secure cloud infrastructure and application systems.
- **Supplier management**: Evaluate and manage cloud service providers, and select the most appropriate cloud service provider and its advantageous services based on service requirements to maximize the value of cloud computing.
- **Cross-department collaboration promotion**: As coordination center across departments, facilitate communication and collaboration between business departments, IT departments, and other related departments to ensure smooth progress of cloud transformation projects.
- **Continuous improvement and optimization**: Continuously track the progress of cloud transformation. Adjust and refine the projects as needed to ensure that the cloud transformation objectives are achieved.

To maximize the effectiveness of CCoE, the CCoE members must include the stakeholders of the enterprise, including professionals from the business, IT, finance, and HR departments, to set up a cross-functional team. This not only ensures that cloud transformation is closely aligned with business objectives, but also effectively solves various problems during transformation. The CIO should directly lead the CCoE organization. This can effectively improve the reporting and decision-making efficiency during cloud transformation. Huawei has its own proven success in cloud transformation. It has also accumulated extensive experience by assisting numerous customers in their cloud transformation journeys. Based on this expertise, we recommend the following CCoE organization structure and key roles. You can tailor and refine the CCoE organization structure based on your IT organization, employee skill levels, and cloud transformation objectives to design the most suitable CCoE organization structure for your enterprise.

Figure 3-1 CCoE organization structure



The following sections describe the responsibilities and skill requirements of the teams and roles in the CCoE organization structure. We also provide suggestions on obtaining suitable workforces. An employee can assume one or more roles, depending on their skills and the enterprise's workforce budget.

3.2.2 Steering Committee

A steering committee provides suggestions, strategic guidance, and decision-making support for cloud transformation projects. It plays a crucial role in cloud transformation. Members of the steering committee must be appointed by key stakeholders (such as the CEO, CIO, CTO, CFO, CISO, and business directors) of cloud transformation. The members should include at least the business director, IT director, finance director, and HR director. The steering committee makes collective decisions based on the parliamentary system and shares the following responsibilities:

- **Cloud strategy development**: Develop a cloud strategy that aligns with the enterprise's business objectives, assess the maturity of cloud transformation, specify the cloud transformation objectives and expected benefits, and plan the specific implementation roadmap.
- **CCoE preparation**: Prepare for and establish the CCoE, specify the responsibilities and skill requirements of each role in the CCoE organization, and coordinate related departments to quickly obtain various human resources required by the CCoE.
- **Top-level planning of cloud transformation**: Guide CCoE members including the cloud architecture team, cloud O&M team, cloud security team, and cloud governance team to perform top-level planning for cloud transformation, including the optimization of the application cloud transformation process, landing zone design, platform engineering design, and cloud operation mode design. The steering committee takes the final responsibility for the effect of the top-level designing.
- **Business requirements clarification**: Specify business-oriented requirements and expectations for cloud transformation, such as improving business continuity and agility, and boosting revenues.
- **Budget approval and oversight**: Review and approve budgets related to cloud transformation. Supervise the budgets and ensure that budgets are spent appropriately and effectively.
- **HR management**: Recruit, select, train, and retain CCoE members to build a stable and high-performance CCoE organization.
- **Cross-department collaboration**: Promote communication and collaboration among business departments, IT departments, and other related departments to ensure that all departments understand and support the cloud transformation and avoid conflicts and obstacles between departments.
- **Cloud transformation effect evaluation**: Evaluate the effect and value of cloud transformation, such as cost saving, efficiency improvement, and service innovation. Adjust and optimize the transformation strategies based on the evaluation results.
- **Decision-making on key issues**: The steering committee is the highest decision-making body in cloud transformation. It is responsible for making decisions on key issues in cloud transformation, such as selecting the cloud service provider, specifying the technical solution, and adjusting the implementation plan.

3.2.3 Application Team

An enterprise usually has multiple business departments. Each business department is responsible for the investment, setup, and O&M of their own business systems, so they usually set up their own application teams. The application teams need to work with different teams in cloud transformation. They need to work with the cloud implementation team to assess the business system conditions, migrate business systems, modernize applications, and test and verify the results. They need to work with the cloud architecture team to design the cloud-based application architecture for business systems based on cloud technologies and cloud services. They also need to work with the cloud O&M team to ensure that business systems run securely and stably in the cloud. Members of an application team usually come from business departments. Since business departments have their own application teams, there may be multiple application teams. These application teams report to the CCoE organization indirectly. An application team usually consists of application architects, application development engineers, application test engineers, and application O&M administrators. The following table lists their responsibilities and skill requirements.

Table 3-1 Roles and responsibilities of an application team

Role	Responsibility	Skill Requirements	Source
Application architect	● Clarify the benefits of cloud-based business systems, such as business continuity and agility. ● Develop migration strategies (Rehost, Replatform, and Refactor) and migration sequence for business systems. ● Work with the cloud implementation team to assess the business system conditions and provide information such as resources, application architecture, deployment architecture, and dependencies. ● Design and manage the cloud-based application architecture for business systems, including the architecture pattern, technology selection, and deployment mode. Ensure the performance, scalability, security, and reliability of applications. ● Work closely with data architects and cloud architects to ensure that the application architecture is compatible with the data architecture and cloud architecture. ● Guide the development team in developing and deploying applications.	● Have an in-depth understanding of various application architecture patterns and design patterns, such as microservice architecture and event-driven architecture. ● Be familiar with various development languages and frameworks. ● Be familiar with DevOps practices and tools. ● Excel in code design and development. ● Understand application security best practices. ● Excel in communication and teamwork.	Business department

Role	Responsibility	Skill Requirements	Source
Application development engineer	● Migrate existing applications to the cloud platform, including code, data, and database migration. ● Modernize applications, for example, split monolithic applications into microservices or use serverless and event-driven architectures. ● Refactor existing code to make it more maintainable, scalable, and testable, and optimize it for the cloud environment, for example, using cloud native services and APIs.	● Be proficient in at least one mainstream programming language, such as Java, Python, and Go. ● Be familiar with DevOps practices and tools. ● Excel in code design and development. ● Be familiar with mainstream cloud platforms and cloud services. ● Be able to effectively communicate and work with other teams.	Business department
Application test engineer	● Design test cases and develop test plans for cloud-based business systems. The test cases should cover functions, performance, security, and reliability. ● Based on the test cases and test plans, select appropriate test tools to comprehensively test the cloud-based business systems, including the functions, performance, security, and reliability. ● Write and maintain automated test scripts. ● Write test reports and documents.	● Have a solid foundation in test theories, and be familiar with software test theories, methods, and processes. ● Have extensive testing experience and be familiar with various test types, such as function, performance, security, and reliability tests. ● Be familiar with mainstream cloud platforms and cloud services. ● Be proficient in using automated test tools and be able to write automated test scripts. ● Be able to effectively communicate and work with other teams.	Business department

Role	Responsibility	Skill Requirements	Source
Application O&M administrator	● Deploy, monitor, and maintain cloud-based business systems to ensure that they can run securely and stably. ● Handle runtime errors of applications and optimize application performance. ● Work with the development team to update and release application versions. ● Monitor application logs, and analyze and resolve potential problems.	● Be familiar with the APM service on the cloud platform and capable of analyzing performance metrics and logs. ● Master CI/CD tools and container orchestration tools. Be familiar with common application deployment modes (such as containerization and microservice architecture). ● Be familiar with the O&M management of common middleware (such as Nginx, Redis, and Kafka).	Business department

3.2.4 Cloud Architecture Team

The cloud architecture team plays a key role in cloud transformation. Based on The Open Group Architecture Framework (TOGAF) and Well-Architected Framework (WAF), it is responsible for designing the technical and data architectures on the cloud. The team needs to work with application architects to design the cloud-based application architecture for business systems based on cloud technologies and cloud services. This helps enterprises build secure, highly available, high-performance, and cost-optimized cloud infrastructure and application systems. The cloud architecture team usually consists of cloud architects and data architects. The following table lists their responsibilities and skill requirements.

Table 3-2 Roles and responsibilities of a cloud architecture team

Role	Responsibility	Skill Requirements	Source
Cloud architect	● Plan and design the overall architecture of the cloud platform and cloud infrastructure, including landing zone, platform engineering, network, storage, security, and disaster recovery. Ensure the security, reliability, performance, and cost-effectiveness of the cloud infrastructure. ● Select appropriate cloud service providers and cloud service types. ● Develop and promote cloud architecture design principles to empower application architects and data architects to design robust technical architectures on the cloud. ● Lead and guide the cloud implementation team to ensure the implementation of technical solutions.	● Have an in-depth understanding of cloud computing technologies and architectures and be familiar with mainstream cloud platforms. ● Have extensive knowledge and experience in landing zone, platform engineering, network, security, storage, and disaster recovery. ● Be familiar with architecture frameworks such as TOGAF and WAF. ● Have good communication skills, teamwork, and leadership.	Enterprise architecture team or external recruitment

Role	Responsibility	Skill Requirements	Source
Data architect	● Design and manage cloud-based data architecture, including data storage, data processing, data integration, and data governance. ● Select appropriate data storage solutions, such as relational databases, NoSQL databases, and data warehouses. ● Ensure data quality, security, and compliance. ● Work closely with application architects and cloud architects to ensure that the data architecture is compatible with the overall architecture.	● Have an in-depth understanding of concepts and technologies about data modeling, data warehouse, data lake, and data governance. ● Be familiar with various database technologies, including relational databases and NoSQL databases. ● Be familiar with big data technologies, such as Hadoop, Spark, and Flink. ● Have data analysis and mining capabilities. ● Be familiar with regulations and standards about data security and data privacy. ● Excel in communication and teamwork.	Big data department or external recruitment

3.2.5 Cloud Implementation Team

The cloud implementation team is responsible for migrating or directly deploying business systems to the cloud. This requires the designing and implementation of technical solutions based on a detailed survey and evaluation of the enterprise's existing IT infrastructure and business systems. The cloud architecture team is responsible for the designing part, and the cloud implementation team is responsible for the implementation part. The cloud implementation team usually consists of survey and evaluation engineers and migration implementation engineers. The following table lists their responsibilities and skill requirements. The two roles listed in the following table are temporary. After all business systems are migrated to the cloud, these roles are no longer needed. So these roles can be temporarily undertaken by engineers with relevant skills from the IT department or outsourced to professional cloud migration service providers.

Table 3-3 Roles and responsibilities of a cloud implementation team

Role	Responsibility	Skill Requirements	Source
Survey and evaluation engineer	● Status survey: Conduct comprehensive survey and document recording on the enterprise's existing IT infrastructure, business systems, application architecture, data storage, and security policies, including hardware configurations, software versions, dependencies, performance metrics, and security vulnerabilities. ● Requirement analysis: Communicate with business departments to understand their requirements and expectations for cloud transformation, such as performance improvement, cost optimization, and disaster recovery. Then, convert these requirements into specific technical indicators. ● Feasibility evaluation: Evaluate the feasibility of migrating existing systems to the cloud platform, including technical feasibility, cost-effectiveness, and risk assessment. ● Capacity planning: Plan the capacity of cloud resources such as compute, storage, and network bandwidth based on service requirements and future development trends. ● Cost estimation: Estimate the cost of cloud migration based on the pricing model of the cloud service provider and compare it with the cost of the traditional IT architecture.	● Be familiar with mainstream cloud platforms and cloud services. ● Have solid knowledge of IT infrastructure, including servers, networks, storage, databases, and middleware. ● Be familiar with various operating systems and application software. ● Understand different migration policies and methods. ● Have experience in conducting survey and evaluation on IT infrastructure and business systems. ● Excel in communication and teamwork.	IT department or outsourced to a professional cloud implementation service provider

Role	Responsibility	Skill Requirements	Source
	This can provide a basis for decision-making.		
Migration implementation engineer	• Migration solution implementation: Migrate and deploy business systems based on the technical solution designed by the cloud architect and the report provided by the survey and evaluation engineer, including environment setup, data migration, application deployment, and configuration adjustment. • Test and verification: Perform comprehensive testing and verification on the cloud-based system to ensure that the system functions properly, the performance is stable, and the system is secure and reliable. • Troubleshooting: Handle various problems and faults that occur during the implementation in a timely manner to ensure smooth implementation.	• Be proficient in mainstream cloud platforms and cloud services and have relevant certification qualifications. • Be familiar with various migration tools and technologies, such as data migration tools, container-based technologies, and automatic deployment tools. • Be familiar with various operating systems and application software. • Have solid script writing skills (such as Shell and Python) to implement automatic operations. • Excel in communication and teamwork.	IT department or outsourced to a professional cloud implementation service provider

3.2.6 Cloud O&M Team

The cloud O&M team is responsible for the routine management and maintenance and ensures high availability, high security, and high performance of cloud infrastructure. The team needs to work with application O&M administrators to ensure that cloud-based business can run stably and securely. The team also needs to continuously improve O&M efficiency through automation and intelligent technologies. The cloud O&M team usually consists of cloud infrastructure administrators, cloud network administrators, database administrators, and automation engineers. The following table lists their responsibilities and skill requirements.

Table 3-4 Roles and responsibilities of a cloud O&M team

Role	Responsibility	Skill Requirements	Source
Cloud infrastructure administrator	● Perform routine O&M management on infrastructure such as storage, VMs, and operating systems on the cloud platform. ● Monitor and optimize the usage of cloud resources to ensure proper resource allocation. ● Handle VM, storage, and operating system faults to keep the system highly available. ● Periodically update system patches and harden system security.	● Be familiar with VMs and cloud storage services of mainstream cloud platforms. ● Master the management and optimization of Linux and Windows. ● Be familiar with cloud native monitoring and O&M tools. ● Be able to write scripts. ● Excel in troubleshooting and problem-solving capabilities	IT department
Cloud network administrator	● Design, configure, and perform routine O&M on the cloud network architecture to ensure network stability and security. ● Manage network components such as VPNs, private lines, VPCs, subnets, network ACLs, routes, load balancers, and firewalls. ● Monitor network performance, rectify network faults, and optimize network latency and bandwidth usage. ● Ensure network security and prevent network threats such as DDoS attacks.	● Be familiar with network services (such as VPC, VPN, private line, load balancer, and firewall services) and their configurations on the cloud platform. ● Be familiar with network protocols such as TCP/IP, HTTP, DNS, and TLS. ● Have the capability of troubleshooting network faults. ● Be familiar with network security technologies (such as firewall rule configuration and intrusion detection).	IT department

Role	Responsibility	Skill Requirements	Source
Middleware administrator	- Install, configure, and maintain message queue services (such as Kafka and RabbitMQ), web servers (such as Nginx and Apache), application servers (such as Tomcat and JBoss), and cache services (such as Memcached and Redis). - Monitor performance metrics, identify performance bottlenecks, and improve performance and efficiency of middleware services. - Quickly diagnose and rectify faults and problems of middleware services to ensure business continuity.	- Master common middleware technologies, such as Kafka, RabbitMQ, Nginx, and Tomcat. - Be familiar with the deployment and management of middleware services on mainstream cloud platforms. - Be familiar with operating systems such as Linux and Windows Server. - Understand DevOps concepts and practices. - Be able to write scripts. - Excel in troubleshooting and problem-solving capabilities	IT department

Role	Responsibility	Skill Requirements	Source
Database administrator	● Deploy, configure, monitor, and maintain cloud databases. ● Ensure high availability and data security of databases and periodically perform backup and recovery drills. ● Optimize database performance and solve problems such as slow query and lock waiting. ● Manage database permissions and access control to ensure data compliance.	● Be familiar with database services and database management services on the cloud platform. ● Be familiar with the management of mainstream databases (such as MySQL and PostgreSQL). ● Master database performance optimization technologies (such as index optimization and database sharding/partitioning). ● Have O&M experience in database backup and restoration, primary/secondary synchronization, and distributed architecture. ● Be familiar with database security policies and data encryption technologies.	IT department
Automation engineer	● Develop and maintain automation O&M tools to improve O&M efficiency. ● Implement automatic deployment, monitoring, and expansion of cloud resources. ● Write scripts or code to automate routine O&M tasks. ● Promote the application of intelligent O&M technologies, such as AIOps.	● Be familiar with automation tools (such as Ansible, Terraform, and SaltStack). ● Master script languages (such as Python and Shell) and cloud platform APIs. ● Understand DevOps and be familiar with the CI/CD process and tools. ● Understand AIOps technologies.	IT department

3.2.7 Cloud Security Team

The cloud security team is responsible for the security assurance of cloud infrastructure and cloud-based business systems. It is mainly responsible for cloud platform security solution design, access control and permission management, security monitoring and threat detection, vulnerability scanning and fixing, data encryption and privacy protection, compliance review and risk assessment, and emergency response and security incident handling. This can ensure the security, compliance, and stability of cloud-based business systems. The cloud security team usually consists of cloud security experts and security operations engineers. The following table lists their responsibilities and skill requirements.

Table 3-5 Roles and responsibilities of a cloud security team

Role	Responsibility	Skill Requirements	Source
Cloud security expert	● Design and optimize the overall security solution of the cloud platform, and formulate security policies and standards. ● Evaluate security risks of cloud infrastructure and business systems, and provide improvement solutions. ● Design and implement identity security, network security, data security, application security, host security, and security O&M solutions. ● Guide and review the work of security operations engineers and provide technical support. ● Keep up with the latest security technologies and develop countermeasures.	● Have an in-depth understanding of cloud security services and security configuration baselines of the cloud platform. ● Be familiar with identity security, network security, data security, application security, host security, and security O&M. ● Master security evaluation tools and penetration testing technologies. ● Have experience in security compliance management (such as DJCP 2.0 and ISO 27001). ● Have excellent security policy formulation and technical guidance capabilities.	IT department

Role	Responsibility	Skill Requirements	Source
Security operations engineer	● Design and implement continuous security operations solutions. ● Be responsible for routine security monitoring and O&M of the cloud platform, and detect and handle security events in a timely manner. ● Perform vulnerability scanning, patch management, and security configuration hardening. ● Implement access control, permission management, and log audit to ensure system compliance. ● Work with cloud security experts to implement and optimize security technology solutions. ● Write security O&M scripts to improve security operations efficiency.	● Be proficient in using security operations services and various security monitoring tools of the cloud platform. ● Master threat detection technologies, vulnerability scanning tools, and patch management processes. ● Be familiar with log analysis tools and automatic script languages (such as Python and Shell). ● Understand security configurations (such as security groups and firewall rules) of the cloud platform. ● Be able to quickly respond to and handle security incidents.	IT department

3.2.8 Cloud Governance Team

The cloud governance team identifies risks in the enterprise cloud transformation, and develops and implements effective governance frameworks, policies, and processes. The purpose is to minimize risks and maximize business benefits of cloud transformation. The cloud governance team usually consists of cloud governance experts, auditors, and cloud trainers. The following table lists their responsibilities and skill requirements.

Table 3-6 Roles and responsibilities of a cloud governance team

Role	Responsibility	Skill Requirements	Source
Cloud governance expert	● Identify and evaluate various risks in cloud transformation and develop mitigation measures. ● Develop and maintain the cloud governance framework, including policies, standards, processes, and guidelines. Promote the implementation and execution of cloud governance best practices. ● Ensure that cloud governance policies are aligned with business objectives. ● Continuously optimize the cloud governance framework to adapt to changing business needs and technology trends. ● Monitor the compliance and security of the cloud environment.	● Have an in-depth understanding of cloud architecture, cloud security, and cloud cost optimization. ● Be familiar with cloud services and best practices of mainstream cloud platforms. ● Have extensive experience in risk management, compliance management, and IT governance. ● Have excellent cross-department communication, collaboration, and problem-solving skills.	IT department
Public service administrator	● Identify public IT services and resources required by each business unit, such as NTP servers, AD servers, self-built DNS servers, OBS buckets, and container image libraries, or PaaS services such as CodeArts. ● Deploy and maintain these public IT services and share them with all business units in the enterprise. ● Ensure that public IT services run securely and smoothly.	● Be familiar with IaaS and PaaS services of mainstream cloud platforms and be able to deploy these services. ● Be familiar with technical solutions for resource sharing on cloud platforms, such as network-based sharing, resource permission policy-based sharing, and Huawei Cloud Resource Access Manager **(RAM)**. ● Have good cross-department communication, collaboration, and problem-solving skills.	IT department

Role	Responsibility	Skill Requirements	Source
Auditor	- Regularly audit the cloud environment to evaluate whether it complies with related regulations, standards, and best practices. - Identify security vulnerabilities and compliance risks in the cloud environment. - Write audit reports and provide improvement suggestions. - Work with cloud governance experts and related teams to resolve identified risks. - Track and monitor the implementation of improvement measures.	- Be familiar with cloud security, compliance audit, and risk assessment methods. - Be familiar with related regulations and standards, such as DJCP 2.0 and ISO 27001. - Be able to analyze data and write reports. - Have good communication and interpersonal skills. - Have basic knowledge of cloud technologies.	IT department
Cloud trainer	- Develop and deliver cloud computing training courses covering cloud infrastructure, cloud architecture design, cloud O&M, and cloud security. - Promote best practices of cloud adoption journey to reduce cloud migration risks.	- Have solid cloud technology knowledge and practical experience. - Have excellent teaching and communication skills to clearly convey complex cloud technology concepts to others. - Be familiar with different training methods and tools.	Outsourced to a cloud service provider

3.2.9 FinOps Team

The FinOps team is responsible for cost lifecycle management and continuous cost operations. It aims to help teams efficiently use cloud resources within the budget and continuously improve the cost-effectiveness of cloud resources to maximize business value. The FinOps team usually consists of FinOps coaches and cloud cost operations engineers. The following table lists their responsibilities and skill requirements.

Table 3-7 Roles and responsibilities of a FinOps team

Role	Responsibility	Skill Requirements	Source
FinOps coach	● Guide and train team members to understand and apply FinOps principles and best practices. Continuously learn and promote new methods of optimizing cloud costs. ● Help develop and implement cloud cost management strategies to ensure that each department can efficiently use cloud resources within the budget. ● Promote cross-department collaboration, drive cost optimization, and improve resource utilization. ● Promote the FinOps culture and concepts within the organization.	● Have an in-depth understanding of the FinOps framework and best practices of cloud cost management. ● Be familiar with the billing modes and cost management tools of mainstream cloud platforms. ● Be familiar with common cloud cost optimization methods. ● Have project management capabilities and be able to drive cross-department collaboration and transformation.	IT department or external recruitment
Cloud cost operations engineer	● Monitor and analyze the usage of cloud resources to identify cost-saving opportunities. ● Generate detailed cost analysis reports and provide data support for decision-making. ● Work with the cloud O&M team and application team to optimize the cost-effectiveness of application systems. ● Implement cost optimization strategies, such as changing billing modes, offering resource packages, and disabling idle resources.	● Be familiar with cost management tools of the cloud platform. ● Be familiar with the billing modes of various cloud services. ● Have data analysis capabilities and be able to extract valuable insights from a large amount of data. ● Have good communication skills and be able to effectively collaborate with technical and financial teams.	IT department or external recruitment

3.2.10 Cloud Project Manager

The cloud project manager is authorized by the steering committee to lead and manage the entire cloud transformation project. They need to ensure that the project is completed on time within the budget and meets quality standards. The main responsibilities of the cloud project manager are as follows:

- **Project planning and goal setting**: Develop an overall plan for cloud transformation, including the project scope, goal, schedule, budget, and acceptance requirements.
- **Project execution and progress management**: Supervise the project execution process to ensure that tasks in each stage are completed as planned; track the project progress, identify potential problems, and take measures in a timely manner to ensure that the project is delivered on time.
- **Budget and cost control**: Supervise and control the usage and costs of cloud resources to ensure that the project operates within budget; identify and implement cost optimization measures to improve the economic benefits of cloud transformation.
- **Quality management**: Ensure that the deliverables meet the specified quality standards and business requirements; organize quality reviews to ensure the stability, security, and scalability of technical solutions.
- **Communication and collaboration**: Communicate effectively with stakeholders (including business, IT, and finance departments, as well as cloud service providers) to ensure project information transparency and gain their support.
- **Risk management**: Identify, evaluate, and manage project risks, develop contingency plans, and minimize risks.
- **Change management**: Manage project changes, ensure that changes are properly evaluated and approved, and minimize the impact on the project.

The cloud project manager is a comprehensive management role. In addition to strong project management expertise, the cloud project manager also needs to have certain cloud technology knowledge and business understanding capabilities. The skill requirements are as follows:

- **Project management skills**: Have solid project management knowledge and experience, be familiar with the Project Management Body of Knowledge (PMBOK) project management methodology, and be proficient in project management tools and technologies.
- **Cloud technology knowledge**: Deeply understand cloud computing concepts, architectures, and services, be familiar with different cloud deployment models (public cloud, private cloud, and hybrid cloud), and understand the characteristics and advantages of mainstream cloud platforms.
- **Communication and coordination capability**: Excel in communication, coordination, and interpersonal skills, be able to effectively communicate and work with different teams and stakeholders, and have cross-department communication and coordination capabilities.
- **Problem-solving capability**: Be able to quickly identify problems, analyze problems, develop solutions, and promote implementation.
- **Leadership**: Be able to lead and motivate the entire cloud transformation project team, ensure efficient collaboration, and create a positive team atmosphere.
- **Business understanding capability**: Understand business requirements and convert business requirements to technical solutions to ensure that business can benefit from cloud transformation.

- **Cost management capability**: Have cost awareness and be able to effectively control project costs and optimize costs during the project lifecycle.

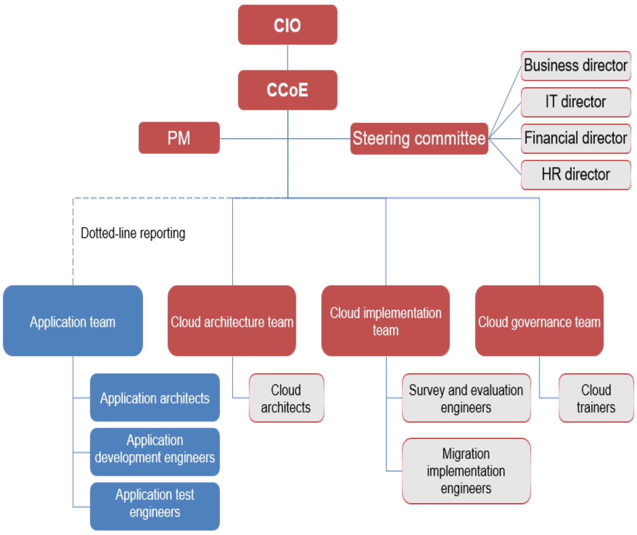
3.2.11 CCoE Evolution

A CCoE organization guides enterprises through their cloud journey. In the early stage of cloud transformation, enterprises do not need to establish a complete CCoE organization.

In the early stage of cloud transformation, the goal is to migrate or directly deploy the first batch of business systems to the cloud. A lean CCoE organization with essential roles can be established to meet the requirements.

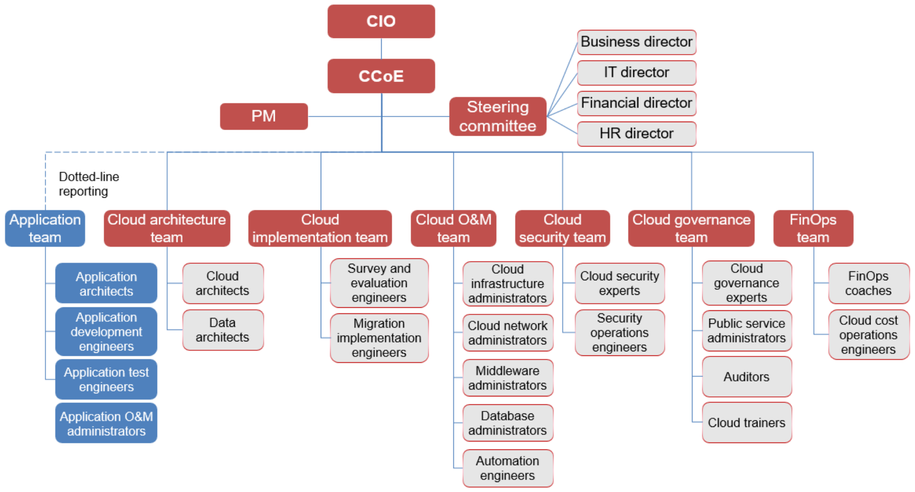
The essential roles include the steering committee, cloud project manager, application architect, application development engineer, application test engineer, cloud architect, survey and evaluation engineer, and migration implementation engineer. These roles work together to help the first batch of business systems go cloud and quickly gain benefits. This can drive the enterprise to move more business systems to the cloud.

Figure 3-2 Lean CCoE organization structure



As enterprise cloud transformation expands and enters the O&M governance stage, more key roles can be added, such as the cloud infrastructure administrator, cloud network administrator, database administrator, application O&M administrator, cloud governance expert, security operations engineer, and cloud cost operations engineer. The lean CCoE gradually evolves into a full-function CCoE.

Figure 3-3 Full-function CCoE organization architecture

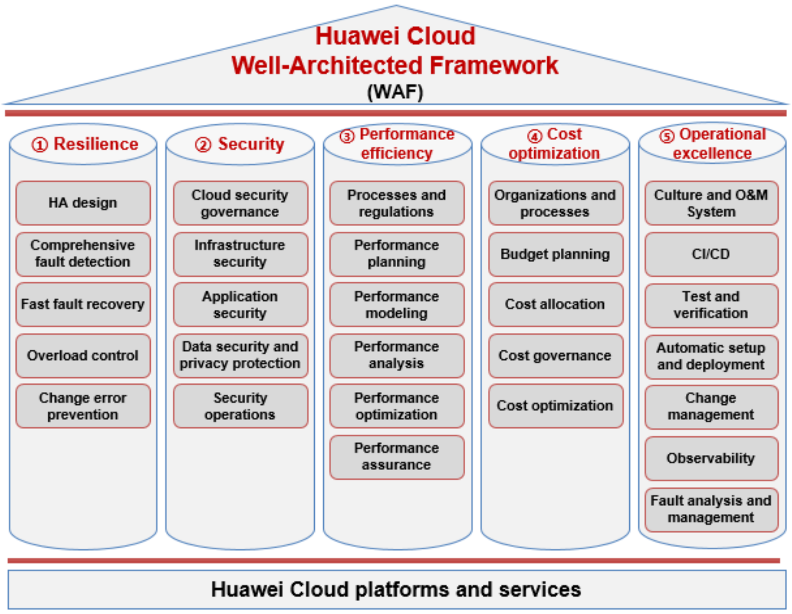


3.3 CCoE Architecture Design

Based on Huawei's cloud architecture design experience and industry best practices, we have developed Well-Architected Framework (WAF) to provide a set of cloud architecture design principles and best practices for architects, software engineers, and O&M engineers. The purpose is to help enterprises design, build, and operate resilient, secure, high-performance, and cost-effective application systems on the cloud.

WAF covers five pillars, as shown in the following figure.

Figure 3-4 Huawei Cloud WAF



- **Resilience**

  Resilience focuses on the continuous availability and reliability of systems in response to faults, pressure, and changes. It uses technologies and processes such as redundant components, comprehensive fault detection, overload control, and fast fault recovery to ensure that application systems can quickly recover from various faults and continue to provide services even in unexpected events or heavy loads. Enterprises need to plan and implement failover as well as backup and recovery policies, and regularly perform disaster recovery drills to verify the resilience of application systems.

- **Security**

  The security pillar is dedicated to protecting the confidentiality, integrity, and availability of information, systems, and assets. It covers application security, data security and privacy protection, infrastructure security, and security operations. Enterprises should use strong security policies, including identity authentication and authorization, encryption of data transmission and storage, network security measures, and continuous monitoring and auditing to detect and handle security threats in a timely manner.

- **Performance efficiency**

  The performance efficiency pillar focuses on how to efficiently use cloud resources to meet system performance requirements and adapt to business changes. It emphasizes performance planning, modeling, analysis, and optimization to ensure that the system can maintain optimal performance under different loads. Enterprises need to continuously monitor system performance metrics, optimize applications and infrastructure configurations, use technologies such as caching to improve response speed, and periodically evaluate and adjust architecture design.

- **Cost optimization**

  Cost optimization aims to improve the cost-effectiveness of cloud resources and eliminate unnecessary overhead and resource waste. It encourages enterprises to allocate resources cost-effectively by optimizing billing modes, resources, and architectures. Enterprises should establish a continuous cost operational model to keep analyzing and monitoring cloud costs and identify cost-saving opportunities. This can avoid over-allocated or idle resources and improve return on investment (ROI) and financial efficiency.

- **Operational excellence**

  Operational excellence focuses on efficiently operating and monitoring systems, continuously improving processes, and delivering business value. It emphasizes DevOps, infrastructure as code, automated deployment, testing and verification, and automated O&M tasks, as well as a setup of comprehensive monitoring, logging, and alarm reporting mechanism. With carefully designed operation processes, change management, and continuous improvement methods, enterprises can quickly respond to changes, reduce errors, improve team collaboration efficiency, and ensure the achievement of business objectives.

3.4 Landing Zone Design

3.4.1 Challenges of All-Cloud IT Governance

Large companies have complex organizational structures and dozens or even hundreds of business units (such as subsidiaries, divisions, product lines, departments, and project teams). Each business unit is responsible for building one or more application systems. The cloud transformation of these application systems will result in hundreds of service systems on the cloud and massive cloud resources. In addition, a large number of users, including enterprise employees, outsourced employees, and partners, need to access and operate these cloud resources. Risks such as resource idleness, misoperations, malicious operations, data leakage, and permission misconfiguration increase exponentially with the scale of cloud use. Large businesses must build a lean, centralized, and structured IT governance system to effectively control these risks, maximizing business benefits. CIOs and CTOs need to design a cloud IT governance system before migrating service systems to the cloud. In practice, the following challenges are often encountered:

- Business unit security and fault isolation. Businesses need to ensure that cloud resources, applications, and data are isolated between service units.
- Reduced impact scope of a single fault
- Flexible cloud resource adjustment adaptive to frequent changes in organizational and service architectures
- Network architecture across business units and controllable network connection channels
- Centralized management and control over the border network ingress and egress of multiple business units
- Production, development, and test environment planning
- Public resource sharing across multiple business units
- Centralized monitoring, O&M, and management of cloud resources across multiple business units
- Centralized budgets and costs management across business units Cloud costs optimization
- Prevention of cloud resources overuse
- User grouping User group authorization
- Compliance with enterprise, industry, and national security standards for cloud resources, data, and applications
- Mitigation of risks raised by misoperations, malicious operations, and permission misconfigurations
- Prevention of data leakage caused by user credentials losses
- Cloud migration of the original IT governance model as much as possible

To address these challenges, a comprehensive cloud IT governance solution and best practices are required to effectively manage business units, users, permissions, cloud resources, data, applications, costs, and security. Huawei Cloud uses the Landing Zone solution to address these challenges. Landing Zone is an aviation term, which refers to the area where helicopters and other aircraft can land safely. The solution is named Landing Zone because it can safely and smoothly migrate enterprise service systems to the public cloud. The purpose is to systematically address the IT governance and security compliance challenges brought in large-scale cloud migration.

3.4.2 Why Landing Zone?

To isolate faults in business units, Huawei Cloud recommends that application systems of different business units be deployed in different accounts. Huawei Cloud accounts have the following characteristics:

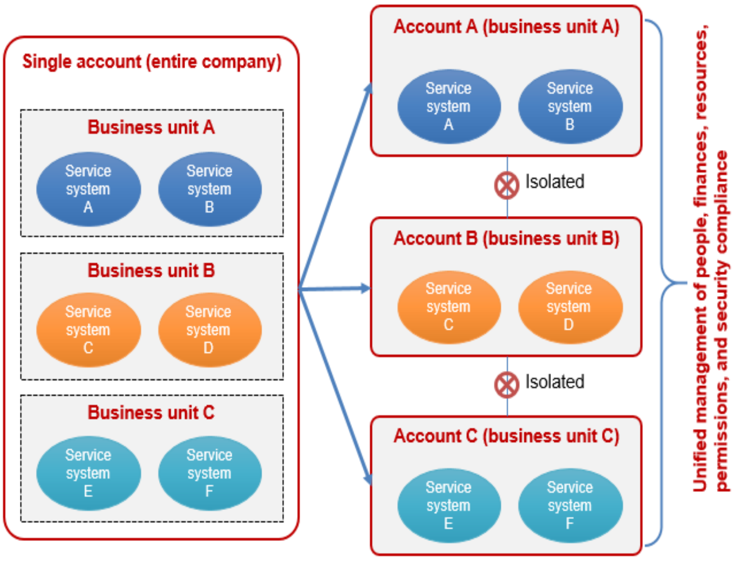
- Huawei Cloud accounts are resource containers. You can deploy cloud resources and upper-layer service application systems in an account. Different accounts are isolated from each other. Faults and security risks in one account do not affect or spread to other accounts.
- Huawei Cloud accounts are also security management boundaries. Each account has an independent identity and permissions management system. Without explicit authorization, users in an account cannot access resources, data, and applications in other accounts.
- Huawei Cloud accounts can also be used as independent billing entities. Each can be used to top up account, purchase cloud resources, settle bills, and issue invoices on Huawei Cloud.

Therefore, Huawei Cloud accounts can be used to effectively isolate faults and security risks and also to achieve efficient financial management and isolation. Using a single account to manage all resources may cause two major issues:

- A single fault in a single account will lead to the breakdown of all service systems.
- As there is a resource limit for cloud accounts, using only one account may hinder capacity expansion.

To minimize the impact of a single failure, do not deploy all service systems and cloud resources in a single account. Deploy different service systems in different accounts, as shown in the following figure.

Figure 3-5 Deployment across accounts



In the case, businesses need a multi-account architecture when migrating all services to the cloud. According to **Conway's law**, the multi-account architecture of a business is usually consistent with its organizational or service architecture. That is, accounts are divided by business unit, geographical unit, and functional unit. The multi-account architecture enables separation of duties. Different accounts are responsible for different tasks and carry different services. The administrator of each account can manage resources in the account independently. From the perspective of IT governance, no single account can be an information silo. Unified IT governance must be achieved within the company to manage identities and permissions, O&M, security, network, finances, and public resources. To meet these requirements, Huawei Cloud proposed the Landing Zone solution to help businesses build a secure, compliant, and scalable multi-account cloud environment. This solution enables resource sharing across accounts and unified management of people, finances, resources, permissions, and security compliance.

- **People**: business units, accounts, users, user groups, and roles
- **Finances**: funds, budgets, costs, invoices, and discounts
- **Resources**: cloud resources, including compute, storage, network, data, and applications
- **Permissions**: access permissions to implement the principle of least privilege (PoLP)
- **Security compliance**: compliance with the enterprise-specific, industry, and national security standards and all-round data perimeters to prevent sensitive data leakage

Landing Zone helps enterprises eliminate risks in cloud management, security, and costs during large-scale cloud migration. It helps establish a separated but unified IT governance system and a complete security compliance system to address all IT challenges.

- **Separated but unified IT governance system**: permission- and domain-specific hierarchical management as well as centralized O&M and security management
- **Complete security compliance system**: compliance with the enterprise-specific, industry, and national security standards in the cloud environment, including cloud resources, data, and applications

3.4.3 Landing Zone Design Principles

Huawei Cloud has summarized the following principles based on its practices and successful delivery of many Landing Zone projects. You can use these principles as a starting point to develop design principles that meet your enterprise requirements.

- **Conway's law**: According to **Conway's law**, the technical architecture of a system reflects the architecture of the organization that owns it. The organizational unit (OU) and account architecture of Landing Zone should be consistent with that of the company. It is recommended that the OU and account system of Landing Zone be planned based on the service architecture, geographical area structure, and IT function of the company.
- **Correlation**: The mapping should only cover the OUs, such as departments and branches that manage IT systems and users of IT resources. For example, there is no need to create an organization that maps to the administrative department if they do not manage, view, or operate any IT resources on the

cloud. Also, there is no need to create a user with financial management permissions for financial personnel who is not responsible for cost accounting, analysis, and budget management of IT systems.

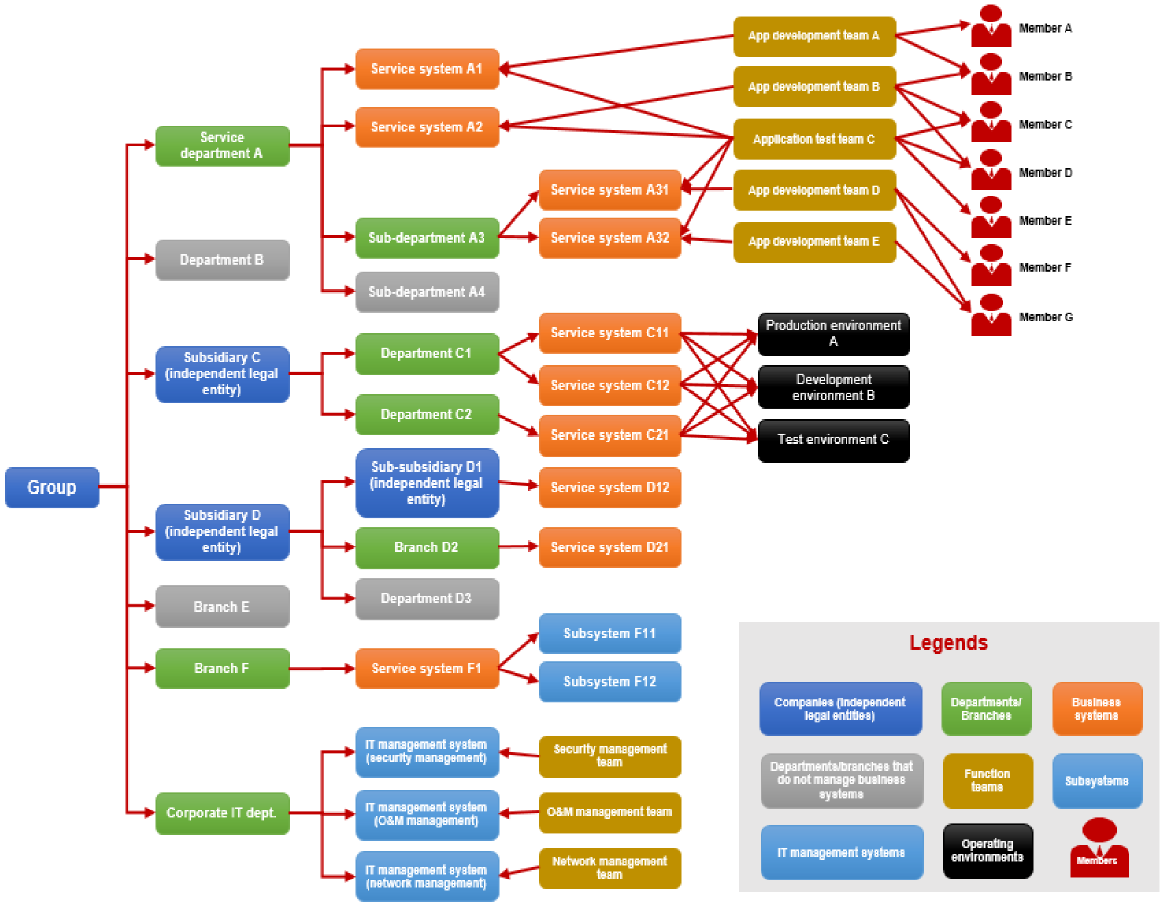
- **OU design**: Accounts that require the same control policies (including **SCPs** and **tag policies**) should be placed in the same OU. Control policies can be applied to that OU, and the policies will be inherited by each member account and lower-level OUs under that OU.
- **Operating environment isolation**: The production environment must be stable, reliable, and secure, while the development and test environments emphasize flexibility. The production environment must be isolated from the development and test environments. Stricter control policies should be used for the production environment, and looser control policies for the development and test environments.
- **Service account design**: For service departments, member accounts should be created based on the business units (such as subsidiaries, business units, product lines, departments, and project teams) defined by your organization.
- **IT management account design**: IT managerial member accounts should be created for IT departments based on their functions, such as security operations, O&M monitoring, network operations, and data platforms.

3.4.4 Landing Zone Reference Architecture

3.4.4.1 Enterprise IT Governance Architecture

Large companies have a wide range of businesses in different industries and regions. To support the long-term stable operations and effective management of the entire company, they usually adopt a group-based and hierarchical management model. As the business scope and scale continue to expand, subsidiaries and branches need to be established continuously. Subsidiaries establish their own subsidiaries, and large departments are gradually split into multiple small departments, leading to more and more organizational levels. The IT governance architecture of large companies is also affected by the organizational structure. The following figure shows a typical IT governance architecture of large businesses. (The figure does not list all levels and diagram elements.) The Landing Zone architecture described in this document is based on the IT governance architecture shown in the following figure. This architecture can be mapped to Huawei Cloud and run properly.

Figure 3-6 IT governance architecture of large businesses



In the preceding IT governance architecture, the meanings of each level are as follows:

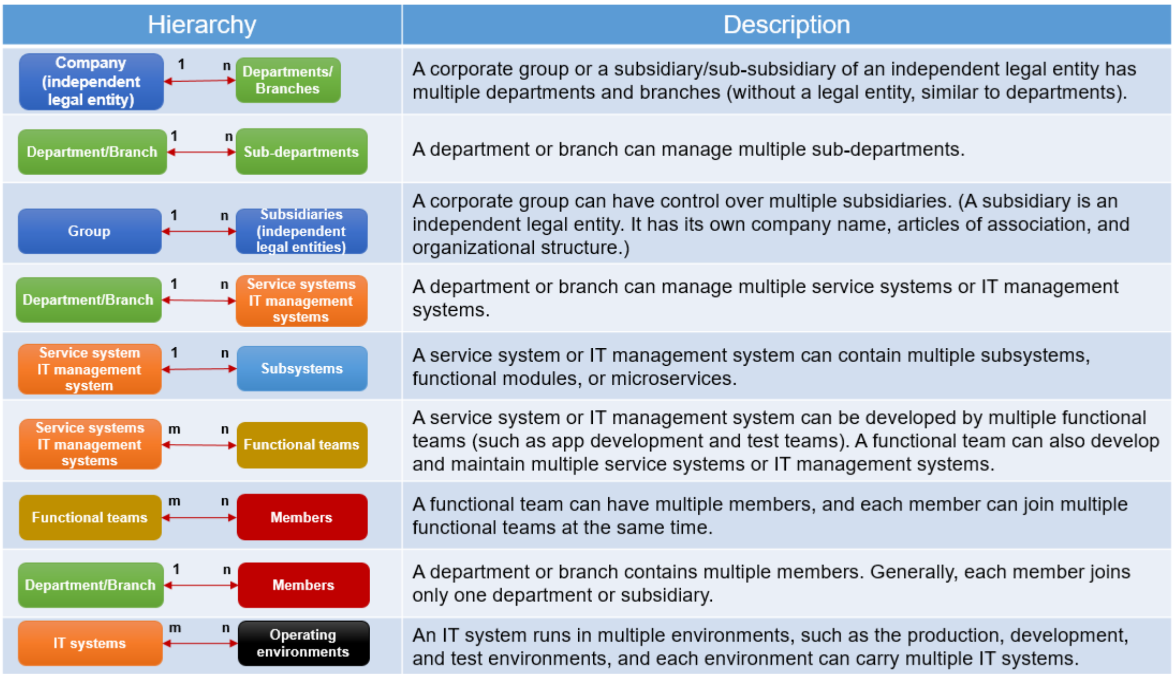
- **Group**: a corporate legal entity comprising a parent company, its subsidiaries, and affiliated members, unified by shared capital, governed by the parent company, and bound by a collective group charter as the standard framework for operations.
- **Subsidiary**: a company in which a parent company holds a controlling share of ownership and exercises operational control. The parent company has the decision-making authority over all major matters of the subsidiary. Legally, the subsidiary remains an autonomous entity with separate legal personality, conducting business operations under its own authority. A subsidiary can establish its own subsidiaries or branches based on its operation and management requirements.
- **Branch**: a branch operating under the authority of its parent company, and functioning as an independent entity established outside the company's primary location to conduct business activities, such as regional sales offices across provinces and cities. A branch lacks the status of an independent corporate entity, with its civil liabilities being assumed by the parent company.
- **Department**: The parent company, along with its subsidiaries and branches, may establish departments tailored to operational and managerial needs, such as distinct divisions for various product lines in a software firm or functional units like R&D, production, procurement, sales, and services within

an industrial manufacturing enterprise. A large department can further divide into small departments.

- **Service system**: a software system designed to complete specific tasks or solve specific problems to support business processes and scenarios in an organization, such as ERP, CRM, and marketing management systems. The development, testing, and operating of service systems require certain resources, such as compute, storage, network, security, database, middleware, big data, and AI services. A large service system can contain multiple subsystems.
- **IT management system**: an IT support and management system established to support the long-term secure and stable operation of service systems, such as the security operations center, IAM, and monitoring and O&M system.
- **Subsystem**: a large service system or IT management system with multiple decoupled and associated subsystems, functional modules, or microservices. These subsystems collaborate with each other to implement the functions of the entire system.
- **Functional team**: Members who participate in the development and O&M of the service system or IT management system are divided into different functional teams based on their responsibilities, such as the network management team, security management team, O&M management team, and application development team.
- **Member**: a person who participates in the development and O&M of a service system or IT management system. A member can join multiple functional teams, but cannot join multiple departments.
- **Operating environment**: The service system and IT management system are usually deployed in different operating environments, such as production, development, and test environments.

The following figure shows the hierarchy of IT governance for large enterprises.

Figure 3-7 Hierarchy of IT governance for large enterprises



Hierarchy	Description
Company (independent legal entity) 1 ↔ n Departments/ Branches	A corporate group or a subsidiary/sub-subsidiary of an independent legal entity has multiple departments and branches (without a legal entity, similar to departments).
Department/Branch 1 ↔ n Sub-departments	A department or branch can manage multiple sub-departments.
Group 1 ↔ n Subsidiaries (independent legal entities)	A corporate group can have control over multiple subsidiaries. (A subsidiary is an independent legal entity. It has its own company name, articles of association, and organizational structure.)
Department/Branch 1 ↔ n Service systems IT management systems	A department or branch can manage multiple service systems or IT management systems.
Service system IT management system 1 ↔ n Subsystems	A service system or IT management system can contain multiple subsystems, functional modules, or microservices.
Service systems IT management systems m ↔ n Functional teams	A service system or IT management system can be developed by multiple functional teams (such as app development and test teams). A functional team can also develop and maintain multiple service systems or IT management systems.
Functional teams m ↔ n Members	A functional team can have multiple members, and each member can join multiple functional teams at the same time.
Department/Branch 1 ↔ n Members	A department or branch contains multiple members. Generally, each member joins only one department or subsidiary.
IT systems m ↔ n Operating environments	An IT system runs in multiple environments, such as the production, development, and test environments, and each environment can carry multiple IT systems.

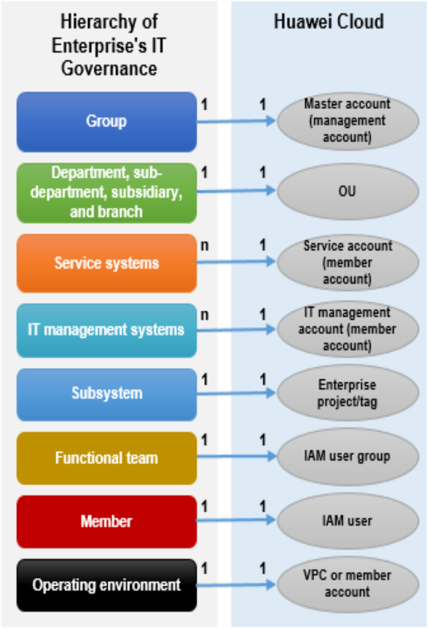
The IT governance architecture needs to map to Huawei Cloud objects. The following figure shows the recommended mappings from the perspective of lean governance. The group maps to the master account (or management account) of Huawei Cloud. Each subsidiary, branch, or department maps to an OU. One or more service systems map to a service account (member account). Generally, all service systems of a business unit map to a service account. One or more IT management systems map to an IT management account (member account).

Subsystems can map to **enterprise projects** or tags on Huawei Cloud.

Functional teams map to IAM user groups, and their members map to IAM users.

Production, development, and test environments can map to different VPCs. To strictly isolate these environments, they can also map to independent IAM users. Note that you do not need to map the subsidiaries, branches, or departments that do not develop or maintain service systems or IT management systems to Huawei Cloud.

Figure 3-8 Enterprise IT governance architecture mapping to Huawei Cloud


Hierarchy of Enterprise's IT Governance
Huawei Cloud
Group
1
1
Master account (management account)
Department, sub-department, subsidiary, and branch
1
1
OU
Service systems
n
1
Service account (member account)
IT management systems
n
1
IT management account (member account)
Subsystem
1
1
Enterprise project/tag
Functional team
1
1
IAM user group
Member
1
1
IAM user
Operating environment
1
1
VPC or member account

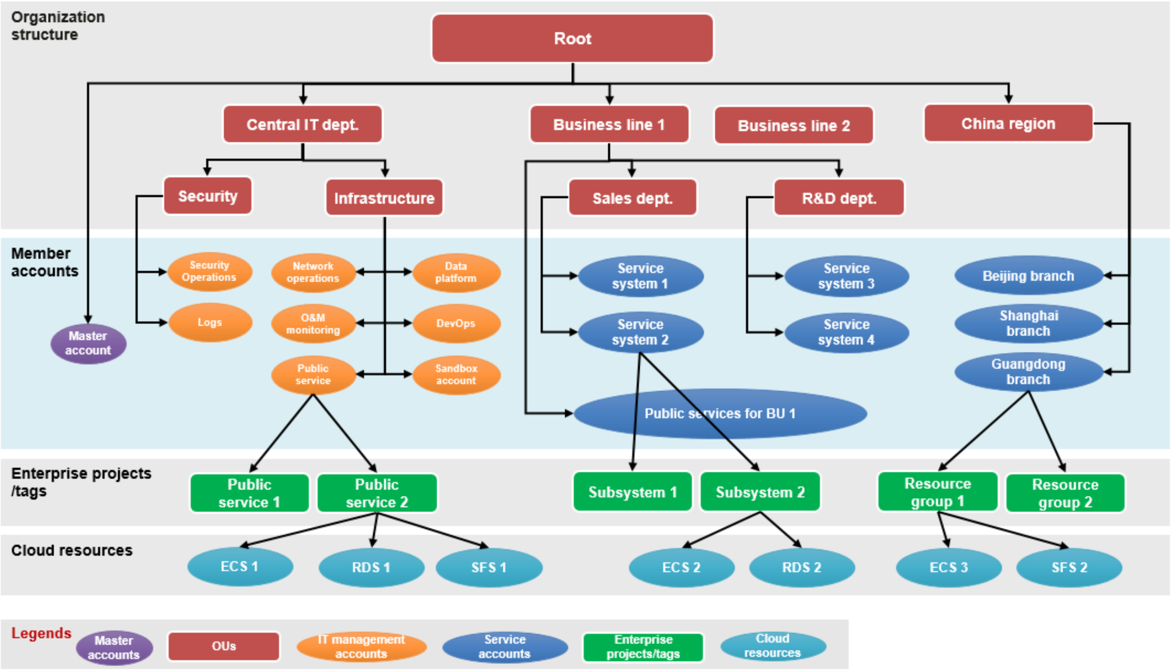
3.4.4.2 Organization and Account Design

The Landing Zone solution aims to build a secure, compliant, and scalable multi-account operating environment on the cloud. The first step is to plan the organization and account architecture. According to **Conway's law**, the account structure on Huawei Cloud should be consistent with the overall organization and business architecture of the company, but not necessarily identical. The mapping should only cover the OUs (such as subsidiaries, branches, and departments) that develop and maintain service systems or IT management systems, and the users of the IT resources.

For example, there is no need to create an OU that maps to the HR department if they do not manage, view, or operate any IT resources on the cloud. Also, there is no need to create a user with financial management permissions for financial personnel who are not responsible for cost accounting, analysis, and budget management of IT systems.

Huawei Cloud provides the following OUs and accounts in the architecture for your reference. You need to design the hierarchical structure of OUs and account groups on Huawei Cloud based on the business architecture, geographical area structure, and IT functions.

Figure 3-9 Reference architecture for OUs and accounts

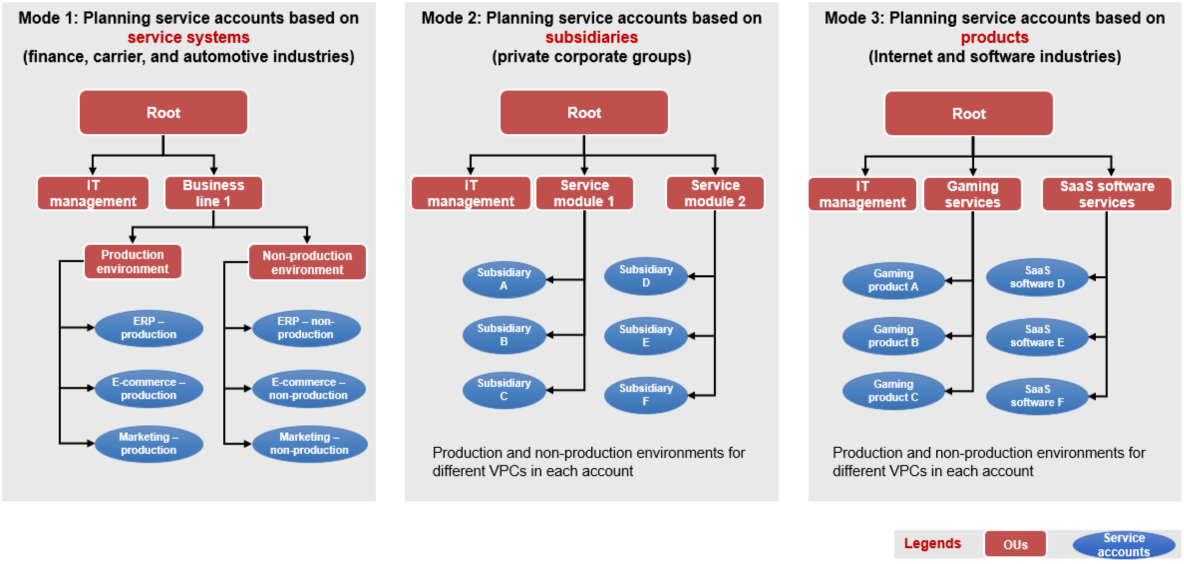


Service Account Planning

You can create OUs at different levels based on the business architecture. Each service OU can contain multiple service accounts, and each service account can be used for one or more service systems.

In principle, service accounts must be consistent with the business units defined by the organization. Business units can be subsidiaries, divisions, product lines, departments, project teams, or service systems. Finer granularity of business units matches higher-level lean governance requirements. Thanks to Huawei Cloud's experience in designing and implementing the Landing Zone solution for a large number of enterprise and government customers, we provided the following typical service account plans for your reference.

Figure 3-10 Service account planning


Mode 1: Planning service accounts based on service systems (finance, carrier, and automotive industries)
Root
IT management
Business line 1
Production environment
Non-production environment
ERP – production
E-commerce – production
Marketing – production
ERP – non-production
E-commerce – non-production
Marketing – non-production
Mode 2: Planning service accounts based on subsidiaries (private corporate groups)
Root
IT management
Service module 1
Service module 2
Subsidiary A
Subsidiary B
Subsidiary C
Subsidiary D
Subsidiary E
Subsidiary F
Production and non-production environments for different VPCs in each account
Mode 3: Planning service accounts based on products (Internet and software industries)
Root
IT management
Gaming services
SaaS software services
Gaming product A
Gaming product B
Gaming product C
SaaS software D
SaaS software E
SaaS software F
Production and non-production environments for different VPCs in each account
Legends
OUs
Service accounts

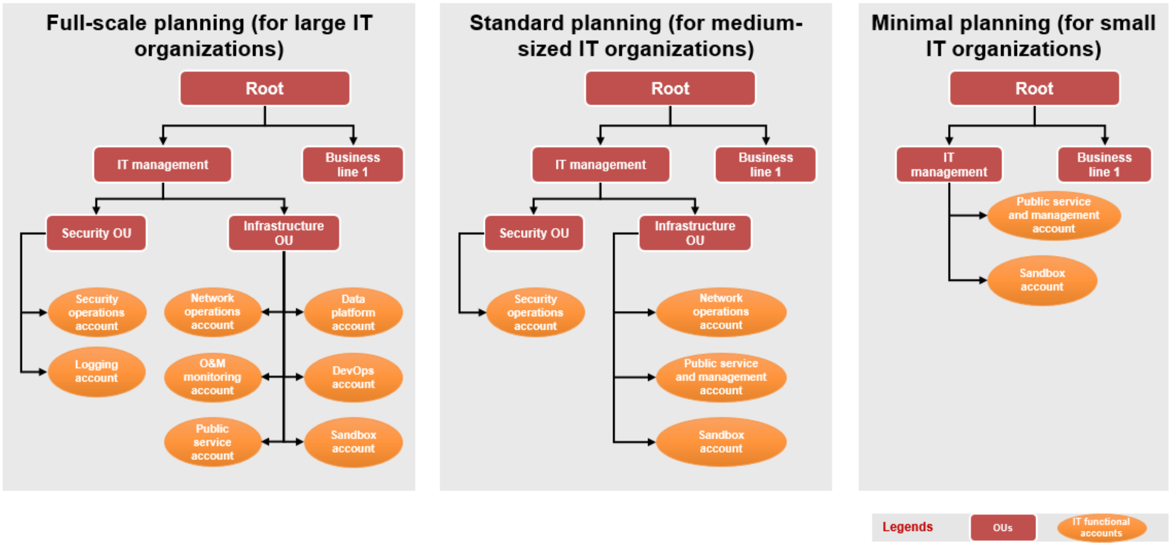
- In mode 1, service accounts are planned based on the service systems of each business unit. Service systems are isolated by member accounts. Member accounts are also used for isolating production and non-production environments. This mode is suitable for companies demanding strict control, such as financial institutions, carriers, and automakers.
- In mode 2, service accounts are planned based on subsidiaries. Each is an isolated service unit. Service systems of each subsidiary are deployed in the member accounts created for it. Within a member account, you can plan VPCs for production and non-production environments of each service system. This mode is suitable for private enterprise groups. They usually have dozens or even hundreds of subsidiaries. From the perspective of corporate governance, strict business isolation is required between subsidiaries, but service systems do not need to be isolated.
- In mode 3, service accounts are planned based on product planning. It breaks down business units into products. Products here refer to software systems that provide services for customers, such as live streaming systems, online games, and SaaS offerings. Different products are deployed and isolated in different member accounts. Within a member account, you can plan VPCs for production and non-production environments of each product. This mode is especially suitable for Internet and SaaS enterprises.

IT Management Account Planning

An OU is required for the central IT department of a company, and two lower-level OUs are required for security management and infrastructure management, respectively. IT management accounts are used to centrally manage all member accounts of the company.

In principle, the planning of IT management accounts must be consistent with the responsibilities and scale of the IT departments to maintain the benefits of each IT group. Thanks to Huawei Cloud's experience in designing and implementing the Landing Zone solution for a large number of enterprise and government customers, we provided the following typical IT management account plans for your reference.

Figure 3-11 IT management account planning


Full-scale planning (for large IT organizations)
Root
IT management
Business line 1
Security OU
Infrastructure OU
Security operations account
Logging account
Network operations account
O&M monitoring account
Public service account
Data platform account
DevOps account
Sandbox account
Standard planning (for medium-sized IT organizations)
Root
IT management
Business line 1
Security OU
Infrastructure OU
Security operations account
Network operations account
Public service and management account
Sandbox account
Minimal planning (for small IT organizations)
Root
IT management
Business line 1
Public service and management account
Sandbox account
Legends
OUs
IT functional accounts

- The full-scale planning mode is suitable for large IT organizations. These IT organizations usually have dozens or hundreds of IT managers. Independent IT functional teams have been divided based on responsibilities such as security operations, log audit, network operations, O&M monitoring, public services, data platforms, and DevOps. To ensure separation of duties (SOD) within the IT department, map these IT functional teams to independent accounts. These accounts are used to fulfill independent IT responsibilities, so they are called IT management accounts (or functional accounts), as shown in the following table.

Table 3-8 IT management accounts

Account	IT Function	Responsible Team	Recommended Cloud Service
Master account or management account	Centrally manage organizations and accounts, finances, governance policies, and identities and permissions.	CIO or IT director	Organizations, Resource Governance Center (RGC), Cost Center, and IAM Identity Center

Account	IT Function	Responsible Team	Recommended Cloud Service
Network operations account	Centrally deploy and manage enterprise network resources (including resources used for protecting network boundaries), and ensure VPC connectivity in a multi-account environment. In particular, manage ingress and egress in a unified manner for Internet and on-premises IDCs.	Network management team	Enterprise Router, Domain Name Service (DNS), NAT Gateway, Elastic IP (EIP), Virtual Private Cloud (VPC), Direct Connect, Cloud Connect, Virtual Private Network (VPN), Cloud Firewall (CFW), Web Application Firewall (WAF), and Anti-DDoS Service (AAD)
Public service account	Centrally deploy and manage the enterprise public resources, services, and application systems, and share them with other member accounts of the enterprise.	Public service management team	Image Management Service (IMS), SoftWare Repository for Container (SWR), Scalable File Service (SFS), Object Storage Service (OBS), in-house NTP servers, and in-house Anti-DDoS servers
Security operations account	Centrally manage and control security policies, rules, and resources in all accounts of the company, set security configuration baselines for member accounts of the company, and be responsible for information security of the entire company.	Security management team	Deploy services that support cross-account security management and control, such as SecMaster, Host Security Service (HSS), Data Security Center (DSC), Data Encryption Workshop (DEW), Cloud Certificate Manager (CCM), and CodeArts Inspector.
O&M monitoring account	Centrally monitor and maintain resources and applications under each member account, manage alarms, handle events, manage changes, and provide O&M security assurance measures	O&M team	Application Operations Management (AOM), Cloud Operations Center (COC), Log Tank Service (LTS), Application Performance Management (APM), and Cloud Bastion Host (CBH)

Account	IT Function	Responsible Team	Recommended Cloud Service
Logging account	Centrally store and view audit logs and security-related logs (such as VPC flow logs and OBS access logs) of all accounts.	Compliance audit team	Cloud Trace Service (CTS), Log Tank Service (LTS), Config, and Object Storage Service (OBS)
Data platform	Centrally deploy big data platforms and collect service data from other accounts to store, process, and analyze them on the data platforms.	Data processing team	Data lake, big data analysis platform, Data Ingestion Service (DIS), and DataArts Studio
Sandbox account	Test functions and governance policies of cloud services.	Test team	Resources and services to be tested

You can create more member accounts such as application integration accounts as needed.

- The standard planning mode is suitable for medium-sized IT organizations. These IT organizations usually have only 10 to 30 IT management engineers. They do not divide IT responsibilities as detailed as large IT organizations. However, they usually divide IT functions into independent security operations, network operations, and O&M monitoring teams. You are advised to create independent IT management accounts for security operations, network operations, public services, and O&M management, respectively. The security operations account in the standard planning mode combines the functions of the security operations account and logging account in the full-scale planning mode. In the standard planning mode, the public service and management accounts combine the functions of the O&M monitoring account, public service account, data platform account, and DevOps account in the full-scale planning mode.
- The minimal planning mode is suitable for small IT organizations with only a few IT management engineers. In this case, only a public service and management account needs to be created. This account combines the functions of the security operations account, logging account, network operations account, O&M monitoring account, public service account, data platform account, and DevOps account in the full-scale planning mode.

Region-based Account Planning

The planning of accounts for different regions is simple. You can divide different levels of OUs on Huawei Cloud based on the geographical area structure. You can create member accounts under each OU for countries or regions, and deploy local customer relationship management systems, customer service systems, and operations management systems under the member accounts. In the preceding reference architecture, the organization in the China region is mapped to an OU on Huawei Cloud, and independent member accounts are created for branches such as Beijing and Shanghai to carry local application systems.

Master Account Planning

A master account (also called management account) is created under the root of the organization by default. It is recommended that no cloud resources be deployed in this master account. You can use this master account to do the following:

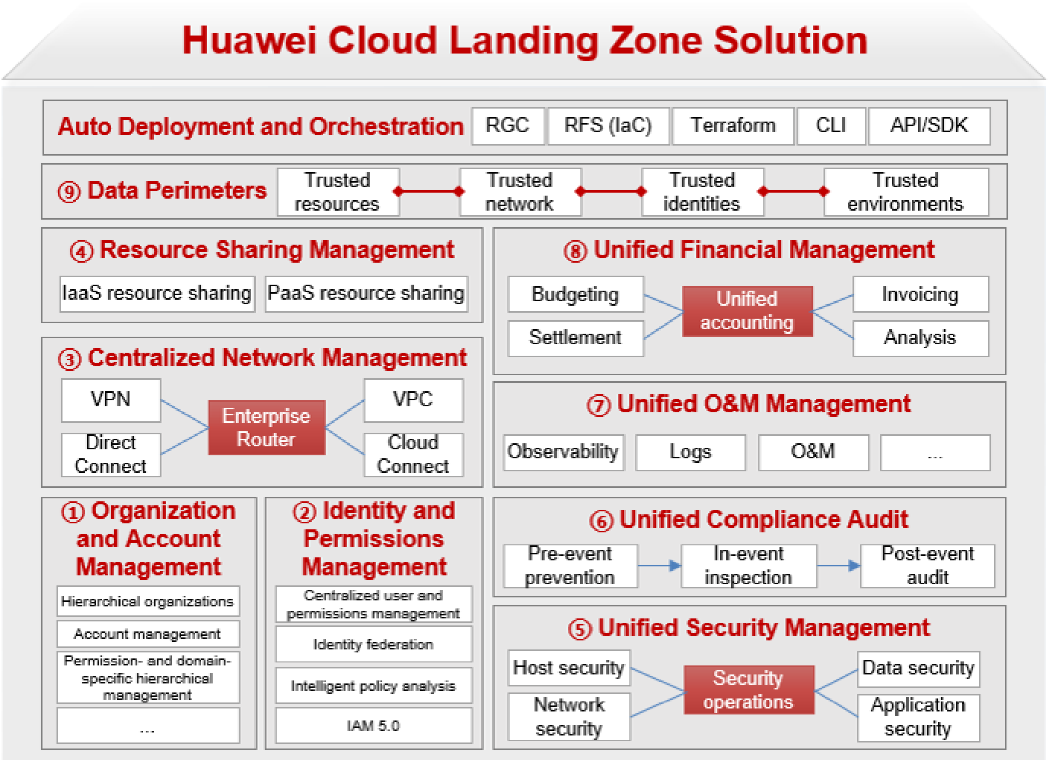
1. **Organization and account management**: Create and manage the organization structure and OUs, create member accounts for OUs, or invite existing accounts as the member accounts of OUs.
2. **Financial management**: Centrally manage all accounts of a company, including unified budget management, cost analysis, bill management, fund management, coupon application, cost settlement, and invoicing.
3. **Policy management**: Configure policies (including **SCPs** and **tag policies**) for OUs and member accounts, forcibly restrict user permissions (also for account administrators) under member accounts to mitigate security risks caused by excessive permissions. If you apply a control policy to a specific OU, the policy will apply to all member accounts and lower-level OUs in that OU.
4. **Identity management**: Create users and user groups based on IAM Identity Center, or configure identity federation with an external identity provider (IdP). Then, configure permissions for these users to access cloud resources in multiple accounts based on the PoLP.

You can also use enterprise projects or tags to logically group resources under each member account. For example, you can map a subsystem of an application system to an enterprise project or tag on Huawei Cloud. You can also allocate costs and grant fine-grained permissions based on enterprise projects or tags.

3.4.4.3 Overall Architecture

We provide the following Landing Zone reference architecture based on Huawei Cloud's practices and extensive delivery experience. This architecture involves nine domains: organization and account management, identity and permissions management, centralized network management, resource sharing management, unified security management, unified compliance audit, unified O&M management, unified financial management, and data perimeters.

Figure 3-12 Landing Zone reference architecture



The resources of the nine domains are managed by specific accounts. For example, organization and account management is implemented in the master account (management account), and centralized network management is implemented in the network operations account. The following table lists the accounts for these domains.

Table 3-9 Domains and corresponding accounts

Domain	Account
Organization and account management	Master account (management account)
Identity and permissions management	Master account (management account)
Centralized network management	Network operations account
Resource sharing management	Public service account
Unified security management	Security operations account

Domain	Account
Unified compliance audit	Security operations account and logging account
Unified O&M management	O&M monitoring account
Unified financial management	Master account (management account)
Data perimeters	Master account (management account) and sandbox account (used to test various control policies)

The previous sections detailed the design of organizations and accounts. The following sections will describe the designs of the other eight domains.

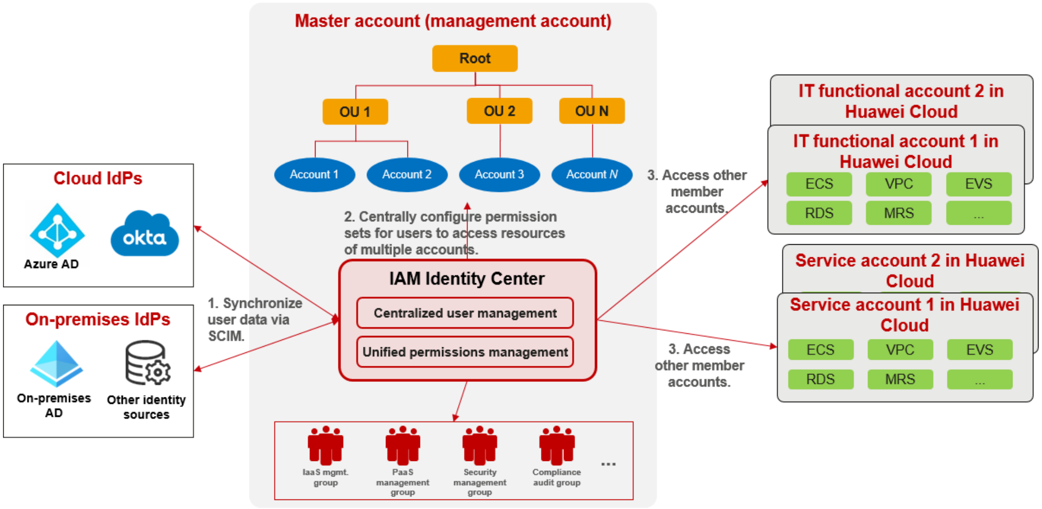
3.4.4.4 Identity and Permissions Design

Thanks to Huawei Cloud's extensive delivery experience, we provide the following best practices in user and permissions management.

Unified Identity and Permissions Management

The identity management system of your company is already optimal in controlling the permissions of employees as they are recruited and revoking permissions from employees who have transferred to different departments or have resigned. You are advised to use your own identity management system to implement federated identity authentication with Huawei Cloud IAM Identity Center, and synchronize users from your identity management system to IAM Identity Center based on the System for Cross-domain Identity Management (SCIM). In IAM Identity Center, you can centrally configure permissions for these users to access resources in multiple accounts. Then, the users can log in to the Huawei Cloud console using Single Sign-on (SSO), view the accounts they can access, and click **Access Console** to access the cloud resources in these accounts.

Figure 3-13 Unified identity and permissions management


Master account (management account)
Root
OU 1
OU 2
OU N
Account 1
Account 2
Account 3
Account N
Cloud IdPs
Azure AD
okta
On-premises IdPs
On-premises AD
Other identity sources
1. Synchronize user data via SCIM.
2. Centrally configure permission sets for users to access resources of multiple accounts.
IAM Identity Center
Centralized user management
Unified permissions management
IaaS mgmt. group
PaaS management group
Security management group
Compliance audit group
...
3. Access other member accounts.
IT functional account 2 in Huawei Cloud
IT functional account 1 in Huawei Cloud
ECS
VPC
EVS
RDS
MRS
...
Service account 2 in Huawei Cloud
Service account 1 in Huawei Cloud
ECS
VPC
EVS
RDS
MRS
...
3. Access other member accounts.

User Group and Permissions Plan

You can plan IAM Identity Center user groups based on the role division of cloud center of excellence (CCoE) and add employees to the user groups that match their responsibilities. The following table lists the recommended user groups, along with their responsibilities, accounts, and permissions granted based on the principle of least privilege (PoLP). You can refer to this table to plan user groups and permissions that meet enterprise requirements.

Table 3-10 IAM Identity Center user groups

User Group	Group Responsibility	Recommended Permissions for Multiple Accounts
Financial management group	Manage financial elements such as bills, costs, discounts, and invoices of member accounts in a unified manner.	BSS Administrator and BSS Finance permissions for the management account
IT governance group	Create and manage OUs, member accounts, and SCPs.	Organizations FullAccess permission for the management account
Identity and permissions management group	Centrally create and manage users and user groups, and configure permissions, agencies, and SSO.	● IdentityCenter FullAccess and Security Administrator permissions for the management account ● Security Administrator for all other accounts
Security management group	Centrally manage and control security policies, rules, and resources for the entire company, and set security configuration baselines for other accounts of the company.	Management permissions of security resources for all accounts, such as SecMaster, Host Security Service (HSS), Data Security Center (DSC), and Database Security Service (DBSS)
Compliance audit group	Centrally view audit logs and security-related logs (such as VPC flow logs and OBS access logs) of all accounts.	● Tenant Administrator permission for the logging account ● Tenant Guest for all other accounts

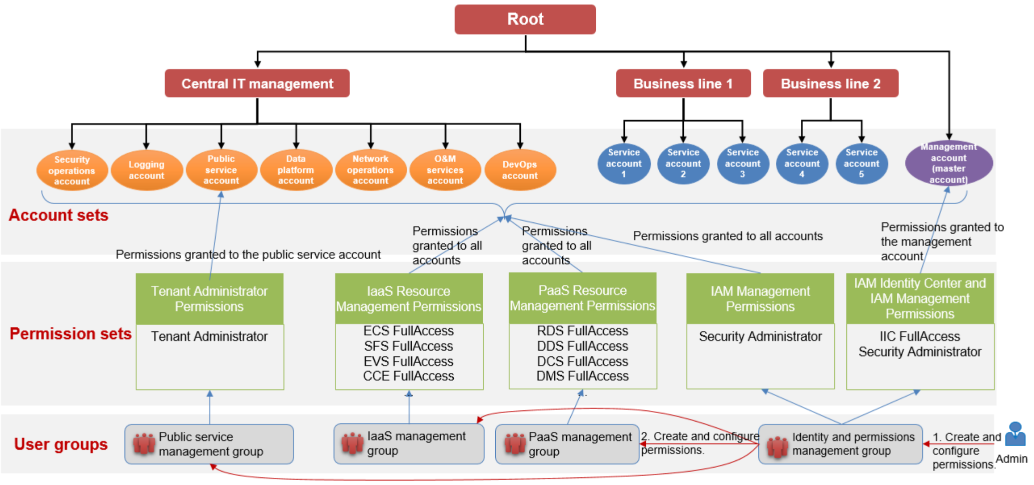
User Group	Group Responsibility	Recommended Permissions for Multiple Accounts
Network management group	Centrally deploy and manage enterprise network connection resources, such as Enterprise Router, Virtual Private Network (VPN), Direct Connect (DC), and NAT Gateway. Centrally create and manage VPCs, subnets, and network access control lists (ACLs) for each account. Centrally deploy and manage network boundary protection resources, such as Web Application Firewall (WAF) and Cloud Firewall (CFW).	● Tenant Administrator permission for the network operations account ● Management permissions of network resources, such as VPCs, subnets, and network ACLs, for all other accounts ● Management permissions of network security resources, such as Web Application Firewall (WAF) and Cloud Firewall (CFW), for network operations account
IaaS management group	Centrally manage IaaS resources of all accounts as the cloud infrastructure administrator.	● Management permissions of IaaS resources for all accounts ● Management permissions of O&M monitoring services, such as Application Operations Management (AOM), Cloud Eye, and Application Performance Management (APM), for all other accounts
PaaS management group	Centrally manage middleware resources of all accounts as the middleware administrator.	● Management permissions of middleware resources for all accounts ● Management permissions of O&M monitoring services, such as Application Operations Management (AOM), Cloud Eye, and Application Performance Management (APM), for all other accounts
Automated O&M group	Centrally monitor and maintain resources of all accounts.	● Tenant Administrator permission for the O&M monitoring account ● Management permissions of Cloud Operations Center (COC) for all other accounts
Data management group	Centrally deploy and manage data platforms and collect service data from other member accounts to store, process, and analyze them on the data platforms.	Tenant Administrator permission for the data platform account

User Group	Group Responsibility	Recommended Permissions for Multiple Accounts
Public service management group	Centrally deploy and manage the enterprise public resources, services, and application systems, and share them with other member accounts of the company.	Tenant Administrator permission for the public service account
Application development group	Develop applications and manage the development environment.	● Tenant Administrator permission for the development account ● Developer permissions for the DevOps account
Application test group	Test applications and manage the test environment.	● Tenant Administrator permission for the test account ● Tester permissions for the DevOps account

Permissions Configuration

The root user or Admin user of the master account is the administrator with the highest permissions. It is recommended that the CIO or IT director of the company keep the password of the root user or Admin user. The Admin user should not perform routine management and O&M operations, including creating users and configuring permissions. You are advised to use the master account's Admin user to create an identity permissions management group and its users in IAM Identity Center, and grant the required permissions to the group. Then, the users in this group can create other users and user groups and grant permissions to them. The following figure shows the recommended user group and permission configuration.

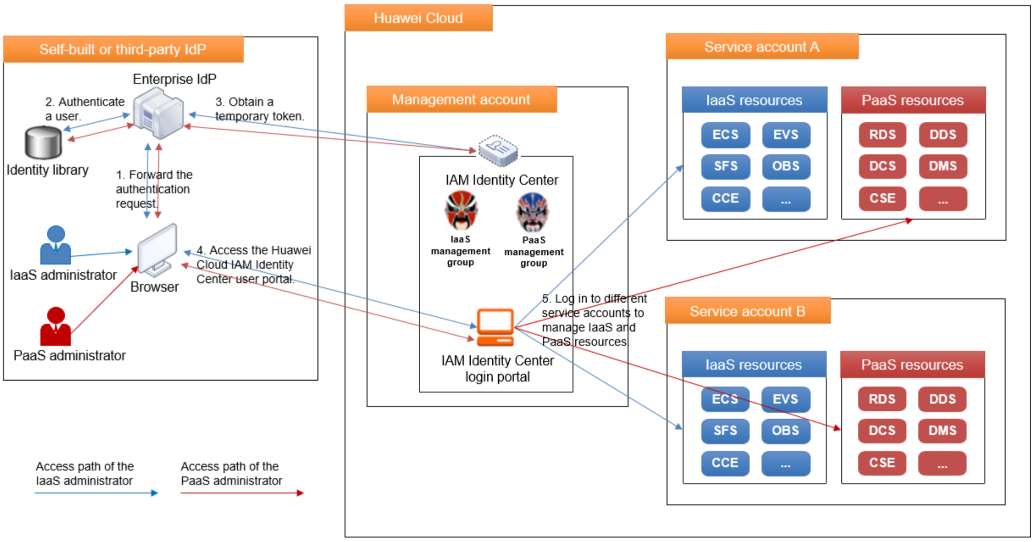
Figure 3-14 Recommended permission configuration


Root
Central IT management
Business line 1
Business line 2
Security operations account
Logging account
Public service account
Data platform account
Network operations account
O&M services account
DevOps account
Service account 1
Service account 2
Service account 3
Service account 4
Service account 5
Management account (master account)
Account sets
Permissions granted to the public service account
Permissions granted to all accounts
Permissions granted to all accounts
Permissions granted to all accounts
Permissions granted to the management account
Permission sets
Tenant Administrator Permissions
Tenant Administrator
IaaS Resource Management Permissions
ECS FullAccess
SFS FullAccess
EVS FullAccess
CCE FullAccess
PaaS Resource Management Permissions
RDS FullAccess
DDS FullAccess
DCS FullAccess
DMS FullAccess
IAM Management Permissions
Security Administrator
IAM Identity Center and IAM Management Permissions
IIC FullAccess
Security Administrator
User groups
Public service management group
IaaS management group
PaaS management group
2. Create and configure permissions.
Identity and permissions management group
1. Create and configure permissions.
Admin

Permissions Use

For unified management, IT management personnel need to use the multi-account authorization method of IAM Identity Center to access and manage cloud resources of other accounts. For example, if the IaaS and PaaS administrators need to centrally manage IaaS and middleware resources of each account in a company, they can use the multi-account authorization method of IAM Identity Center to access these resources, as shown in the following figure.

Figure 3-15 Unified multi-account management of IaaS and middleware resources by IaaS and PaaS administrators



Other Best Practices for Identity and Permissions Management

- Grant groups only the permissions essentially required to perform specific tasks by following the PoLP. If the responsibilities of a group were changed, adjust the granted permissions immediately.
- To simplify authorization, grant permissions to user groups instead of users.
- Ensure that permission configuration, permission use, and permission audit are performed by different natural persons or teams who are not part of the community of shared interests.
- Perform secondary authentication for sensitive operations such as deleting and updating key resources and using a large amount of money.
- Do not share your password. Instead, create a user for each person who needs to manage or use Huawei Cloud resources, and assign permissions to that user. In this way, all operations performed on Huawei Cloud can be tracked and audited.
- The enterprise CTO or CIO keeps the password of the root user (with the same name as the account) of the master account. The owner of the business unit to which the member account belongs keeps the password of the member account's root user.
- The root user (with the same name as the account) has high permissions. Do not use the root user to access Huawei Cloud. Instead, create one or more common users and grant permissions to them by following the PoLP, and then use these users to perform routine management.

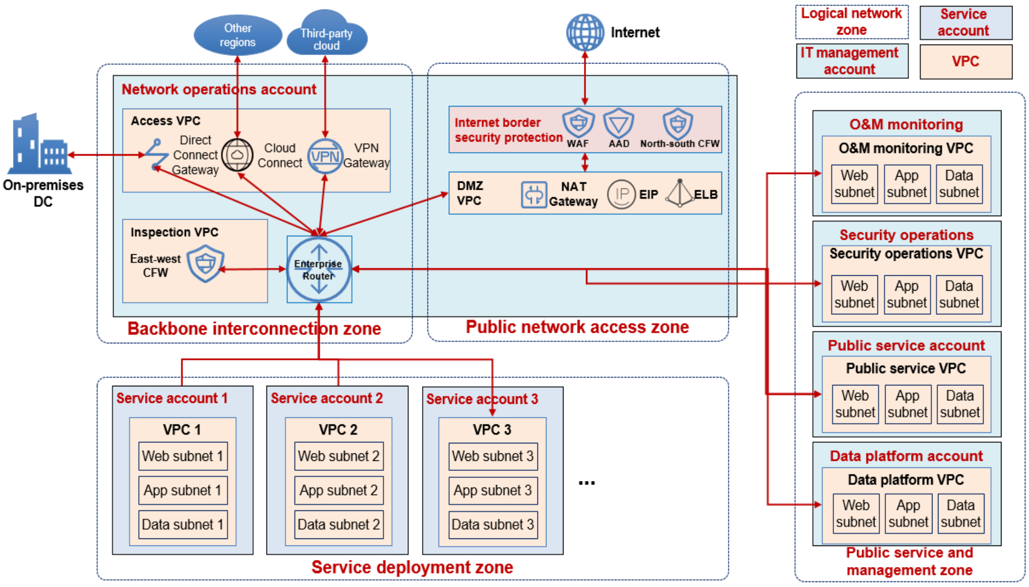
- Use service control policies (SCPs) to control the permissions of the root user of the member account.

3.4.4.5 Overall Network Architecture

The following figure shows the overall network architecture of Landing Zone. The network operations account functions as the network hub of the Landing Zone to centrally manage the border network ingress and egress and enable connections between VPCs across multiple accounts. Enterprise routers are deployed in the network operations account to interwork with VPCs of service accounts. This allows multiple accounts to communicate with the on-premises IDC through Direct Connect and VPN, and communicate with the Internet through the public NAT gateway. They can also communicate with accounts in other regions through Cloud Connect.

You can manage network resources in this account to reduce the management workloads and develop and implement unified network security policies. For example, you can deploy internet-oriented Advanced Anti-DDoS (AAD), CFW, and WAF services and configure specific security policies.

Figure 3-16 Overall network architecture



The Landing Zone network architecture consists of four logical network zones: backbone interconnection zone, public network access zone, service deployment zone, and public service and management zone.

Main Functions of the Backbone Interconnection Zone

- Centrally deploy enterprise routers to build network hubs for cloud and on-premises interconnection, multi-account and multi-VPC interconnection on the cloud, and cross-region interconnection on the cloud.
- Centrally deploy VPN or Direct Connect to interconnect the cloud with local data centers for all accounts.
- Centrally deploy Cloud Connect to interconnect with other regions on Huawei Cloud for all accounts.

- Centrally deploy VPN to interconnect with third-party clouds for all accounts.

Main Functions of the Public Network Access Zone

- Centrally set up demilitarized zones (DMZs) and deploy and maintain resources such as NAT Gateway, Elastic IP (EIP), proxy servers, and Elastic Load Balance (ELB) to provide Internet access capabilities for other accounts.
- Deploy security services such as WAF, CFW, and Anti-DDoS to protect Internet connection resources.
- Expose explicit IP addresses and ports, prohibit other ports, and terminate public networks.

Main Functions of the Service Deployment Zone

- Create VPCs and subnets and deploy cloud resources required by service systems.
- Divide VPCs for different operating environments, such as production, development, and test environments.
- Divide subnets for the layers of the application architecture, including web, application, and data subnets.

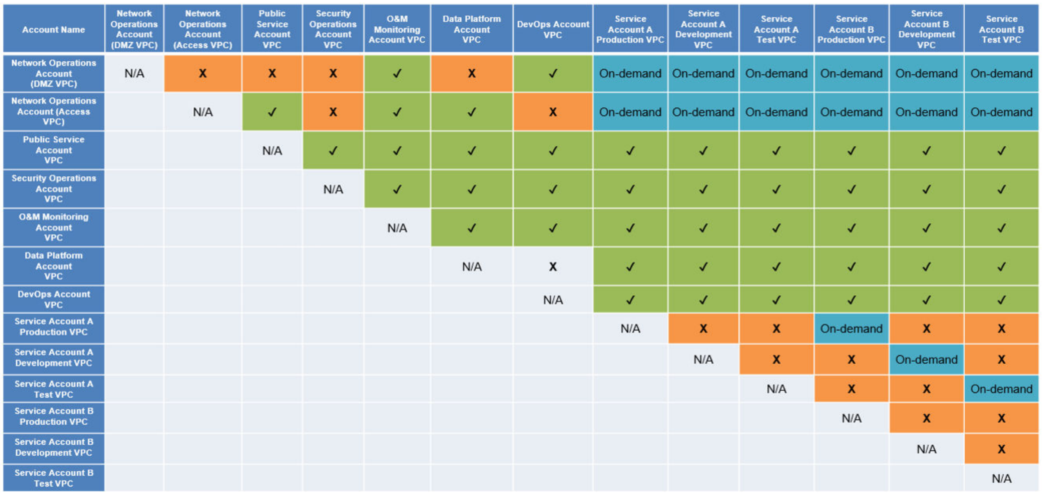
Main Functions of the Public Service and Management Zone

- Create VPCs and subnets and deploy cloud resources required by the public services and IT management systems. Public services include Active Directory (AD), Domain Name System (DNS), file systems, Object Storage Service (OBS) buckets, and data platforms. IT management systems include O&M and security management systems.
- Divide VPCs for different operating environments, such as production, development, and test environments.
- Divide subnets for the layers of the application architecture, including web, application, and data subnets.

The core of this network architecture is the network operations account, which serves as the network hub for connecting other accounts. Other accounts must communicate with each other through the enterprise router of this account. Huawei Cloud organizes the connectivity matrix between VPCs in each account according to the following assumptions and account responsibilities. With this matrix, you can then configure routing rules on the enterprise router to control communication between VPCs.

- The O&M monitoring account needs to maintain resources in third-party clouds and local DCs.
- The security operations account needs to obtain system patch packages from the public network.
- The data platform needs to obtain data from third-party clouds and local DCs.
- The DevOps account needs to download code from GitHub and deploy software artifacts to each service account.
- The public service account needs to connect to the local IDCs.
- The production, development, and test environments must be isolated from each other.

Figure 3-17 Connectivity matrix of the VPC network of each account

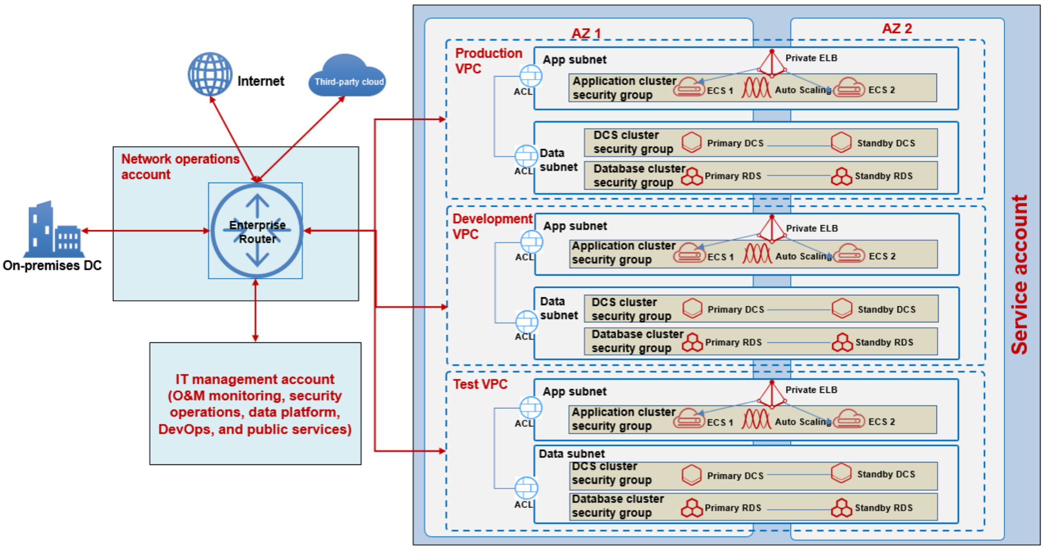


Account Name	Network Operations Account (DMZ VPC)	Network Operations Account (Access VPC)	Public Service Account VPC	Security Operations Account VPC	O&M Monitoring Account VPC	Data Platform Account VPC	DevOps Account VPC	Service Account A Production VPC	Service Account A Development VPC	Service Account A Test VPC	Service Account B Production VPC	Service Account B Development VPC	Service Account B Test VPC
Network Operations Account (DMZ VPC)	N/A	X	X	X	✓	X	✓	On-demand	On-demand	On-demand	On-demand	On-demand	On-demand
Network Operations Account (Access VPC)		N/A	✓	X	✓	✓	X	On-demand	On-demand	On-demand	On-demand	On-demand	On-demand
Public Service Account VPC			N/A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Security Operations Account VPC				N/A	✓	✓	✓	✓	✓	✓	✓	✓	✓
O&M Monitoring Account VPC					N/A	✓	✓	✓	✓	✓	✓	✓	✓
Data Platform Account VPC						N/A	X	✓	✓	✓	✓	✓	✓
DevOps Account VPC							N/A	✓	✓	✓	✓	✓	✓
Service Account A Production VPC								N/A	X	X	On-demand	X	X
Service Account A Development VPC									N/A	X	X	On-demand	X
Service Account A Test VPC										N/A	X	X	On-demand
Service Account B Production VPC											N/A	X	X
Service Account B Development VPC												N/A	X
Service Account B Test VPC													N/A

The logging account is used to centrally store audit logs and run logs through Huawei Cloud Cloud Trace Service (CTS), Log Tank Service (LTS), and OBS. These services do not have IP addresses for the tenant plane, so you do not need to interconnect this account with VPCs of other accounts. A sandbox account is used to test Huawei Cloud resources and control policies, including VPC functions and connectivity with other accounts. Therefore, you do not need to preconfigure the connection between this account and other accounts in enterprise router.

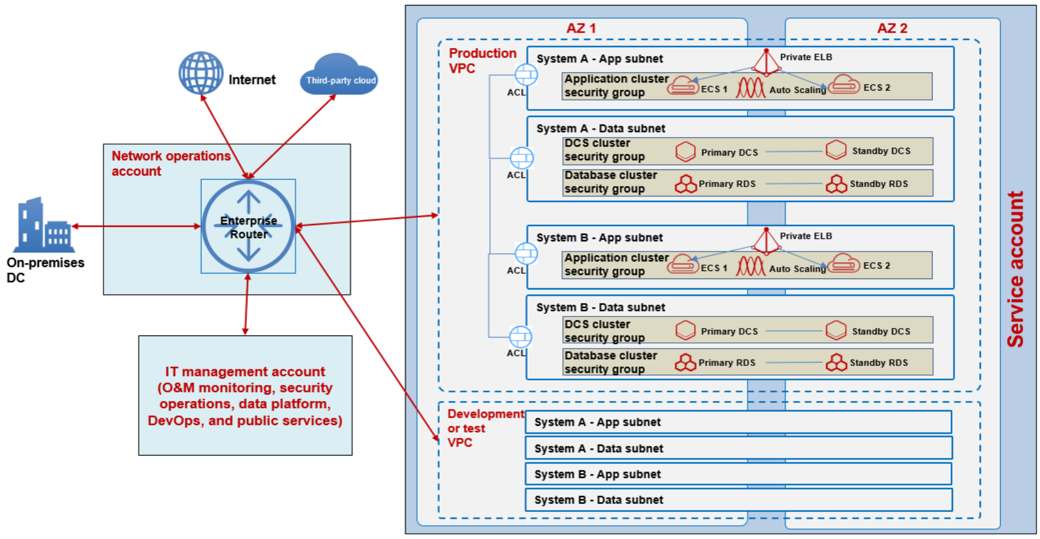
If a service system has an independent account and the production and non-production environments do not need to be strictly isolated, you are advised to create separate VPCs (production VPC, development VPC, and test VPC) for different operating environments of the service system. You need to deploy at least two subnets for each VPC: application subnet for the application layer and data subnet for the data layer. You can use network ACLs to control access between subnets. You can also add cloud servers and Relational Database Service (RDS) resources to security groups and configure security group rules for instance-level access control. You can deploy application server clusters across availability zones (AZs) to achieve high availability for applications, and use Huawei Cloud active/standby database clusters and cache clusters across AZs to achieve high availability for data use. The following figure shows an example.

Figure 3-18 Network planning in a service account: one member account for one service system



If multiple service systems use the same member account, you are advised to create three independent VPCs which are isolated from each other in the account: production VPC, development VPC, and test VPC. These service systems are deployed in these VPCs and isolated by subnets. Each service system has an independent application subnet and data subnet. You can create ACLs for these subnets to control internal network traffic between different subnets, as shown in the following figure.

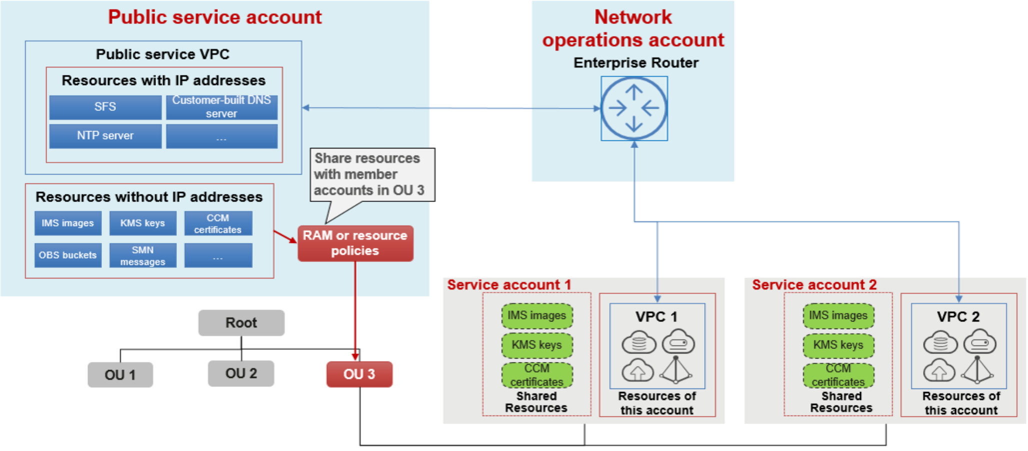
Figure 3-19 Network planning in a service account: one member account for multiple service systems



3.4.4.6 Public Service Management

You need to identify the public IT services and resources required by each business unit, such as the NTP server, Scalable File Service (SFS), self-built DNS servers, OBS buckets, VM images, certificates, and PaaS services such as CodeArts. Then, you can centrally deploy and maintain these public IT services and share them with all business units in the company. Huawei Cloud provides three resource sharing methods.

Figure 3-20 Resource sharing solution



- **Network-based sharing**: You can use Enterprise Router or VPC Peering to connect networks between accounts and share resources. This method applies only to resources that have IP addresses visible to tenants, such as NTP servers, self-built DNS servers, and SFS.
- **RAM-based sharing**: You can use **Resource Access Manager (RAM)** to configure resource sharing and authorize other organizational units (OUs) and accounts to use the shared resources. This sharing method is more secure. For details about the resources that can be shared across accounts through RAM, see **Sharable Cloud Services and Resource Types**
- **Sharing based on resource policies**: You can use resource policies to grant other accounts permissions to access resources, such as OBS bucket policies, Image Management Service (IMS) **image sharing**, and Cloud Backup and Recovery (CBR) **backup sharing**.

3.4.4.7 Unified Multi-account Management

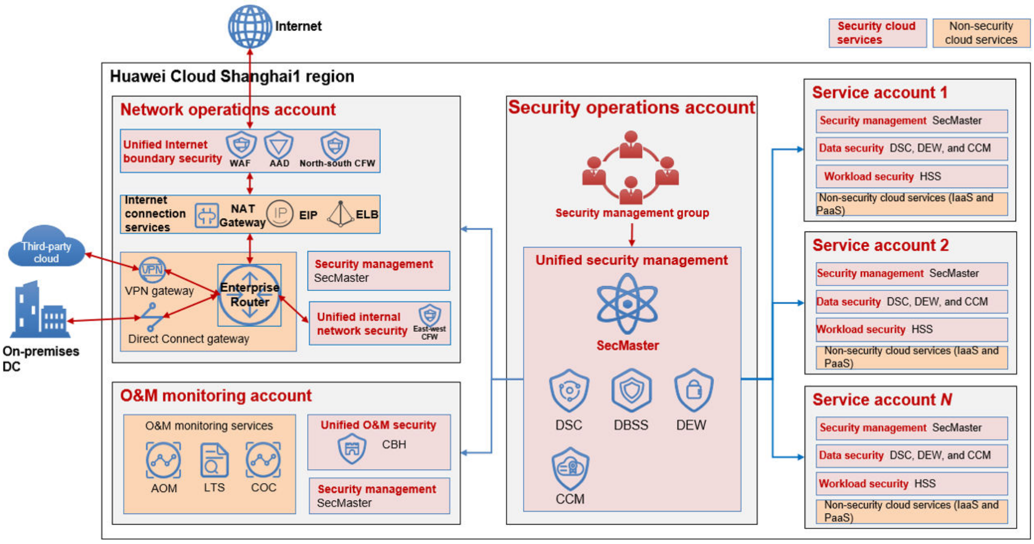
Unified multi-account management covers security, compliance audit, O&M, and finance. It helps enterprises significantly improve management efficiency and consistency, reducing management costs. The more the accounts, the greater the benefits from unified management.

Unified Security Management

SecMaster, Data Security Center (DSC), Database Security Service (DBSS), Data Encryption Workshop (DEW), and Cloud Certificate Manager (CCM) are centrally deployed in the security operations account. This account is used to manage the security of other member accounts.

SecMaster in the security operations account can collaborate with the SecMaster and HSS deployed in other accounts. You can use the security operations account to perform unified security operations on other accounts without logging in to these accounts, including managing cloud assets, security posture, security information and events, and security orchestration and response. DSC in the security operations account can centrally protect the data of all member accounts, including identifying data security risks and protecting data (data watermarking and data masking). DBSS in the security operations account can perform cross-account database audit and centrally display audit information collected by agents when the network is reachable. CCM in the security operations account can centrally apply for SSL certificates and share them with other accounts via RAM. DEW in the security operations account can centrally create Key Management Service (KMS) keys and share them with other accounts via RAM.

Figure 3-21 Unified security management for multiple accounts


Internet
Security cloud services
Non-security cloud services
Huawei Cloud Shanghai1 region
Network operations account
Unified Internet boundary security
WAF
AAD
North-south CFW
Internet connection services
NAT Gateway
EIP
ELB
Third-party cloud
On-premises DC
VPN
VPN gateway
Enterprise Router
Direct Connect gateway
Security management
SecMaster
Unified internal network security
East-west CFW
O&M monitoring account
O&M monitoring services
AOM
LTS
COC
Unified O&M security
CBH
Security management
SecMaster
Security operations account
Security management group
Unified security management
SecMaster
DSC
DBSS
DEW
CCM
Service account 1
Security management SecMaster
Data security DSC, DEW, and CCM
Workload security HSS
Non-security cloud services (IaaS and PaaS)
Service account 2
Security management SecMaster
Data security DSC, DEW, and CCM
Workload security HSS
Non-security cloud services (IaaS and PaaS)
Service account N
Security management SecMaster
Data security DSC, DEW, and CCM
Workload security HSS
Non-security cloud services (IaaS and PaaS)

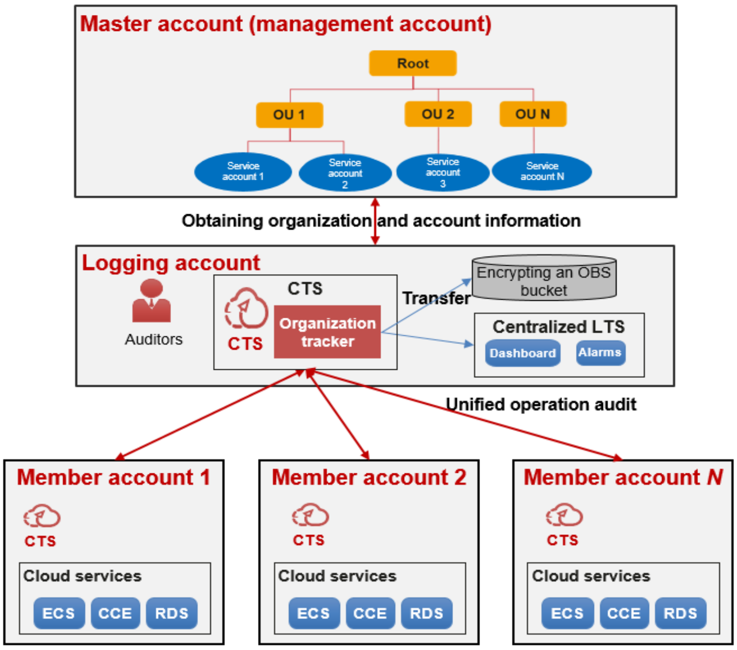
Network security protection services, such as WAF, Anti-DDoS, and CFW, are centrally deployed in the network operations account following the principle of proximity to protect network connection resources such as NAT gateways and EIPs.

Unified Compliance Audit

Auditors use the logging account to audit the operations, including configuring trackers and key operation notifications, of all member accounts. They do not need to log in to member accounts one by one.

- You can create an organization tracker in CTS of the logging account to aggregate audit logs collected by CTS of all member accounts and transfer the audit records to LTS of the logging account.
- You can centrally view the audit records of all member accounts in LTS.
- You can also configure alarm notifications for key operations, such as creating and deleting resources, in LTS.

Figure 3-22 Unified operation audit for multiple accounts



Auditors can audit member accounts' resource configurations based on the **organization rules** and **organization conformance packages** provided by Config, and centrally view non-conforming resource configurations.

Unified O&M Management

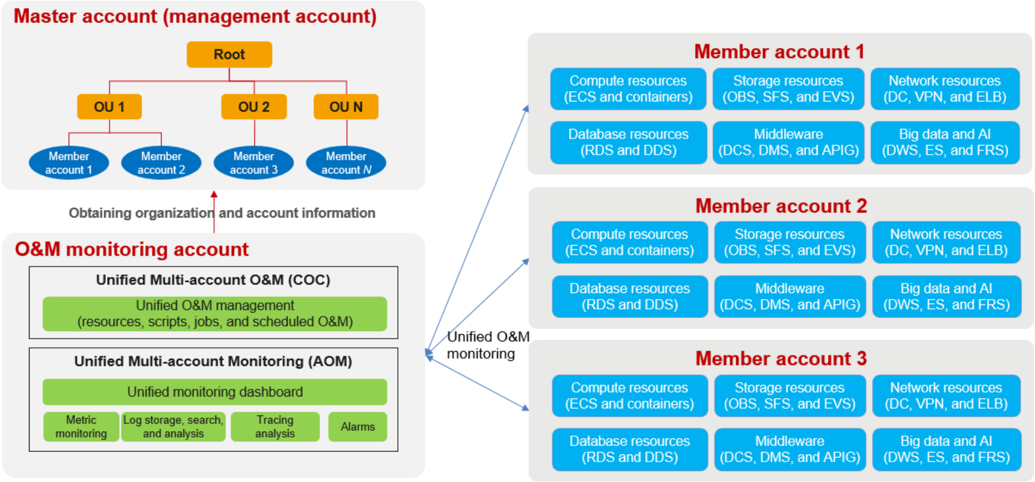
Cloud Operations Center (COC) and Application Operations Management (AOM) are centrally deployed in the O&M monitoring account to monitor and maintain other member accounts in a unified manner, as shown in the following figure.

AOM in the O&M monitoring account can collaborate with AOM in other accounts. You can access the monitoring metrics of each cloud service in other accounts, view these metrics with the O&M monitoring account, and configure alarm rules in a unified manner.

For details, see **Unified Metric Monitoring**.

COC in the O&M monitoring account can centrally manage cloud resources of other accounts and deliver O&M instructions to other accounts.

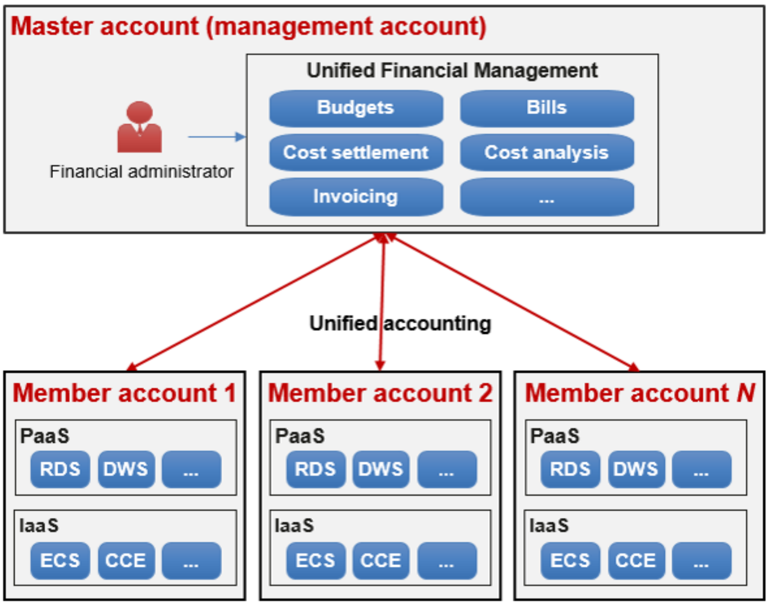
Figure 3-23 Unified O&M management for multiple accounts


Master account (management account)
Root
OU 1
OU 2
OU N
Member account 1
Member account 2
Member account 3
Member account N
Obtaining organization and account information
O&M monitoring account
Unified Multi-account O&M (COC)
Unified O&M management (resources, scripts, jobs, and scheduled O&M)
Unified Multi-account Monitoring (AOM)
Unified monitoring dashboard
Metric monitoring
Log storage, search, and analysis
Tracing analysis
Alarms
Unified O&M monitoring
Member account 1
Member account 2
Member account 3
Compute resources (ECS and containers)
Storage resources (OBS, SFS, and EVS)
Network resources (DC, VPN, and ELB)
Database resources (RDS and DDS)
Middleware (DCS, DMS, and APIG)
Big data and AI (DWS, ES, and FRS)

Unified Financial Management

You are advised to choose unified accounting when creating member accounts in the Enterprise Center. After this function is enabled, the financial administrator can centrally manage the funds, bills, and invoices of the member accounts with the master account. The master account pays for the cloud resources used by the member accounts. Huawei Cloud only issues invoices for the master account because the master account acts as the transaction entity for Huawei Cloud. The following figure shows the relationships between the master account and member accounts.

Figure 3-24 Unified financial management for multiple accounts



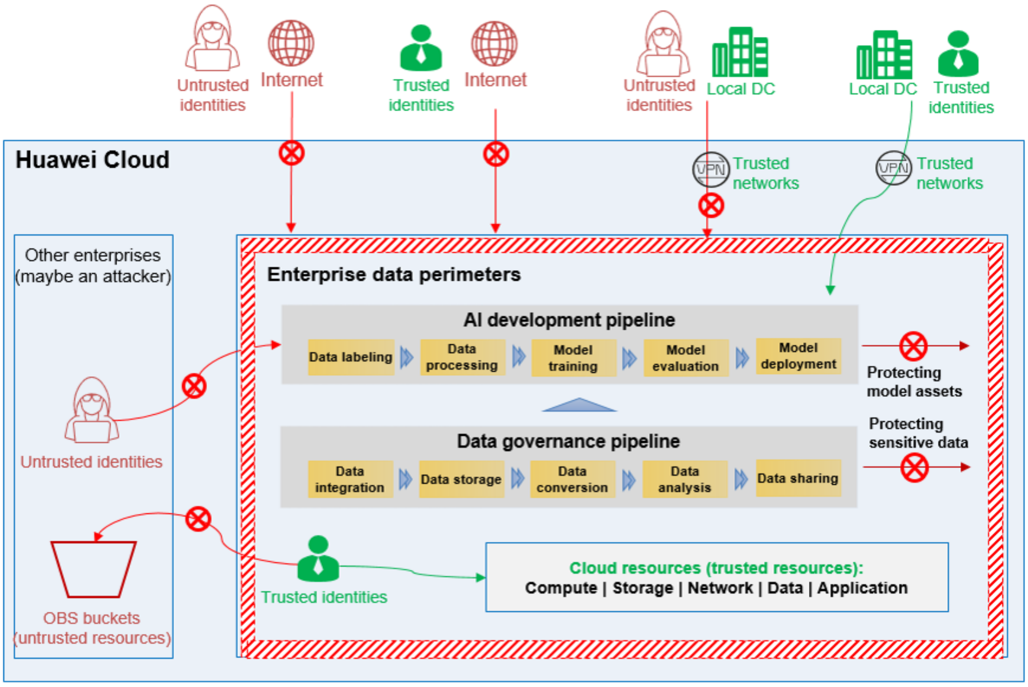
In the unified accounting mode, the master account can perform the following financial management operations for member accounts:

- **Shared discounts**: The master and member accounts share discounts by default. This eliminates the need for customers to repeatedly apply for discounts for each member account and significantly reduces customer costs.
- **Unified payment**: The master account does not need to manually allocate cash, credit, or cash coupons to member accounts for resource usage. Instead, the master account pays the expenditures of all member accounts. This significantly eases the financial workload.
- **One-stop bill management**: The master account can query bills of all member accounts, and check all these bills in one place.
- **Unified invoicing**: The master account can issue invoices for the expenditures of a single member account or all member accounts together.
- **Unified cost management**: The master account can manage the costs of all member accounts in a unified manner, including unified budget management and cost monitoring, analysis, forecasting, and optimization. This greatly improves the cost management efficiency for enterprise customers.

3.4.4.8 Data Perimeters

Huawei Cloud provides all-round data perimeters to protect your sensitive data through identity control policies, network control policies, and resource governance policies. Permissions are granted only to authenticated and trusted identities to access specific resources in a trusted network environment that meets security standards. As shown in the following figure, access requests from trusted identities to cloud resources using the Internet (untrusted network) are rejected. Access requests from untrusted identities to cloud resources using the local data center network (trusted network) are rejected. Access requests from trusted identities to object storage buckets (untrusted resources) of other enterprises are also rejected. Only access requests from trusted identities to cloud resources using the local data center network (trusted network) are allowed.

Figure 3-25 All-round data perimeters



The all-round data perimeters provide the following data protection capabilities:

- Service accounts are not allowed to directly access the Internet. Only the DMZ network of the network operations account can be used to provide Internet services or access the Internet.
- Users can only access the Huawei Cloud management console from the intranet to prevent sensitive data from being transmitted over the Internet.
- You can restrict the regions that can be used by users and restrict data transfer within certain regions to meet compliance requirements such as General Data Protection Regulation (GDPR).

3.5 Security Architecture Design

3.5.1 Introduction to Security Architecture Design

Cloud security and traditional IT security have the same aim to protect data and system, but they significantly differ in infrastructure, security responsibilities, security management, compliance, and audit.

- **In terms of infrastructure**, traditional IT security protects enterprise-built physical hardware and network facilities. Security measures aim to protect physical environment and internal network using tools like firewalls, intrusion detection systems, and antivirus software. Cloud security, however, relies on virtualization technologies and cloud service providers' infrastructure. It must address challenges like securing the virtualization layer, ensuring data isolation in shared systems, and protecting APIs.
- **In terms of security responsibilities**, in traditional IT environments, enterprises take full responsibility for security at all layers, including physical hardware, networks, operating systems, applications, and data. In the cloud environment, the shared responsibility model for security is used. Under this model, cloud service providers secure the infrastructure layer, covering data center safety, network safety, and virtualization platform safety. Enterprises, as tenants of cloud services, are responsible for the security configuration and management of the operating systems, applications, and data on the cloud.
- **In terms of security management and technical implementation**, traditional IT security relies more on hardware devices. Security policies need to be manually implemented and updated, which takes a long time. Cloud security, however, leverages the rich security tools and services provided by cloud service providers, such as IAM, virtual firewalls, security groups, and encryption services, to support automated and programmable security management. Cloud security can quickly respond to and adjust security policies, improving security management efficiency.
- **In terms of compliance and audit**, traditional IT enterprises need to ensure security compliance and invest a large number of resources in audit and certification. Cloud providers often hold various global security certifications. Enterprises can build on this compliance framework while managing the compliance of their own applications and data.

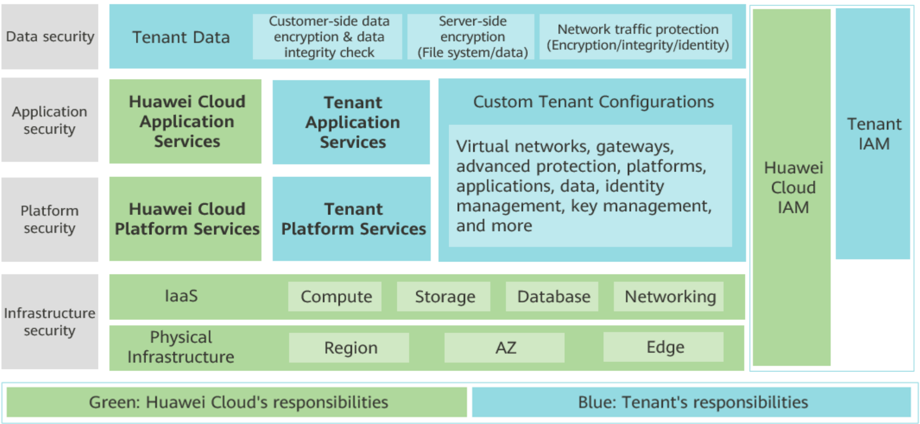
Huawei Cloud's overall cloud security design and practices offer complete, multi-dimensional, flexible, and customizable security and privacy features, covering infrastructure, platform, application, and data security. In addition, Huawei Cloud security services let you customize a variety of advanced security settings. These security services boast deep integration with cloud service security features, settings, and controls across a multi-layered architecture. They present the seamless orchestration of a number of siloed technologies and increasingly automated cloud security O&M.

In conclusion, cloud security and traditional IT security differ greatly in their focuses and execution. During cloud migration, enterprises need to review and update their security policies and architecture. They should leverage built-in cloud-native security features from cloud service providers, align with cloud-based security practices, and safeguard their services and data.

3.5.2 Shared Responsibility Model

Huawei Cloud prioritizes security compliance above all. Both Huawei Cloud and its customers share this responsibility. To define each party's role in protecting cloud security, Huawei Cloud has defined a shared responsibility model for security, as shown in the following figure.

Figure 3-26 Shared responsibility model for security


Data security
Tenant Data
Customer-side data encryption & data integrity check
Server-side encryption (File system/data)
Network traffic protection (Encryption/integrity/identity)
Application security
Huawei Cloud Application Services
Tenant Application Services
Custom Tenant Configurations
Virtual networks, gateways, advanced protection, platforms, applications, data, identity management, key management, and more
Platform security
Huawei Cloud Platform Services
Tenant Platform Services
Infrastructure security
IaaS
Compute
Storage
Database
Networking
Physical Infrastructure
Region
AZ
Edge
Huawei Cloud IAM
Tenant IAM
Green: Huawei Cloud's responsibilities
Blue: Tenant's responsibilities

Huawei Cloud is responsible for security of its cloud platform and cloud services, as well as the physical environments of Huawei Cloud data centers where the IaaS, PaaS, and SaaS operate. Huawei Cloud is responsible for providing secure and efficient, and ensuring secure O&M of, cloud infrastructure and services, and responsible for complying with relevant compliance requirements.

- Huawei Cloud securely develops, configures, and deploys cloud technologies. It also provides secure O&M by quickly detecting, isolating, and responding to incidents to restore cloud services promptly. Meanwhile, Huawei Cloud uses an effective vulnerability management system for its cloud services. The system promptly responds to cloud service security vulnerabilities, ensures fast updates during CSP O&M windows, and maintains uninterrupted service for users. The approaches involve optimizing default security settings for cloud services, prioritizing patch installation over R&D, and streamlining patch release schedules. In addition, Huawei Cloud is developing easy-to-use cloud native security services that are competitive in the market.
- Huawei Cloud takes infrastructure security and privacy protection as the top priority of secure O&M. The infrastructure primarily consists of the physical environment for deploying cloud services, including Huawei-developed

software and hardware and all types of cloud service system facilities for O&M such as computing, storage, networking, databases, platforms, applications, and security, for example, IAM and high-level security services. In addition, Huawei Cloud integrates third-party security technologies or services and is responsible for their secure O&M.

- Huawei Cloud is also responsible for the security configuration and version maintenance of the cloud services it supports.
- Huawei Cloud provides tenants with comprehensive data protection functions such as privacy, integrity, availability, durability, authentication, authorization, and non-repudiation, and is responsible for the security of related functions. Huawei Cloud keeps the data, but tenants retain full ownership and control of their information. Huawei Cloud never allows O&M personnel to access tenant data without authorization. Huawei Cloud stays up-to-date with internal and external compliance requirements, complies with security laws and regulations required for running Huawei Cloud services, evaluates security standards of the industries it served, and shares compliance practices with tenants.
- Huawei Cloud works with cloud security partners to provide consulting services for tenants. For example, Huawei Cloud assists tenants in configuring security settings for virtual networks and VMs (including host VMs and guest VMs), managing security patches for systems and databases, customizing configurations for virtual network firewalls, API gateways, and advanced security services, and performing DoS/DDoS attack defense drills, emergency response to security incidents, and disaster recovery drills.

Tenants of Huawei Cloud are responsible for the secure and effective management of the configurations of cloud services including IaaS, PaaS, and SaaS. This includes but is not limited to virtual networks, the OSs of guest VMs and host VMs, virtual firewalls, API gateways, advanced security services, all types of cloud services, tenant data, identity accounts, and key management.

- Tenant-specific security responsibilities are ultimately based on the cloud services that tenants use. These responsibilities are tied to the specific default or customized security configurations they apply. Huawei Cloud offers the necessary resources, features, and performance for tenants to perform security tasks. Tenants must configure security settings for their resources.
- Tenants need to: (1) Configure policies for firewalls, gateways, and advanced security services of their virtual networks. (2) Manage security settings, including updates and patches, for virtual networks, host VMs, guest VMs, and cloud services like containers. Manage container security, including configuring access control for container clusters, nodes, and containers. Configure security policies for the big data analysis platform. (3) Configure other security policies for cloud services they lease. (4) Manage security of any applications or tools they deploy on Huawei Cloud.
- Tenants are responsible for testing security configurations before deploying services to their production environment to prevent negative impacts on their applications and businesses. For the security of most cloud services, tenants only need to control account access to resources and keeps credentials secure. A few cloud services require other tasks to achieve the desired security. There are many security configurations in monitoring, management, and advanced security services. Tenants can seek technical support from Huawei Cloud and its partners to complete these configurations and ensure security.

- Let's take Huawei Cloud MapReduce Service (MRS) as an example. Tenants need to (1) Configure EIP and virtual network firewalls for the MRS big data cluster they purchase. (2) Control access to the big data cluster, for example, by only allowing trusted networks or hosts to access the EIP port. (3) Manage big data cluster users, configure security policies for big data components, and properly keep related account credentials. (4) Manage the security of applications deployed on the big data cluster.
- Let's take a database service as an example. Tenants need to manage the lifecycle and security of their database engines, including (1) using the latest instance version by default and upgrading the version in a timely manner based on the official website prompt and vulnerability notice; (2) sorting out asset categories and formulating database instance protection policies, such as designing active/standby instances or clusters, planning data backup and recovery, configuring VPCs and security groups, managing internet access, encrypting connections, setting database authentication and authorization policies, enabling the audit service, and applying other necessary security settings.
- Regardless of which Huawei Cloud service is used, tenants are always the owners and controllers of their data. Tenants are responsible for configuring data security, and ensuring the confidentiality, integrity, and availability of data, and authenticating and authorizing data access. For example, when using Identity and Access Management (IAM) and Data Encryption Workshop (DEW) services, tenants are responsible for keeping their service accounts, passwords, and keys safe, and shall comply with industry best practices in configuring, updating, and resetting passwords and keys. Tenants need to set up personal accounts and multi-factor authentication (MFA), properly use secure transfer protocols to communicate with Huawei Cloud resources, and set up user activity logs for monitoring and auditing.
- Tenants are responsible for ensuring compliance of applications and services deployed on Huawei Cloud that are not provided by Huawei Cloud, and evaluating the security standards of the industries they serve.

Note that the security responsibilities depend on the service type (IaaS, PaaS, SaaS), service mode (cloud service or cloud software), and the security service provider. In 2024, Huawei Cloud and China Academy of Information and Communications Technology (CAICT) jointly released the *Shared Responsibility Model for Cloud Security* in 2024. This white paper introduces the shared responsibility for cloud security 2.0 in detail and offers guidance for applying this model effectively.

3.5.3 Security Design Principles

Huawei Cloud has summarized the following ten security design principles based on its security practices and successful projects. You can follow these principles to design effective cloud security solutions.

- **Zero Trust Principle**

  Follow the "Never Trust, Always Verify" concept and assume that no one or program is trustworthy, whether it is an internal user, external user, or network device. Components in a system must be explicitly verified before any communication to reduce the attack risks. Zero Trust transforms the existing static trust model (black-and-white) based on authentication and default authorization into a dynamic trust model based on continuous risk evaluation and adaptive authorization. Zero Trust does not determine the credibility based on the network location. It focuses on protecting resources, not CIDR blocks. Compared with the traditional security concepts, Zero Trust shifts the focus of network defense from static network-based boundaries to users, devices, and resources. All resources (such as people, objects, devices, applications, networks, data, and supply chains) require continuous identity authentication and trust evaluation, and dynamic security policies must be applied globally. Zero Trust reduces the attack surface and ensures system security through dynamic and continuous risk evaluation.

- **Principle of Least Privilege**

  Assign only the minimum permissions to users or applications to do their tasks based on the fine-grained authorization. This limits access and reduces the risk of attacks. If passwords or applications are hacked, attackers will not get wider access. If user or application tasks change, you need to update their permissions promptly to ensure that the tasks can be completed.

- **Defense in Depth**

  Do not rely on a single security layer. Use multiple layers of security instead. If one layer fails, the rest will still provide protection. Think of it as a castle with a moat, walls, and gates working together for defense. Establish a defense-in-depth mechanism covering the entire technology stack and apply multiple types of security control measures to all technology stacks, including network edges, VPCs, cloud storage, ECSs, operating systems, application configurations, and code.

- **Balance Between Security and Cost**

  The defense-in-depth mechanism is recommended, but a more comprehensive security protection solution may cost too much. To balance security and cost, design a cost-effective security protection solution based on the compliance requirements (such as classified information security protection) of service systems and sensitive data classification. Focus only on necessary protections instead of using all-round and high-level security protection solutions everywhere.

- **Cloud Native Security**

  Use cloud native security services. Cloud service providers provide plenty of cloud native security services, such as Web Application Firewall (WAF), Anti-DDoS, Cloud Firewall (CFW), and Data Encryption Workshop (DEW). These services are deeply integrated with the cloud platform, and have excellent performance, elasticity, and convenience. Cloud service providers are experienced in managing security operations and consistently improve cloud-native security features. For details about the cloud native security services provided by Huawei Cloud, see **Cloud Native Security Service**.

- **Attacker Perspective**

  Evaluate service system security by thinking like an attacker, identify vulnerabilities, and enhance defenses. In this way, you can reduce attack risks, strengthen protection, and increase the costs of attacks.

- **Continuous Security Operations**

  Security depends less (30%) on technology and more (70%) on operations. Continuously improve security management processes, conduct regular security operations, and monitor and assess cloud environment compliance to maintain long-term system stability and safety.

- **Barrel Principle**

  Security operates like a system where every part matters. A single weak point lowers overall protection.

- **Separation of People and Data**

  Use tools and management systems to limit user access to data, minimizing human errors like accidental deletions or changes when handling sensitive information.

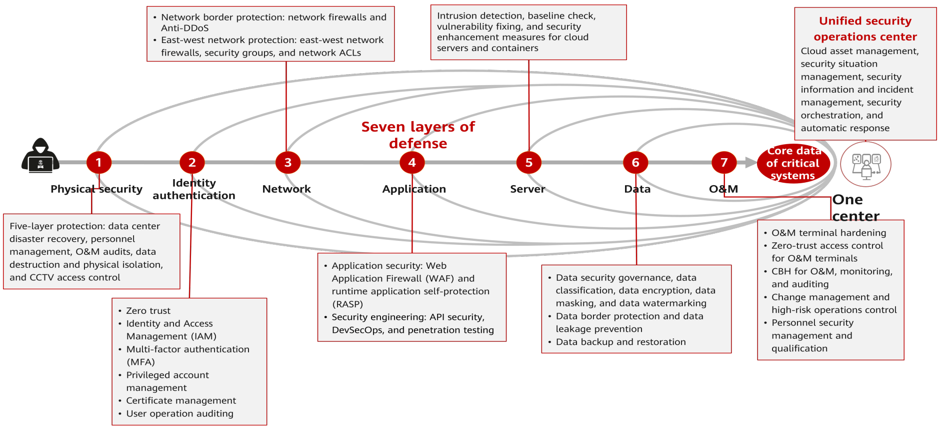
- **DevSecOps**

  Integrate security throughout the entire software development lifecycle, including requirement analysis, design, development, testing, deployment, maintenance, and operations, to maintain system safety and reliability. To achieve this, DevSecOps is recommended. It integrates security checks into DevOps automation, enabling faster vulnerability detection and fixes while boosting software development speed and quality.

3.5.4 Security Reference Framework

Huawei Cloud suggests using its "1 center + 7 layers of defense" security reference framework and cloud-native security services to create a robust cloud protection system, as shown below. The security reference framework follows the defense-in-depth principle and combines various security measures. It tailors specific protection for critical systems and essential enterprise data according to their needs. This creates several layers of defense that work together, make it harder for attackers to break them. These layers also provide enough time for detecting and responding to threats.

Figure 3-27 One center and seven layers of defense


• Network border protection: network firewalls and Anti-DDoS
• East-west network protection: east-west network firewalls, security groups, and network ACLs
Intrusion detection, baseline check, vulnerability fixing, and security enhancement measures for cloud servers and containers
Unified security operations center
Cloud asset management, security situation management, security information and incident management, security orchestration, and automatic response
Seven layers of defense
1
Physical security
2
Identity authentication
3
Network
4
Application
5
Server
6
Data
7
O&M
Core data of critical systems
One center
Five-layer protection: data center disaster recovery, personnel management, O&M audits, data destruction and physical isolation, and CCTV access control
• Zero trust
• Identity and Access Management (IAM)
• Multi-factor authentication (MFA)
• Privileged account management
• Certificate management
• User operation auditing
• Application security: Web Application Firewall (WAF) and runtime application self-protection (RASP)
• Security engineering: API security, DevSecOps, and penetration testing
• Data security governance, data classification, data encryption, data masking, and data watermarking
• Data border protection and data leakage prevention
• Data backup and restoration
• O&M terminal hardening
• Zero-trust access control for O&M terminals
• CBH for O&M, monitoring, and auditing
• Change management and high-risk operations control
• Personnel security management and qualification

- **Physical security**

  Huawei Cloud data centers are protected by five layers of physical security, including data center disaster recovery, personnel management, O&M audits, data destruction and physical isolation, CCTV, and access control. Tenants do not need to pay attention to the physical security of data centers. However, enterprises need to maintain the physical security of their data centers where the dedicated cloud is deployed.

- **Identity authentication**

  Huawei Cloud verifies identities based on the principle of zero trust, and assigns permissions based on the principle of least privilege (PoLP). By default, multi-factor authentication is enabled to prevent unauthorized access. Privileged accounts are properly managed, and any user operations on the cloud platform are recorded and audited.

For details, see **IAM best practices**.

- **Network**

  The core is to protect network borders and control east-west access on the intranet.

  - **Network border protection**: The network border refers to the border where one network connects to another, typically where Internet access, VPN, and private line access are involved. Cloud Firewall (CFW), VPC security groups, and network ACLs provided by Huawei Cloud control traffic flowing through the network border. CFW detects and prevents network intrusions. Whitelists can be used to allow only limited access and traffic over high-risk ports and protocols are not allowed.
  - **East-west network protection**: You can manage services by confidentiality level, for example, deploying services of different confidentiality levels in different VPCs. You can use VPCs to isolate large network security zones, use CFW to control access between east-west VPC networks, and use VPC security groups to protect instances and network ACLs to protect subnets.

- **Application**

  WAF should be deployed to protect application servers that provide Internet-accessible services. To design application security layer, you need to pay attention to software security engineering and improve the built-in security capability of applications. From the perspective of security risks, you should pay more attention to applications connected to external networks. In addition, you need to identify internal core applications and strengthen their security.

- **Server**

  Server-level intrusion detection works better. Install server security products on servers to handle vulnerabilities and configure security settings.

- **Data**

  You can identify and categorize data assets early, manage data security across its entire lifecycle, anonymize, encrypt, and audit critical data usage, and regularly back up essential data. And also, you can build a solid data security border, by applying identity authentication, network access control, and resource control, to prevent sensitive data leaks.

- **O&M**

  Initiate O&M activities only from secure networks. Establish dedicated access channels for O&M activities. For example, let O&M personnel use dedicated O&M services and bastion hosts for O&M to minimize black screen O&M operations and ensure that O&M activities can be audited and traced.

- **One center**

  Security depends less (30%) on technology and more (70%) on operations. Only when security products at each defense layer are correctly configured and well maintained can security be effectively implemented. To achieve this, a unified security operations platform that integrates various security products is needed.

Establishing an in-depth defense system may be time-consuming. You need to balance security, efficiency, cost, and experience throughout the process. You should regularly assess security risks, strengthen defenses against top threats, continuously enhance protective measures, and test system effectiveness through activities like team/blue team drills.

3.6 Platform Engineering

3.6.1 What Is Platform Engineering?

Platform engineering is an engineering discipline that optimizes software delivery and lifecycle management by building and operating an internal developer platform (IDP). Platform engineering simplifies developer interactions with infrastructure by standardizing processes and automating tasks, boosting both efficiency and delivery speed. Gartner has listed platform engineering as one of the top 10 strategic technology trends in 2023 and 2024, and predicts that by 2026, 80% of large software engineering organizations will establish platform teams to provide developers with reusable services, components, and tools. Platform engineering brings the following benefits to enterprises:

- **Enhanced developer experience**: Platform engineering provides self-service tools that simplify infrastructure configurations, application deployment, and management processes. This lets developers focus more on service logic development rather than underlying infrastructure management.
- **Accelerated software delivery**: By providing pre-defined environments, automated processes, and reusable components, platform engineering can significantly accelerate the software delivery and bring products to market faster.
- **Improved operations efficiency**: Platform engineering reduces manual operations and human errors, improves operations efficiency, and cuts operational costs through automated and standardized processes.
- **Strengthened security compliance**: Platform engineering integrates security policies and compliance checks to ensure that applications and infrastructure comply with security standards and regulatory requirements.
- **Promoting innovation**: Platform engineering provides developers with more flexible and convenient development environments, encouraging them to try new technologies and methods.

3.6.2 How Can We Build Platform Engineering?

You can use cloud platform's extensive services and tools to lower building and maintenance expenses while boosting IDP's reliability and scalability. The following are some key steps.

Clearly Define the Objectives and Requirements for Platform Engineering

Platform engineering aims to accelerate software delivery, improve lifecycle management, boost development efficiency through a self-service IDP. To this end, you need to analyze your company's development process and identify the key issues that require improvement. For example, you may want to simplify environment setup, streamline deployments, or reduce resource waiting times. Talk to each team to gather their needs and figure out what services, tools, and features are needed to support the developer's work. Based on the requirement analysis, define the objectives for the platform engineering, including but not limited to:

- Provide a unified platform for application development, testing, and deployment.
- Set up automated CI/CD pipelines for integration and continuous delivery.
- Gather and reuse public components and services across your organization.
- Establish a comprehensive monitoring and O&M mechanism.
- Secure the platform and maintain compliance with regulations.

Set Up an IDP Based on Huawei Cloud

Huawei Cloud provides a variety of cloud services to help you quickly build an IDP.

1. **CodeArts** is a one-stop platform that provides out-of-the-box cloud services for requirement delivery, code commit, check, build, verification, deployment, and release throughout the entire software lifecycle.
2. **ServiceStage** is an application management and O&M platform that lets you deploy, roll out, monitor, and maintain applications all in one place. It supports technology stacks such as Java, PHP, Python, Node.js, Docker, and Tomcat, and supports microservice applications such as Apache ServiceComb Java Chassis (Java chassis) and Spring Cloud.
3. **Cloud Service Engine (CSE)** is a cloud middleware used for microservice applications. It supports ServiceComb engines contributed to Apache and open-source enhanced Nacos engines and application gateways. You can also use other cloud services to quickly build a cloud-native microservice system, implementing quick development and high-availability O&M of microservice applications.
4. **Cloud Application Engine (CAE)** is a serverless hosting service for webs and microservice applications. It provides one-stop application hosting with fast deployment, low costs, and simplified O&M. CAE releases applications from source code, software packages, and image packages quickly and easily, with auto scaling to the second, all billed pay-per-use. The whole application lifecycle is manageable with observable metrics.

Organize and Consolidate Public Services

Organizing and consolidating public components and services across an organization is the key to maximize platform engineering value. To do this, you need to start by finding the software components used across different applications, like authentication, log, and message components. Make them consistent and reusable by setting standard interfaces, data formats, and error-handling methods. Then, encapsulate these public components into independent microservices and deploy them on CSE. CSE's service registry center allows for microservice registry and discovery. Other application systems can call the APIs of these microservices to use these common components, avoid repeated development, improve development efficiency, and ensure code quality consistency. You can also use the service governance function of CSE to further improve the reliability and performance of public components. You need to prepare detailed component user guides and API documentation for developers to access and use. In addition, you need to establish a version control system for components to streamline upgrades, maintenance, and ongoing improvements.

Establish a Reliable Monitoring and O&M System

You can use **Application Operations Management (AOM)** to centrally monitor microservice performance metrics, such as the response time, error rate, and number of calls, helping you quickly locate and rectify faults.

You can use **Log Tank Service (LTS)** to collect and analyze system logs and application logs. LTS supports log search, alarm reporting, and report generation, helping you locate faults and analyze performance.

You can use **Application Performance Management (APM)** to monitor application call traces, analyze performance bottlenecks, locate abnormal requests, and improve application performance.

Improve and Optimize Platform Engineering

Platform engineering is a continuous optimization process driven by user needs and technology updates. Regularly communicate with development, testing, and operations teams to gather their suggestions and feedback on the platform. Analyze monitoring data to identify performance bottlenecks and optimize resource allocation.

Monitor the development of new technologies such as cloud native, containerization, microservices, and serverless, and evaluate their application value on the platform. Choose appropriate projects to test new technologies, then expand their use once proven effective. Improve the platform's user guides, API documentation, and best practices to help new users get started quickly. Organize regular training sessions to enhance the team's understanding and usage of the platform.

3.7 Cloud Operating Models

3.7.1 What Are Cloud Operating Models?

Before the emergence of cloud computing technologies, enterprises had established IT operating models to define how IT supports service development. In a narrow sense, IT operating models refer to how an enterprise manages and operates its IT resources, services, and infrastructure. Enterprises use these models in effectively configuring, managing, and optimizing IT resources to enhance performance and efficiency, reduce costs, and increase flexibility to align with their service objectives and strategies. Broadly, the IT operating models also encompass organizational structure, operational processes, roles, and responsibilities. Simply put, IT operating models define how the IT department functions. A traditional IT operating model is centered on IT infrastructure deployed in on-premises data centers or IDCs. This includes IT hardware and basic software like virtualization. Enterprises typically purchase IT hardware and software assets upfront to ensure the secure, stable operation of service systems. Over time, the performance of IT hardware deteriorates, or the hardware itself may fail, requiring technical personnel to dedicate significant time to manage, maintain, and update it.

With the emergence of cloud computing technologies, enterprises began building IT infrastructure using cloud platforms and services, gradually migrating numerous service systems or directly deploying them on public clouds. This marked the transition to the cloud computing era. The cloud platform-based IT operating models (often called the cloud operating models) shift the enterprises' focus from IT infrastructure to applications and data assets. In cloud operating models, enterprises must efficiently configure, manage, and optimize cloud resources to ensure the secure, stable operation of service systems on the cloud. Simply put, the cloud operating models represent how enterprises leverage cloud technologies and services to drive service development. The objectives of the cloud operating models are identical to those of the traditional IT operating models. Both of them use technologies to help enterprises achieve their business goals and maximize business value. The following table highlights the differences between the cloud operating models and IT operating models.

Table 3-11 Differences between the IT operating models and cloud operating models

Category	IT Operating Model	Cloud Operating Model
Cost model	● Relies on capital expenditure (CapEx). Hardware must be planned and purchased in advance, which takes significant time.	● Operates on a pay-per-use operating expense (OpEx) basis, allowing enterprises to adjust costs flexibly based on actual usage, reducing upfront investments.
Key points of management	● Enterprise management emphasizes IT infrastructure maintenance, server uptime, and physical security of data centers.	● Enterprise management focuses on higher-level operations, such as application performance optimization, data management, and cloud security.
Agility	● Hardware procurement and deployment are time-consuming. Resource expansion requires complex approval and procurement processes, resulting in slow responses. ● Innovation and change are restricted by hardware resources, making adaptation to service needs difficult.	● Cloud resources can be scaled dynamically based on demand, enabling rapid responses to service requirements. ● Deploying new applications or features is faster, supporting agile development and continuous delivery. ● Innovation is no longer tied to hardware procurement cycles, allowing quick testing and launching of new products or services.

Category	IT Operating Model	Cloud Operating Model
Security	● Enterprises are responsible for security. ● Protection focuses on the physical perimeter and internal network security of data centers.	● Security follows a shared responsibility model. Cloud service providers manage the security of cloud platforms and services, while enterprises are responsible for the security of upper-layer applications and data. ● Cloud providers also offer cloud native security services and best practices to support enterprises in protecting their applications and data.
Required skills	● Technical personnel primarily manage and maintain IT infrastructure, spending considerable time on hardware fault resolution, performance optimization, system updates, and more. ● Skills include hardware maintenance, network management, and virtualization.	● Technical personnel require expertise in cloud platform usage, cloud resource configuration, optimization, automatic O&M tools, cloud security management, and more. ● Advanced skills are needed for application performance optimization and data management.

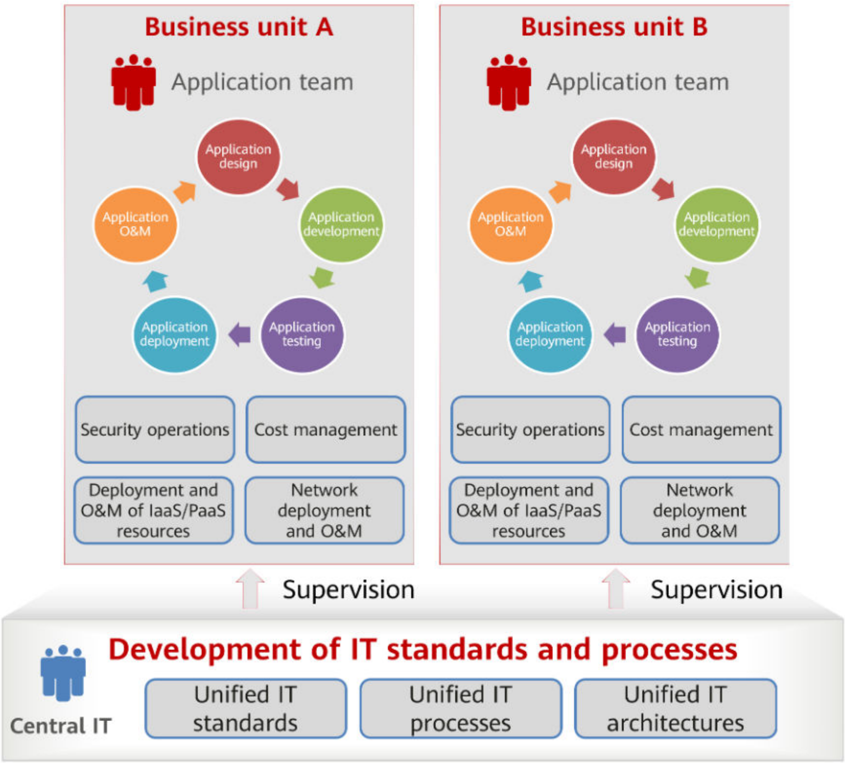
The cloud operating models offer significant advantages in flexibility, agility, and cost-effectiveness. However, they also impose higher demands on personnel skills and security management within enterprises. Enterprises must gradually transition from the traditional IT operating models to the cloud operating models in alignment with their service requirements and development strategies.

The cloud operating modes are not the natural outcome of enterprise cloud transformation. Simply migrating all service systems to the cloud does not automatically result in an effective cloud operating model that fully supports service objectives. A robust cloud operating model is a prerequisite for the successful cloud transformation of an enterprise. To maximize the business value derived from cloud computing, enterprises need to design optimal cloud operating models tailored to their existing IT operating models and the characteristics of their service systems. The responsibility matrix and collaboration mechanism between the CCoE and the application teams must be clearly defined for the cloud operating models. Drawing from the cloud transformation experiences of numerous enterprises, Huawei Cloud has identified three kinds of cloud operating models.

3.7.2 Decentralized Operating Model

The decentralized operating model is the simplest operational structure. In this model, all service systems are managed independently by their respective dedicated application teams. These teams are not only responsible for the design, development, testing, deployment, and O&M of applications, but also for the deployment and O&M of the IaaS and PaaS resources required by the service systems. Additionally, they must ensure the security of service systems and manage the costs of cloud resources. The central IT team is responsible for developing unified IT standards and processes, issuing relevant documentation to service systems for reference and overseeing their implementation. However, the central IT team cannot enforce these standards and processes on the service systems. In this operating model, there is no requirement to establish a dedicated CCoE team.

Figure 3-28 How the decentralized operating model works



The decentralized operating model offers several advantages.

- **High agility**: Each business unit can quickly deploy and scale resources based on their own requirements, accelerating innovation.
- **Service-oriented**: The application teams have a better understanding of service requirements and can customize cloud solutions accordingly.
- **Clear responsibilities**: Each business unit manages their own cloud environment, making it easier to identify and resolve issues and optimize performance.

The decentralized operating model also has its disadvantages.

- **Lack of consistency**: Since each business unit independently deploys and maintains their required cloud environment, the absence of unified IT standards or security policies can lead to inconsistencies. This complicates management efforts.

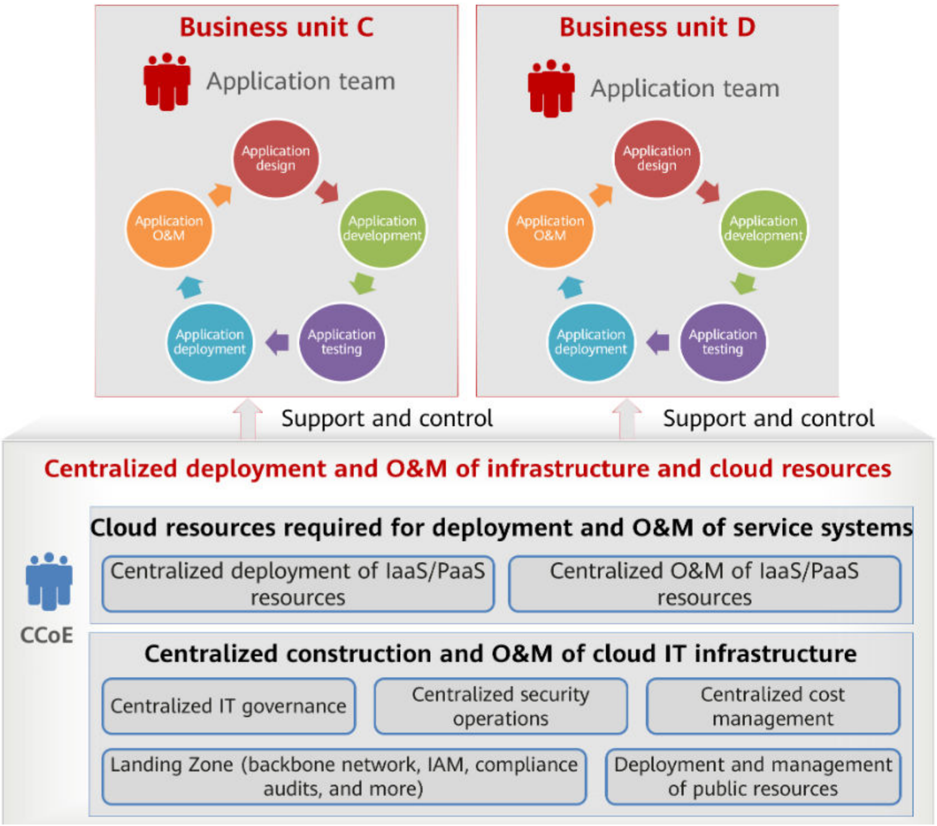
- **Increased costs**: Without centralized coordination, duplicated efforts and resource waste can escalate cloud costs.
- **Lack of an overall view**: Difficulty in obtaining a comprehensive view of cloud resource usage limits strategic decision-making and optimization.

The decentralized operating model is ideal for innovative service systems requiring full control over cloud resource creation and O&M. These systems need to rapidly innovate and iterate based on evolving service requirements.

3.7.3 Centralized Operating Model

The decentralized operating model prioritizes quick innovation over centralized control. The centralized operating model focuses on centralized control rather than rapid innovation. In the centralized operating model, the CCoE team is responsible for centrally building and maintaining Landing Zone, including key components like the backbone network, Identity and Access Management (IAM), and compliance audit systems on the cloud. Furthermore, the CCoE team enforces centralized IT governance across the enterprise-wide cloud environment. All cloud resources required by service systems are centrally deployed and maintained by the CCoE team. This allows the team to efficiently identify public resources needed by each service system, ensuring they are deployed and managed in a unified manner. Additionally, all cloud resources of business units are managed centrally, with centralized security operations and cost management. This alleviates application teams from concerns about deploying and managing infrastructure or cloud resources, enabling them to focus solely on application design, development, testing, deployment, and O&M.

Figure 3-29 How the centralized operating model works


Business unit C
Application team
Application design
Application development
Application testing
Application deployment
Application O&M
Business unit D
Application team
Application design
Application development
Application testing
Application deployment
Application O&M
Support and control
Support and control
Centralized deployment and O&M of infrastructure and cloud resources
CCoE
Cloud resources required for deployment and O&M of service systems
Centralized deployment of IaaS/PaaS resources
Centralized O&M of IaaS/PaaS resources
Centralized construction and O&M of cloud IT infrastructure
Centralized IT governance
Centralized security operations
Centralized cost management
Landing Zone (backbone network, IAM, compliance audits, and more)
Deployment and management of public resources

The centralized operating model offers several advantages.

- **Centralized management**: Centralized management ensures consistent security policies, compliance, and standardized processes, effectively reducing risks. When compared to decentralized management, it simplifies operations and enhances efficiency.
- **Cost optimization**: Unified construction of public resources, centralized procurement, and resource integration improve resource utilization and reduce total cost of ownership (TCO).
- **A global view**: The CCoE team monitors and analyzes enterprise-wide cloud resource usage, allowing for optimized resource configurations.

The centralized operating model also has its disadvantages.

- **Lack of agility**: All cloud resource requests must go through the CCoE team. This will slow down response times and affect service agility.
- **Bottleneck risk**: Overloading the CCoE team may otherwise prevent timely responses to the needs of individual business units, decreasing efficiency.
- **Lack of business understanding**: The CCoE team might not fully understand specific business scenarios, resulting in mismatched resource allocations.
- **Hard to meet diverse requirements**: Unified standards might fail to address the specific needs of different business units.

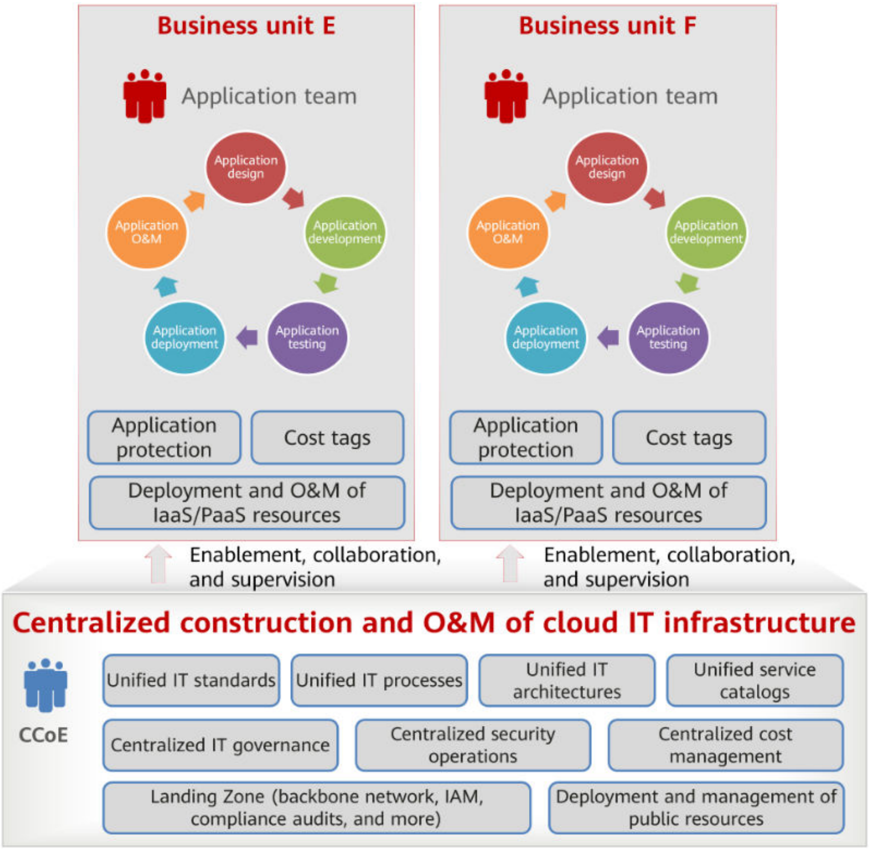
The centralized operating model is most suitable for service systems that are stable and rarely require updates, such as commercial software like Systems Applications and Products (SAP) or mature systems developed internally. This model is also appropriate for service systems that are constructed and maintained in a unified manner by the IT department.

3.7.4 Enablement and Collaborative Operating Model

The enablement and collaborative operating model combines the strengths of both decentralized and centralized operating models, striking a balance between centralized management and service system agility. In the enablement and collaborative operating model, the CCoE team is responsible for centrally building and maintaining Landing Zone, including key components like the backbone network, IAM, and compliance audit systems on the cloud. The CCoE team defines unified IT standards, processes, and architectures while implementing centralized IT governance across the enterprise-wide cloud environment. The CCoE team also enables the application teams to take full ownership of the deployment and O&M of the cloud resources required by service systems. This division of responsibilities eases the workload of the CCoE team, enhances the autonomy of the application teams, and further boosts system agility. To avoid inconsistencies stemming from independent deployments and cloud resource O&M, the CCoE team establishes IT governance policies, mandating compliance across business units. It also creates a unified service catalog to ensure the application teams use only cloud services authorized by the CCoE team for development and deployment.

Close collaboration between the CCoE team and the application teams is essential to maintaining secure, stable service systems while optimizing costs. For O&M, the CCoE team handles routine maintenance of cloud IT infrastructure, including the backbone network, IAM, and compliance audit systems. The application teams manage the routine O&M of applications and their required cloud resources. If there is a service system fault, the CCoE and application teams work together to identify and resolve the issue. In terms of security, the CCoE team manages platform-level protection and centralized security operations. The application teams focus on application-level security, such as preventing SQL injection. Regarding cost management, the CCoE team is responsible for centralized cost management, including centralized cost planning, monitoring, analysis, and optimization, while the application teams add cost tags to cloud resources.

Figure 3-30 How the enablement and collaborative operating model works


Business unit E
Application team
Application design
Application development
Application testing
Application deployment
Application O&M
Application protection
Cost tags
Deployment and O&M of IaaS/PaaS resources
Business unit F
Application team
Application design
Application development
Application testing
Application deployment
Application O&M
Application protection
Cost tags
Deployment and O&M of IaaS/PaaS resources
Enablement, collaboration, and supervision
Enablement, collaboration, and supervision
Centralized construction and O&M of cloud IT infrastructure
CCoE
Unified IT standards
Unified IT processes
Unified IT architectures
Unified service catalogs
Centralized IT governance
Centralized security operations
Centralized cost management
Landing Zone (backbone network, IAM, compliance audits, and more)
Deployment and management of public resources

The enablement and collaborative operating model incorporates the advantages of both centralized and decentralized models.

- **A balance between centralized control and agility**: This model ensures centralized management and unified control and grants business units a degree of autonomy. These features enhance service agility.
- **Close collaboration**: The CCoE team works hand-in-hand with the application teams within business units, preventing buck-passing and boosting overall operational efficiency.
- **Cost optimization**: Unified construction of public resources, centralized procurement, and resource integration improve resource utilization and reduce TCO.
- **A global view**: The CCoE team monitors and analyzes enterprise-wide cloud resource usage, allowing for optimized resource configurations.

The enablement and collaborative operating model also has its disadvantages.

- **Complex implementation**: Complex IT governance measures must be developed to ensure business units comply with unified IT standards.

Alongside this, a unified service catalog must be created and maintained for streamlined operations.

- **High capability requirements**: Effective implementation demands advanced management expertise and extensive experience in IT governance. CCoE team members must possess higher-level skills and act as advocates, guiding application teams to deploy and manage cloud resources in alignment with best practices.

The enablement and collaborative operating model works well for agile systems, like self-developed digital marketing systems, that need quick updates and fast releases of new features to meet fast-changing market demands. Additionally, this operating model is effective for stable service systems supported by independent application teams with robust IT capabilities.

The three kinds of cloud operating models adhere to different types of service systems and organizational structures. In large enterprises, these kinds of models often coexist, working together to facilitate agile iteration while ensuring secure, stable operations for diverse service systems.

The chosen cloud operating models will significantly influence the application lifecycle management process, which will be elaborated on in the following section.

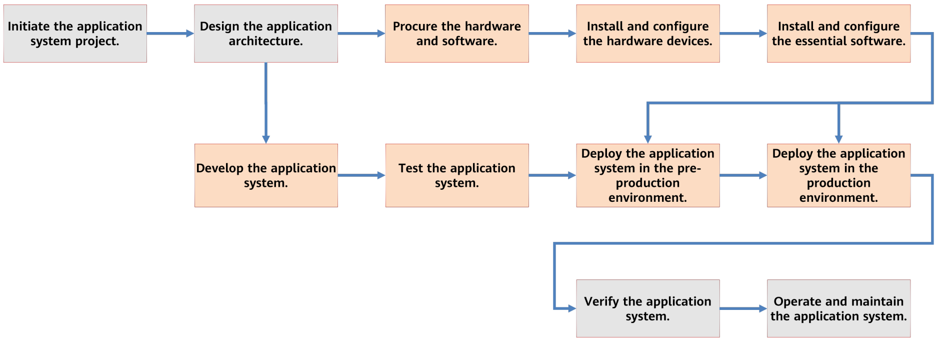
3.8 Application Lifecycle Management

The core of enterprise cloud adoption is to migrate various application systems that support enterprise production and operations to the cloud. The primary goal is to securely and reliably run application systems on the cloud while using cloud computing advantages to improve application system's resilience, agility, security, performance, etc. Additionally, the aim is to develop and update cloud-based applications, adopt new technologies for product, service, and model innovations, create advanced features, streamline operations, enrich user experience, and increase revenue.

However, migrating application systems to the cloud can be complex and systematic. It requires planning through every stage of the application's life based on Application Lifecycle Management (ALM). This ensures the benefits of cloud computing are maximized at each step.

The following figure shows the application lifecycle management process based on traditional IT. Application systems are needed to be deployed on hardware resources. However, the procurement and delivery of hardware resources take a long time. Once an application system project starts, you must purchase both hardware devices and essential software. After the hardware devices are ready, you need to install them in your own or leased data center. Then, install and configure the operating system and virtualization software on the hardware devices to build the pre-production environment and production environment for the application systems. After the application system is developed and tested, it can be directly deployed and run in the pre-production and production environments. Enterprises often buy ready-made commercial software like ERP and CRM. With such software, enterprises do not need to develop custom code but might require integrating the software with peripheral systems. However, delays in buying and delivering hardware can slow down the deployment process.

Figure 3-31 Traditional IT-based application lifecycle

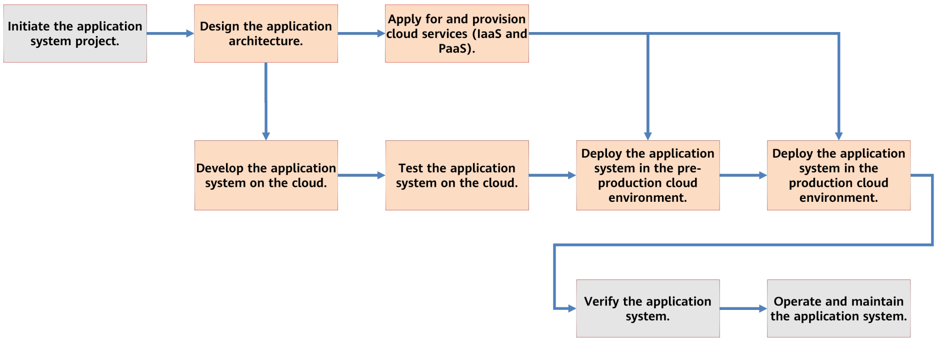

Initiate the application system project.
Design the application architecture.
Procure the hardware and software.
Install and configure the hardware devices.
Install and configure the essential software.
Develop the application system.
Test the application system.
Deploy the application system in the pre-production environment.
Deploy the application system in the production environment.
Verify the application system.
Operate and maintain the application system.

The basic process of constructing cloud computing-based application systems does not change too much. Yet, the processes (marked in light yellow in the above figure) in the application lifecycle need adjustments to fit cloud computing features and maximize its benefits.

The cloud platform has a unified resource pool that offers IaaS and PaaS services for application systems. This means you do not have to buy hardware devices or basic software like operating systems and virtualization software separately. Instead, you can simply buy and provision the IaaS and PaaS services from the cloud platform to set up test, pre-production, and production environments. This saves time on procurement and delivery. The steps, in the preceding figure, for software and hardware procurement, hardware installation and configuration, and essential software deployment and configuration, can be included in the cloud service application and provisioning process in the cloud computing environment.

Additionally, the cloud platform provides a cloud native DevOps development toolchain, like Huawei Cloud's CodeArts and open APIs. These tools and APIs enable you to develop and test application systems, and use pipelines to directly deploy application systems in the pre-production and production environments on the cloud. Such application systems are developed, tested, deployed, and run on the cloud, making them cloud-native applications. In this way, the value of cloud computing can be fully utilized. The following figure shows the lifecycle of cloud computing applications.

Figure 3-32 Cloud computing-based application lifecycle

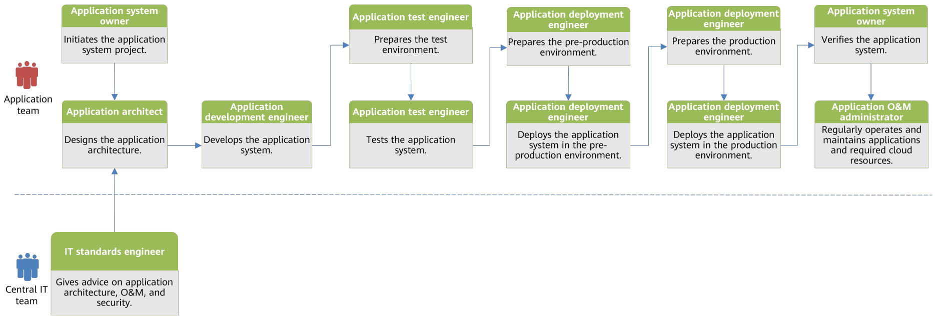

Initiate the application system project.
Design the application architecture.
Apply for and provision cloud services (IaaS and PaaS).
Develop the application system on the cloud.
Test the application system on the cloud.
Deploy the application system in the pre-production cloud environment.
Deploy the application system in the production cloud environment.
Verify the application system.
Operate and maintain the application system.

Cloud computing-based application lifecycle management must match the cloud operating model described in the preceding chapters. The roles of the CCoE (or central IT team) and application teams vary across different models, affecting their responsibilities and how they work together. As a result, the application lifecycle management process varies by cloud operating model.

Decentralized Operating Model

The following figure shows the application life cycle management process under the decentralized operating model. The application team manages all stages of an application's lifecycle. The central IT team formulates unified IT standards. They offer advice on application architecture, operations, maintenance, and security during application system design and oversee their execution. Yet, the team cannot enforce the application system on the compliance with these standards.

Figure 3-33 ALM in the decentralized operating model

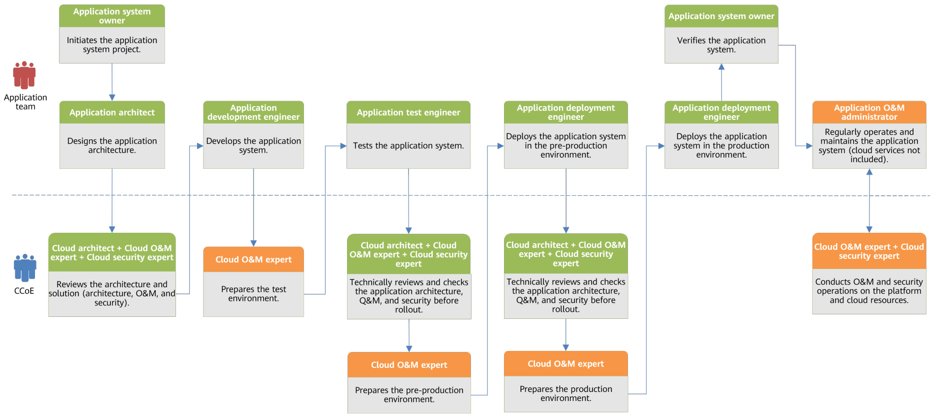

Application system owner
Initiates the application system project.
Application team
Application architect
Designs the application architecture.
Application development engineer
Develops the application system.
Application test engineer
Prepares the test environment.
Application test engineer
Tests the application system.
Application deployment engineer
Prepares the pre-production environment.
Application deployment engineer
Deploys the application system in the pre-production environment.
Application deployment engineer
Prepares the production environment.
Application deployment engineer
Deploys the application system in the production environment.
Application system owner
Verifies the application system.
Application O&M administrator
Regularly operates and maintains applications and required cloud resources.
Central IT team
IT standards engineer
Gives advice on application architecture, O&M, and security.

Centralized Operating Model

The following figure shows the application life cycle management process under the centralized operating model. The CCoE team deploys, operates, and maintains the cloud resources required by the application system. This allows the application team to concentrate solely on designing, developing, testing, deploying, and maintaining their applications without worrying about infrastructure or cloud resource deployment and management. Cloud architects, O&M experts, and security experts from the CCoE team participate in the application architecture design phase. They review the architecture, O&M, and security of the application team's design solution to ensure that the solution complies with the design principles and best practices of cloud technologies, maximizing the value of cloud computing. Cloud architects, O&M experts, and cloud security experts technically review and check the application system before deploying the application system in pre-production and production environments. This ensures the implemented solution matches the original design and prevents deviations between design and development phases. Through the reviews and checks, the CCoE team will be able to help the application team significantly improve the resilience, security, and performance of the application system on the cloud.

In the application O&M phase, the CCoE team handles daily O&M and security operations for cloud resources and the platform layer. The application team manages routine O&M and ensures security operations at the application layer, including measures like blocking SQL injections.

Figure 3-34 ALM under the centralized operating model


Application team
Application system owner
Initiates the application system project.
Application architect
Designs the application architecture.
Application development engineer
Develops the application system.
Application test engineer
Tests the application system.
Application deployment engineer
Deploys the application system in the pre-production environment.
Application system owner
Verifies the application system.
Application deployment engineer
Deploys the application system in the production environment.
Application O&M administrator
Regularly operates and maintains the application system (cloud services not included).
CCoE
Cloud architect + Cloud O&M expert + Cloud security expert
Reviews the architecture and solution (architecture, O&M, and security).
Cloud O&M expert
Prepares the test environment.
Cloud architect + Cloud O&M expert + Cloud security expert
Technically reviews and checks the application architecture, Q&M, and security before rollout.
Cloud architect + Cloud O&M expert + Cloud security expert
Technically reviews and checks the application architecture, Q&M, and security before rollout.
Cloud O&M expert + Cloud security expert
Conducts O&M and security operations on the platform and cloud resources.
Cloud O&M expert
Prepares the pre-production environment.
Cloud O&M expert
Prepares the production environment.

Enablement and Collaborative Operating Model

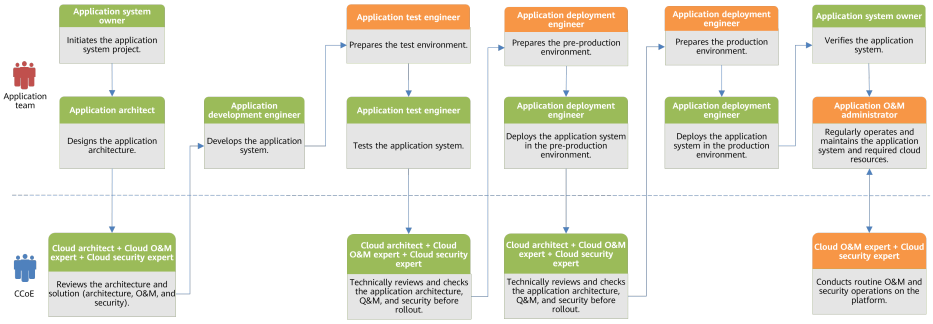
In this model, the application lifecycle management process remains similar to the centralized operating model. It includes all phases, though the owners and duties of some phases are changed. The phases marked in yellow in the following figure are changed.

In this model, the CCoE team enables the application team to take full responsibility for the deployment and O&M of cloud resources. This reduces the burden of the CCoE team, boosts the application team's independence, and enhances the application system's flexibility. To prevent inconsistent standards for independent cloud resource deployment and O&M by each application team, the CCoE team must formulate IT standards and ensure all teams follow them.

The CCoE team and application team must work closely to ensure secure and stable running of service systems on the cloud. In terms of O&M, the CCoE team performs routine O&M of the cloud platform or cloud IT infrastructure, and the application team manages routine O&M of applications and required cloud resources. If there is a service fault, the two teams work together to locate and rectify the fault. In terms of security operations, the CCoE team handles platform-level security and security operations. Application teams manage application system security and protect related cloud resources, including defenses against threats like SQL injections.

Like the centralized operating model, the cloud architects, O&M experts, and security experts from the CCoE team participate in the application architecture design phase. They review and check the application system before deploying the application system in pre-production and production environments.

Figure 3-35 ALM under the enablement and collaborative operating model

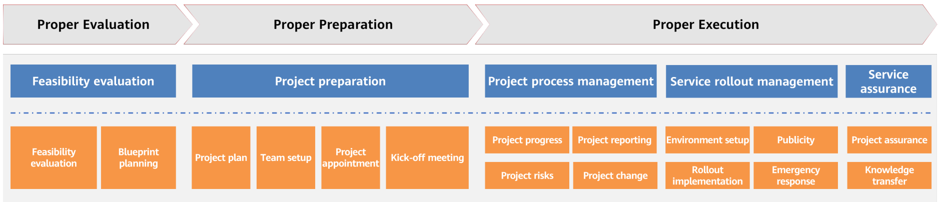

Application team
Application system owner
Initiates the application system project.
Application architect
Designs the application architecture.
Application development engineer
Develops the application system.
Application test engineer
Prepares the test environment.
Application test engineer
Tests the application system.
Application deployment engineer
Prepares the pre-production environment.
Application deployment engineer
Deploys the application system in the pre-production environment.
Application deployment engineer
Prepares the production environment.
Application deployment engineer
Deploys the application system in the production environment.
Application system owner
Verifies the application system.
Application O&M administrator
Regularly operates and maintains the application system and required cloud resources.
CCoE
Cloud architect + Cloud O&M expert + Cloud security expert
Reviews the architecture and solution (architecture, O&M, and security).
Cloud architect + Cloud O&M expert + Cloud security expert
Technically reviews and checks the application architecture, Q&M, and security before rollout.
Cloud architect + Cloud O&M expert + Cloud security expert
Technically reviews and checks the application architecture, Q&M, and security before rollout.
Cloud O&M expert + Cloud security expert
Conducts routine O&M and security operations on the platform.

3.9 Cloud Project Management

To ensure smooth and effective cloud migration, you need to run it as a standard project, and define clear objectives, scope, progress, costs, and quality for the project. Cloud migration is a long-term, complex, systematic project that impacting organizations, processes, and technology. Using effective project management and plans ensures better efficiency, quality, and success in achieving cloud migration objectives.

The cloud migration project management includes feasibility assessment, project preparation, project initiation, project process management, service rollout management, and service assurance. Based on Huawei Cloud project management and delivery experience, the following method is used to manage and control the entire cloud migration of an enterprise. The following describes each part of the cloud migration management process.

Figure 3-36 The management process for a cloud migration project



Feasibility Evaluation

Before starting cloud migration, company decision-makers may want to know how cloud migration can help their organization and whether it will benefit their short- or long-term plans. To obtain the information, the decision-makers usually ask the IT department to evaluate the value and benefits of cloud migration. Yet, the IT team typically relies on conventional IT practices and on-premises data center technologies, with limited knowledge or experience in cloud solutions. In this case, cloud migration experts are needed to assist the IT team in assessing the cloud value and benefits. This phase is called feasibility evaluation and blueprint planning phase. In this phase, cloud migration experts take the lead and the IT department assists them in evaluating the current services, organizations, operations, platforms, security, and O&M. Based on the evaluation conclusion and gap analysis, they estimate the value and benefits that will be brought by cloud migration, and design the post-migration overall blueprint. By doing all these evaluations, the company decision-makers will have a clear view of how cloud migration benefits their short- and long-term business goals, and to what extent the cloud platform will improve service continuity, agility, and innovation capabilities. This will help the decision-makers make scientific and reasonable decisions faster.

Project Preparation

Once the decision-makers approve the cloud migration, the project moves into its preparation stage. Project preparation defines the project objective, scope, plan, management mechanism, and acceptance criteria, and sets up a project team. In this phase, define the cloud migration's scope and goals with the customer. Set up a joint project team based on the impacted organizational areas. Also, coordinate with relevant organizations early about the project schedules, responsibilities, roles and involvement stages, and key tasks. Prepare the project plan before the kick-off meeting. Confirm the schedule, engaged personnel roles, and resources with relevant departments like the service department. The project management system ensures smooth operations. It covers regular meetings, risk control, changes, and reporting processes. The system helps projects run smoothly even in challenging situations. Clarify the project acceptance plan early by defining acceptance cases, metrics, and criteria based on project goals and service needs. This ensures the service system's functions and performance meet requirements after migration to the cloud. The service department and users should agree on the core processes and key metrics beforehand, with the service department providing the final acceptance metrics. Every project relies on people to push forward. Set up the CCoE team during the project preparation phase. Refer to **CCoE** for detailed steps on preparing and organizing this team.

After the project preparations are done, a formal project kick-off meeting must be held. The kick-off meeting officially starts the cloud migration project by setting clear objectives, plans, organizations, appointments, supervision, and evaluation criteria. It ensures team members follow their responsibilities and the project plan to meet its objectives. All CCoE members and the cloud service provider's project team need to attend the kick-off meeting.

An important and key part of the kick-off meeting is to assign roles and responsibilities and grant necessary permissions, and allocate KPIs to each member. It makes sure tasks for team members are proper, clear, and measurable, and maintains staff stability and increases their motivation to achieve objectives.

In addition to assign roles and responsibilities and grant necessary permissions, the project reporting and supervision mechanism must be clarified at the kick-off meeting. A project has multiple implementation phases. Can the result of each phase meet the expectation? Are there any bottlenecks and problems? Does the project team have the resources and capabilities to handle these problems? These are the problems that may be faced by the project team during the execution. Solving these problems quickly and effectively relies on the team's ability to understand the problems and manage resources. Yet, the team alone cannot resolve every problem. In this case, regular meetings and reporting to senior management are essential. In project delivery, agile project management is recommended. This includes daily stand-up meetings and weekly meetings to spot issues early, address them promptly, and reduce delays. Daily stand-up meetings and weekly meetings quickly raise project problems to decision-makers, who then allocate resources promptly to resolve these problems. This process forms the quality supervision mechanism described earlier. The system gathers top-level company expertise to guarantee successful projects, ensuring effective and quality outcomes.

In addition to the functions mentioned above, the kick-off meeting also sets up the daily operation management mechanism (daily and weekly reporting, issue escalation, etc.), risk change mechanism (personnel change, timeline change, environment change, etc.), and cross-team division and cooperation mechanism. You can manage these using standard project management practices.

Project Process Management

This phase includes project progress management, reporting management, risk management, and change management. We have mentioned the key steps of progress management and reporting management. For example, agile management (daily stand-up meetings and weekly meetings) can be used to help keep goals aligned and track the project schedule effectively. Regular updates to senior management enable swift monitoring on progress and risks, ensuring quick resolution of issues and obstacles. Cloud migration projects often face several challenges, including schedule risks, personnel change risks, technical feasibility risks, operation risks, and security risks. The following describes risk management (including change management) and agile management methods.

Cloud migration projects often face delays caused by unexpected issues like new service releases, critical database failures, or virus attacks. To avoid these, the project team must carefully assess potential risks at each phase, plan a proper schedule, and prepare backup plans for some extreme risks to finish the project on time.

Personnel changes often pose risks during a cloud migration. It is necessary to back up necessary roles before the project starts. Key roles held by one person require careful change management. If a company has only one database administrator (DBA), the project manager should create a backup plan before the project starts. In specific scenarios, personnel can be backed up across departments or personnel be reserved in advance. This issue affects both cloud migration projects and the long-term sustainability of core business operations.

The technical risks are manageable. The project team can test feasibility using POC verification. This checks if the functions support current services and if performance metrics like speed, latency, and throughput meet service running needs. In addition, for technical risks during migration, the project team should conduct migration and cutover drills to identify potential risks and issues. They can then create a runbook to address and prevent these risks effectively.

Cloud migration projects face different operating risks compared to traditional IT projects. Traditional IT projects rely on hardware platforms, needing one operator with several supervisors to achieve desired outcomes. Cloud migration operations rely on networks, where services and platforms are highly integrated and a single mistake can impact many components and services. To reduce human errors, automate cloud operations during service system transitions and rollouts. In short, use tools for automation when available, choose scripts only if no suitable tools are available, and avoid manual operations whenever possible.

Cloud migrations demand strict security measures. Teams must follow the "no vulnerable systems to the cloud" rule by thoroughly checking and scanning systems beforehand. This includes reviewing hardware, software, middleware, application status, logs, events, alarms, and using security tools to confirm the system runs smoothly and has no risks.

Cloud migration projects avoid the complexities and lengthy timelines tied to integrating multiple hardware vendors and software providers seen in traditional IT projects. Yet, cloud migrations feature broad involvement, high platform integration, making issue resolution challenging. This leads to potential centralized bottlenecks and risks in managing these projects. Even one missing feature can stall the whole project timeline. Traditional waterfall project management struggles with cloud migration projects, making agile project management more suitable.

The preceding chapters have briefly described the methods for handling project problems, such as daily stand-up meetings and weekly meetings, to quickly review, streamline, and remove project bottlenecks. Fundamentally, these methods are part of agile project management. Agile project management works backward through stages. It sets clear goals for each phase, reviews ongoing progress, identifies challenges, and finds quick solutions aligned with those goals. A cloud migration project manager should clearly know the goals for each phase and stay aligned with those goals. To achieve the goals, they should identify the bottlenecks and find solutions as early as possible.

Agile project management needs to be implemented based on agile management tools. Combining these tools with agile processes creates efficient, fast-closing loops for managing tasks effectively. Common agile management tools include Jira. The cloud native project management tool **CodeArts Req** provided by Huawei Cloud is also recommended. CodeArts Req integrates with other CodeArts tools and cloud native tool chain DevOps provided by Huawei Cloud, boosting end-to-end project management and application delivery efficiency.

Service Rollout Management

Service system rollout management aims to ensure that services can still run smoothly during service system rollouts and reduce or eliminate the impact and risks on services. Service system rollout management involves preparing environments, promoting awareness, handling risks, and executing smooth cutovers.

Before the service system cutover, the cloud environment preparation usually includes deploying the service environment, synchronizing data (like during migrations), configuring connections with related systems, and checking both internal and external connectivity. These preparations create the necessary foundation for the smooth running of the service system.

Publicizing the rollout is a key for companies moving to the cloud for the first time. Publicity activities ensure all team members and stakeholders know their roles and work together to support smooth service rollouts. For example, let them know the impact of service rollouts, roles and responsibilities, implementation contents, timelines, and problem feedback mechanism. Verify every step and metric to guarantee a successful system rollout. In addition, enterprise executives can convey a key message to their employees that cloud migration is essential for the company's future and everyone need to embrace this shift and get ready for the digital transformation.

Risk emergency preparation is a necessary step before each service rollout. Identify risks and problems that may be encountered and formulate solutions. Identifying risks goes beyond technical issues. It also includes systematic risks in terms of organization, process, security, and platform. For example, a system that has been running for years may face risks like irreparable hardware failures, harmful viruses in its environment, or missing critical roles during rollouts. These risks can severely disrupt the system running. To reduce their impacts, identify risks early, prepare backup plans, and conduct necessary drills.

The service rollout is the final and most crucial step. With proper preparation, thorough risk management, and completed verifications, this process typically runs smoothly. This phase involves verifying the system using the updated manual and assessing if the rollout meets success criteria. A key point is that this phase is personnel-intensive. All personnel need to assume their responsibilities in different phases based on the publicity requirements, execute related operations based on the standard requirements, verify related processes and results, and ensure accountability by signing off on acceptance criteria. Then all feedback data will be used to determine whether the service rollout is successful.

Service Assurance

After the service system is rolled out, it enters the assurance period for resolving issues and completing knowledge transfer. The assurance period is usually one week after each rollout. During this time, issues often arise the most frequently. The cloud migration project team must prioritize and manage this critical phase. A dedicated assurance team from the cloud service provider collaborates with the enterprise to maintain stable operations. In this phase, the problems raised by the service department are sorted by urgency and severity, then resolved according to their priority. Knowledge transfer involves training the service department's application O&M team on cloud technologies after the service system is rolled out. It equips them with the skills needed to manage daily O&M tasks and handle incidents effectively on the cloud platform.

3.10 Anti-patterns Against Top-Level Planning

During top-level planning, some common anti-patterns may hinder cloud migration. Identifying and avoiding these anti-patterns is crucial to ensure the success of cloud migration. The following are some common anti-patterns and corresponding suggestions.

The Incomplete CCoE Team

The centralized CCoE team is established by an enterprise for cloud migration. It guides the entire cloud migration journey. It provides best practices, guidance, and resources to boost cloud benefits and achieve smooth cloud migration. CCoE is like the engine of cloud migration. Without essential members, the team will not function effectively, which may even lead to migration failures. The following are examples of this anti-pattern:

- **No cloud architects**: Without professional cloud architects, cloud migration will be like a boat without a rudder. A poorly designed architecture limits system expansion and makes maintenance challenging. This creates a fragmented cloud setup with low resource usage. Choosing the wrong technology can cause poor performance, higher costs, and security issues. This also makes it challenging to integrate existing systems with cloud platforms and make full use of cloud native features. As a result, enterprises struggle to benefit from cloud advantages and even encounter potential security and performance challenges.
- **Insufficient engagement of the application team**: Application teams are the end users of cloud platforms. If they do not fully participate in the work of CCoE, the platform may fail to meet real business needs. Without the support from the frontline personnel, cloud migration designs might fail to address real needs. Poor collaboration with the application team may delay

the migration process. Without the application team, promoting DevOps and automation becomes difficult, stifling innovation and blocking the benefits of cloud migration.

- **No cloud governance experts**: Cloud governance experts are like "housekeepers" of cloud environments. Without them, enterprises face unregulated cloud use, higher costs, and greater security threats. Companies struggle with compliance and risk legal issues without professional governance policies and measures. In addition, without effective monitoring and management mechanisms, problems cannot be detected and resolved in a timely manner, affecting business stability.

For details about how to establish a functional CCoE team, see **CCoE**.

Starting Migration Without Setting Up a Landing Zone

A landing zone is designed to create a multi-account runtime environment on the cloud that features robust architecture, security, compliance, and scalability. It includes the cloud backbone network, IAM, compliance audit, resource organization, and governance policies. Migrating services to the cloud without establishing a landing zone, a unified infrastructure and governance framework, is risky. This is a common anti-pattern, which is like building a house without a solid foundation. This will lead to multiple issues, slowing down the project progress, raising expenses, and potentially compromising safety. To avoid this, plan and deploy a landing zone before migrating any service system. For details, see **Landing Zone Design**.

Inappropriate Cloud Operations Model

Inappropriate cloud operations models are common anti-patterns against cloud migration. A right cloud operations model directly promotes resource management, improves efficiency, ensures security compliance, and controls costs on the cloud. Selecting an unsuitable cloud operations model without evaluating your business requirements, structure, capabilities, or objectives may result in inefficiencies, increased security vulnerabilities, budget issues, and complicated management. This prevents you from gaining the benefits of migration to the cloud. The following are examples of this anti-pattern:

- **Centralized operations model for fast-changing services**: If an enterprise needs to quickly respond to market changes but adopts the centralized operations model, all resource requests and changes need to be approved by the CCoE team. The enterprise may struggle to adapt swiftly to market shifts and often lose valuable business opportunities.
- **Decentralized operations model for businesses with high compliance requirements**: If an enterprise with a mature and stable business system, especially in industries like finance that demand high security and compliance standards, operates under a decentralized operations model where departments function independently, maintaining consistent security measures becomes challenging, raising compliance risks.
- **Empowerment and collaboration operations model for small enterprises with limited resources and budgets**: Building and maintaining complex IT systems and governance policies under this model demands significant

resources. Small businesses with tight budgets often find it too costly and complicated.

4 Surveys

4.1 Overview

The cloud migration survey is not a one-time effort. It needs to be performed for multiple times during the migration to the cloud. The information to be collected varies in each phase. This section describes the approach of survey analysis, which can be used as a reference in any phase of cloud migration. Companies that do not lead their cloud migration can follow this approach to cooperate with third parties. As for the information required in a phase, collect as much of them as possible in a single survey, so as to reduce the number of surveys, especially interviews.

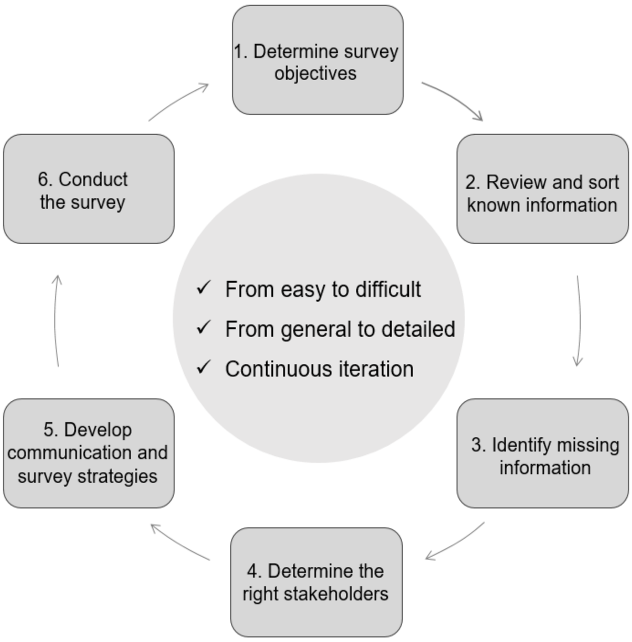
- **Basic environment survey**: The survey is conducted before the cloud architecture design, collecting information about the overall IT technical architecture and IT governance status and requirements.
- **Application survey**: The application survey is performed throughout the cloud migration process. In the evaluation and planning phase, only the service overview is required. In the migration pilot and large-scale cloud migration phases, the detailed technical architecture of each application system needs to be analyzed, and the details about the technical components of each application system needs to be collected, including the versions and parameters of the components.
- **Big data survey**: Survey the overall big data technical architecture and then gradually analyze details.

A survey involves the following steps:

1. Determine the survey objectives and sort out the information to be surveyed in the current cloud migration phase.
2. Align existing information to avoid repeated survey.
3. Based on the survey objectives, identify the information to be collected, why the information is required, and how to collect the information.
4. Determine the stakeholders who can provide the information based on the company's organizational structure and division of responsibilities.
5. Develop the survey interview outline and survey template, and develop the communication strategies and plan.

6. Obtain the required information based on the authorization mode recognized by stakeholders, sort out the information, and complete the survey.

Figure 4-1 Survey methods


1. Determine survey objectives
2. Review and sort known information
3. Identify missing information
4. Determine the right stakeholders
5. Develop communication and survey strategies
6. Conduct the survey
✓ From easy to difficult
✓ From general to detailed
✓ Continuous iteration

A survey should start with easy methods and general information, followed by complex methods and detailed information. It should be continuously iterated. The details are as follows:

- **Start with easy survey methods.** Select simple and quick survey methods to start with.
- **Start with general survey content.** The information obtained in the evaluation and planning phase is general, whereas that obtained in the implementation phase is the most detailed.
- **Continuously iterate the survey process.** A survey is a continuous effort. It needs to be continuously iterated, especially in a large-scale migration. A survey on details can be conducted in multiple stages based on the migration phases.

4.2 Establishing a Survey and Evaluation Team

An efficient survey and evaluation team is critical to enterprise cloud migration. The team is responsible for conducting detailed surveys; evaluating the existing IT infrastructure, service requirements, and cloud migration benefits of companies; and ensuring the effectiveness and feasibility of cloud migration strategies. The survey and evaluation team consists of members from different departments. A company can refer to the preceding CCoE organizational structure and role responsibilities to set up a well-rounded professional team. Necessary team members include:

- **Survey and evaluation engineers**: assigned by the IT director. They are from the IT department. They are responsible for surveying and evaluating the existing IT infrastructure, service systems, application architecture, data storage, and security policies, including the hardware configuration, network architecture, software version, and dependency. They should also analyze the compatibility between these facilities and cloud services, the migration difficulty level, the feasibility of migrating the existing systems to the cloud platform.

- **Service experts**: assigned by service directors. They are from service departments. They are responsible for collecting and analyzing service requirements, interviewing stakeholders, understanding the requirements and expectations of different departments, analyzing service value, providing teams with guidance and suggestions from the service perspective, and helping to quantify the benefits of cloud migration.
- **Application architects**: assigned by the service director. They are from the application team of the service department. They help the cloud implementation team survey the status of the service systems. They provide the survey and evaluation team with the resource status, application architecture, deployment architecture, dependencies, and more information.
- **Financial experts**: assigned by the finance director. They are from the finance department. They are responsible for accounting and analyzing the costs of cloud migration projects, including cloud service fees, migration fees, and O&M fees; and evaluating the economic benefits and return on investment (ROI) of cloud migration projects. These can provide financial information support and suggestions for cloud migration decision-making.
- **Cloud security experts**: assigned by the IT director. They are from the security team of the IT department. They are responsible for evaluating the security, compliance, and data protection capabilities of cloud services; identifying and handling potential security risks, and ensuring compliance with relevant laws, regulations, and industry standards during cloud migration.
- **Project manager**: a member of the Project Management Office (PMO). This role manages the survey project progress, ensures in-time task completion, coordinates communication and collaboration between departments, promotes information flow, and resolves problems in the project in a timely manner.
- **Cloud architects**: A cloud expert from the IT department or cloud vendor. This role provides technical support and guidance for cloud migration, including migration methodologies and the best practices of a survey.

The way the survey and evaluation team works differs by scenario. If a company leads its cloud migration, the survey should be led by the preceding roles in the company, and the cloud vendor should provide necessary technical support. If a company purchases third-party professional services and the third party leads the cloud migration, the third party should lead the survey, and the company's team should provide related service and technical information.

4.3 Infrastructure Survey

The basic environment survey focuses on the current status of a company's IT infrastructure and cloud migration requirements, including resource information, networking, security architecture, O&M architecture, access permission control, and resource metering and billing. The survey is conducted based on the information exported from the IT system (such as CMDB, CMP, and virtualization management software) and questionnaires. The basic environment survey is conducted by the company's O&M team. The survey personnel work with the company's O&M team to collect data and information about the basic environment.

A common survey method is to export information from the internal IT system of a company, such as the configuration management database (CMDB), cloud management platform (CMP), or virtualization management software. These systems provide information about hardware devices, network topologies, OSs, applications, and related configurations and versions, helping survey personnel understand the company's IT infrastructure.

Questionnaires and interviews are also common survey methods. Survey personnel can design questionnaires to collect information about resource usage, networking configuration, security architecture, O&M process, and access permission control. They can also conduct face-to-face interviews with the O&M team to deeply understand the actual company situation, challenges, and requirements.

During the survey, the survey personnel should fully communicate with the O&M team to ensure that the IT basic environment information of the company is accurately obtained and understood. They should properly handle sensitive and confidential information, and comply with the security and confidentiality requirements of the company.

Based on the survey results, the company can better understand the IT infrastructure status and cloud migration requirements, and make targeted planning and decisions. By evaluating resource usage, networking configuration, and security architecture, the company can formulate appropriate cloud migration policies, optimize resource configuration, improve O&M efficiency, and ensure access permission control and accurate resource metering and billing.

In a word, the basic environment survey aims to deeply understand the status quo of the company's IT infrastructure and cloud migration requirements. By working with the O&M team, the survey personnel collect required information from the IT systems, questionnaires, and interviews. Such surveys can provide valuable reference for future decision-making on company's IT planning and migration.

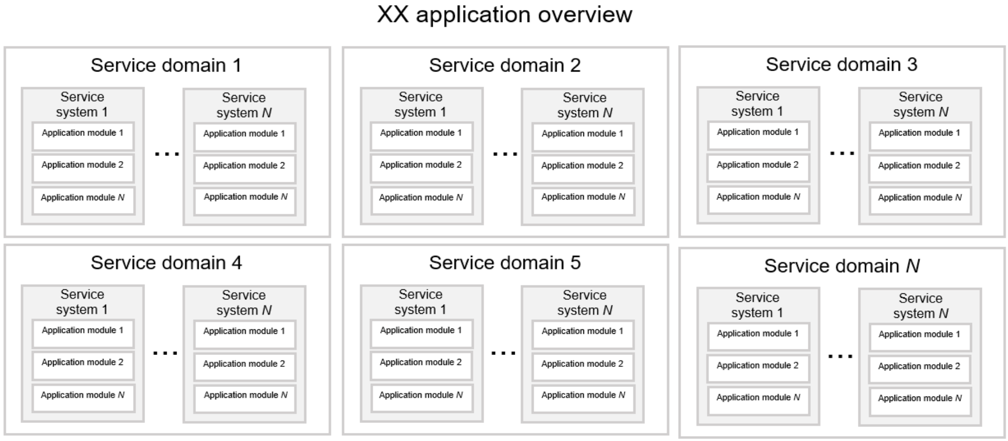
4.4 Application System Survey

4.4.1 Survey Application Overview

An application migration survey should start with general information collection, followed by detail collection. It should be continuously iterated and lasts for the entire cloud migration period. In the early stage, the survey focuses on the application overview. In the migration phase, each application needs to be surveyed thoroughly to learn the detailed deployment architecture and component information. The application survey needs to be conducted by the application architects and application O&M administrators of all the service domains.

The survey of the application overview is conducted in the evaluation and planning phase. Generally, the applications are surveyed layer by layer in the following sequence: service domains, service systems, and application modules, as shown in the following figure.

Figure 4-2 Application overview example



The application overview survey methods, arranged from easiest to hardest, are as follows:

- **Knowledge base**: Some companies have well-developed knowledge bases and already have documents describing the application overview. In this case, you can directly obtain the information. Note that the document information in the knowledge bases may be outdated. Align and confirm with the service owners on it.
- **CMDB**: Some companies' CMDB systems have information about all applications. You can export application information from the CMDB, classify the information by service domain and service system, and align and confirm the information with the service owners.
- **Observable platform**: Some companies have built application observability platforms, such as Datadog and Huawei Cloud APM. On these platforms, you can check the invocation relationships between applications and develop the application overview.
- **Survey and interview**: Interview the service domain owners and record the system and application information of the domains. Develop an overview after completing all the service domain surveys.

4.4.2 Application Deployment Architecture Survey

The application deployment architecture survey is conducted during the pilot migration or large-scale migration. The survey is based on a single application, and mainly focuses on four layers, that is access layer, application layer, middleware layer, and data layer, of the deployment architecture. The survey also focuses on the detailed information about technical components, such as the specifications, version, and capacity, at each layer. The specific survey content is as follows:

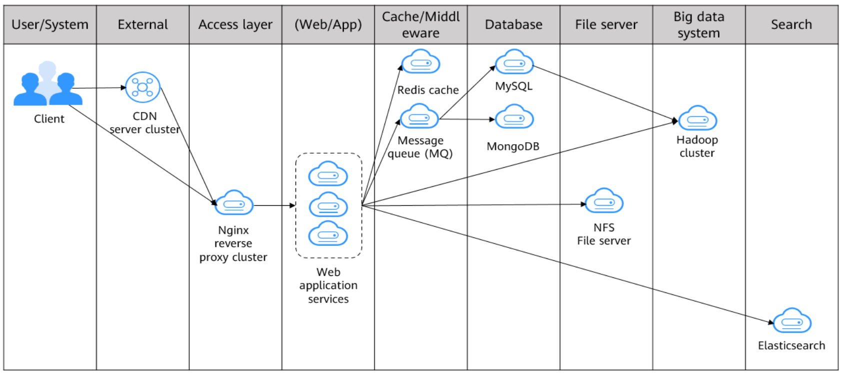
Four Layers of the Application Deployment Architecture

Collect detailed information about the access layer, application layer, middleware layer, and data layer, and collect three types of associations (shared data, shared server, and application interaction communication dependency). You can refer to the following table to collect the details about the deployment architecture of an application.

Table 4-1 Application survey table

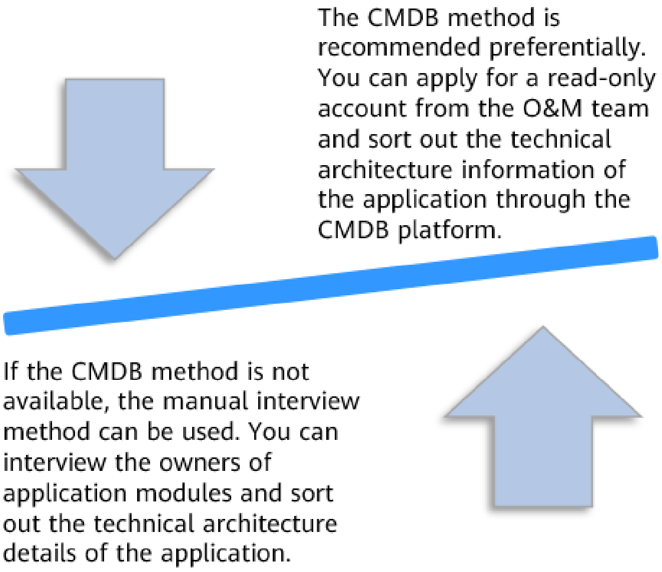
Application Type	Access Layer		Application Layer		Middleware Layer			Data Layer		Access Domain Name		Remarks
Application name	NAT	NGINX	Host quantity	IP addresses	Redis	Kafka	MQ	MySQL	Mongo	Internal/External domain names	WAF	Remarks

You can also refer to the following figure to draw the application deployment architecture.



The following figure shows the survey methods.

Figure 4-3 Survey methods



The CMDB method is recommended preferentially. You can apply for a read-only account from the O&M team and sort out the technical architecture information of the application through the CMDB platform.

If the CMDB method is not available, the manual interview method can be used. You can interview the owners of application modules and sort out the technical architecture details of the application.

Technical Component Details

This section describes the details about each technical component (including hosts, databases, and middleware) involved in the deployment architecture of an application. The details include the resource specifications, version, capacity, and configuration, as shown in the following table.

Table 4-2 Host information survey example

Host Name	Host Type (ECS/Physical machine)	Specifications	CPU (Core)	Memory (GB)	OS Version	System Disk Type	System Disk Size (GB)	Data Disk Type	Data Disk Size (GB)	Private IP Address	Public IP Address
The table header information provided here is for reference only. Fill in the table with real business data.											

Table 4-3 Database information survey table

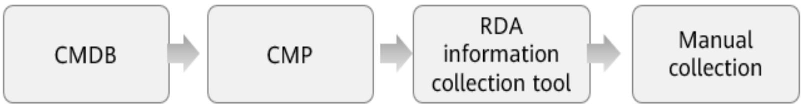
Application Name	Region	Instance Name	Architecture Type	IP Address:Port	Version	Instance Specifications	CPU	Memory	Storage Type	Disk Capacity
The table header information provided here is for reference only. Fill in the table with real business data.										

Table 4-4 Middleware information survey table

Application Name	Region	Name	Version	Connection Address	Specifications	Topics	Partitions
The table header information provided here is for reference only. Fill in the table with real business data.							

The following figure shows the survey methods.

Figure 4-4 Survey methods

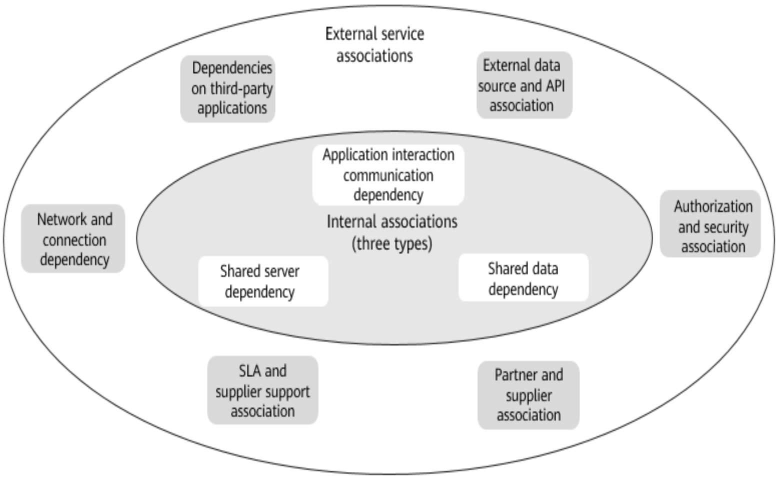


1. The CMDB method is preferred.
2. If CMDB is unavailable, we can also obtain information from the cloud management platform (CMP) or virtualization management software on the live network.
3. If neither of them (CMDB and CMP) is feasible, we can install an information collection tool (for example, Huawei Cloud RDA) to collect information.
4. If none of the preceding methods is feasible, we can collect information through interviews.

4.4.3 Application Association Survey

When applications of an enterprise are migrated to the cloud, both internal and external correlations of the enterprise need to be considered. Based on internal associations, migration waves are planned, and switchover solutions are formulated. Based on external correlations, migration impact is evaluated, proper time windows are selected for shutdowns, and switchover solutions are formulated.

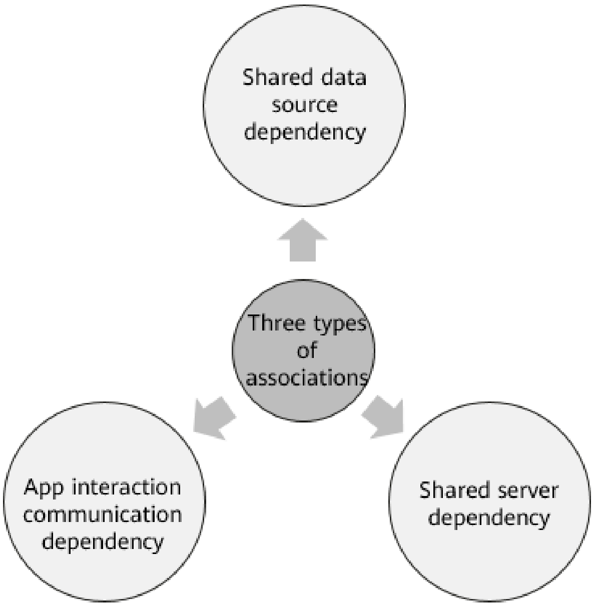
Figure 4-5 Association relationship survey diagram


External service associations
Dependencies on third-party applications
External data source and API association
Application interaction communication dependency
Network and connection dependency
Internal associations (three types)
Authorization and security association
Shared server dependency
Shared data dependency
SLA and supplier support association
Partner and supplier association

Internal Association Survey

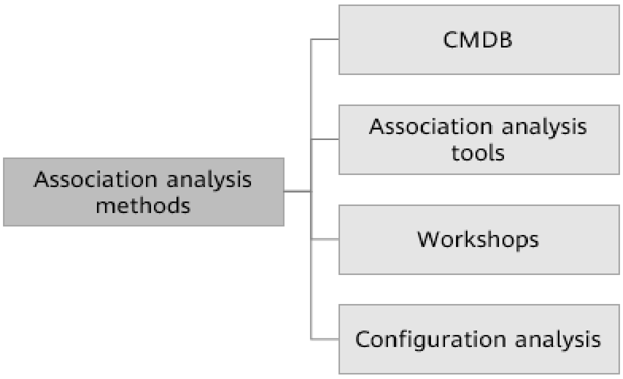
Association analysis is an important input for batch planning and solution switchover. It is also a difficult point in cloud migration. The following figure shows the three types of associations that affect cloud migration.

Figure 4-6 Three types of association relationships


Shared data source dependency
Three types of associations
App interaction communication dependency
Shared server dependency

There are four association analysis methods. During cloud migration, enterprises can select a proper analysis method based on their actual conditions.

Figure 4-7 Association analysis methods



- **CMDB method**: This method is applicable to customers who have a CMDB system. The CMDB system usually has dependencies between applications, between applications and databases, and between applications and middleware. You can directly obtain the dependencies from the CMDB system.

Figure 4-8 CMDB method

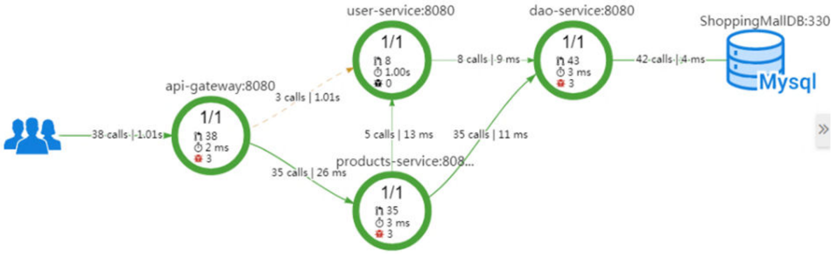
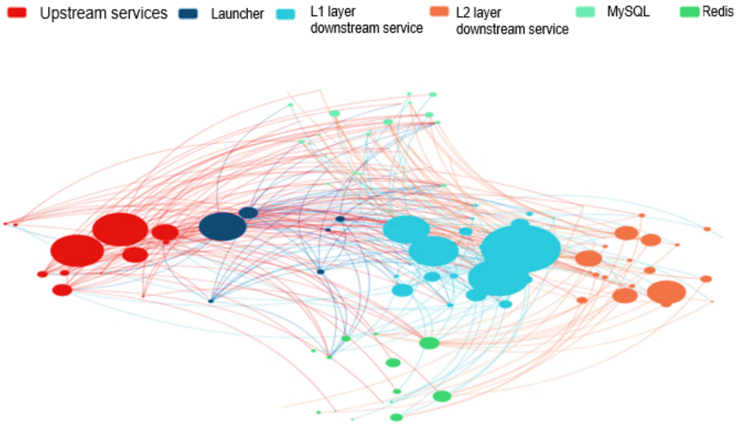


Figure 4-9 CMDB method example



- **Association analysis tool**: You can use dedicated association analysis tools, such as Huawei Cloud **Migration Center (MgC)**, or other association analysis tools in the industry.

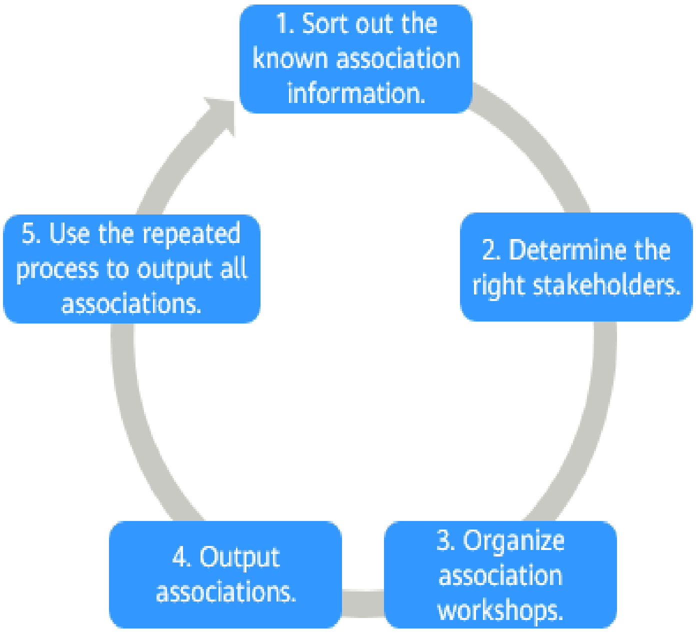
Table 4-5 Association analysis tool method

Software Name	Commercial Use	Description
Dynatrace	Yes	Dynatrace includes excellent tools for managing application performance and provides automatic application dependency mapping. It discovers and monitors microservices and applications, and even microservices and applications running in containers. It collects performance data and communication time data, and highlights poorly performing services and applications.
Cyberspace mapping	Yes	Cyberspace mapping is a tool for discovering applications and servers. It supports agentless automatic discovery on MS, Linux, Unix, cloud vendors, management programs, hardware, virtualization, and application layer. It is also applicable for remote data collection across multiple data centers.

Software Name	Commercial Use	Description
Device42	Yes	Device42 is a discovery and mapping tool used to collect and organize data across the entire IT environment. It enables device discovery, asset management, and dedicated application mapping and management. It can also check network devices, such as switches, load balancers, and power and environment devices, including PDUs, UPSs, and CRAC devices.
Airlync iSRG dynamic intelligent system	Yes	Airlync iSRG provides dynamic resource management for traditional networks, clouds, microservices, containers, and virtual systems. It enables resource discovery and collection, template management, resource management, and view editing.
ManageEngine Applications Manager	Yes	ManageEngine is an application manager tool. It is a monitoring tool for general-purpose servers and applications. It can monitor servers, databases, VMs, applications, web services, and other components.
Datadog	Yes	Datadog is a performance monitoring tool that provides application monitoring and mapping functions. It collects information from the entire infrastructure, including matching characteristic customers and tracing endpoints or errors. It automatically maps data flows and organizes services based on dependencies.
Pinpoint	Open-source	Pinpoint can trace transactions between distributed applications to check the overall structure and running status. Pinpoint can monitor applications in real time and understand the application topology clearly and rapidly.

- **Workshops** can be organized to guide key personnel who are familiar with the business system to sort out the associations.

Figure 4-10 Workshop method


1. Sort out the known association information.
2. Determine the right stakeholders.
3. Organize association workshops.
4. Output associations.
5. Use the repeated process to output all associations.

- **Configuration analysis** is a method that explores associations by analyzing configuration files of application systems. It helps us understand the mutual invocation relationships between applications, connections between applications and databases, and other associations. The following describes the basic procedure of the configuration analysis method:
  - **Collect configuration files**. You need to collect and obtain configuration files related to the target application system. This may include configurations of DNS, ELB, NAT, and Nginx.conf.
  - **Parse configuration files**: For each configuration file, you need to compile scripts or use existing tools to parse the content. The scripts can extract key information and process the information based on the file format and syntax rules.
  - **Extract associated information**. When a configuration file is parsed, you need to identify information related to other components or resources. For example, you can search for the mutual calling relationship between applications, such as the URL or API calling from one application to another application. You can also search for the connection information between applications and databases, such as the database addresses, usernames, and passwords.
  - **Create an association graph**. You can organize the extracted association information into a graph or an association model. The graph or model may be a directed graph, an undirected graph, or other proper data structures, and is used to represent associations and dependencies between applications.
  - **Analyze associations**. For the association graph created earlier, we can use the graph theory algorithm or other analysis methods to explore the associations. This helps us find hidden dependencies.

By using the configuration analysis method, we can gain a deeper insight into the internal association of application systems and better understand the overall architecture and operation mode. This is of great value to system migration. It should be noted that configuration files may be subject to changes and updates, so the configuration information must be updated and verified in a timely manner to ensure its accuracy during association analysis.

External Association Survey

Before migrating applications, you need to survey and evaluate the following common external associations. Ensure that you fully understand the external dependencies of applications and take appropriate measures to ensure that external services are running properly after the migration.

- **Dependencies on third-party applications**

  Identify third-party applications and services, including their versions and integration modes, related to the target application. Determine whether these dependencies need to be adjusted or reconfigured.

- **External data source and API dependencies**

  Analyze and record the external data sources and APIs on which the target application depends, such as external databases, file systems, message queues, or third-party services. Ensure that these dependencies can work properly after the migration.

- **Authorization and security associations**

  Identify the authorization and security associations related to the target application. These associations include external services and mechanisms related to identity authentication, access control, token management, and IP address whitelists.

- **Partner and supplier associations**

  If the target application is integrated with offerings from partners or suppliers, you need to investigate associations and ensure that they can work properly after the migration.

- **Service Level Agreement (SLA) and supplier support associations**

  Check existing SLAs and supplier support agreements and evaluate the impact of migration to the cloud platform on these associations. Ensure that the service requirements can be met and the expected support and services remain available in the cloud environment.

- **Network and connection dependencies**

  Identify the network connections and transmission protocols required by the target application. Determine whether network configuration and access control are required after migration to ensure that the application can communicate with external systems.

  The following methods can be used to survey external associations. You can use multiple methods to improve the efficiency and result integrity.

- **Documents and existing materials**

  Read existing documents and technical materials, including the application architecture diagrams, deployment descriptions, and O&M manuals. These materials can identify key dependencies and integration points of the application.

- **Communication with the development and O&M teams**

  Communicate with the application development and O&M teams to learn about their understanding of system dependencies. They may provide detailed information about the application, description of dependencies, and integration with other systems.

- **Code analysis**

  Carefully check the source code of the application, especially the external dependencies involved in the configuration file and code. Some dependencies may be specified by the code.

- **System scanning and monitoring**

  Use system monitoring tools and network scanning tools to scan the entire system and identify dependencies related to applications.

- **Communication with related teams**

  Communicate with other departments or teams to understand the integration relationships between applications and offerings of other companies, suppliers, or partners. These relationships may include data sharing, interface invoking, and permissions control.

- **Service providers and documentation**

  If the application relies on external service providers, review their documentation, API references, and support resources for more information about the dependencies.

4.4.4 Application Cloud Migration Survey

The survey includes the following information:

- Migration time window
- Cutover time window
- Target architecture requirements (function, performance, availability, security, cost, scalability, and O&M)
- Rollback requirement
- Confirmation of associations between services

This survey can be conducted through manual interviews or workshops. This survey can be combined with the other two surveys to reduce survey times.

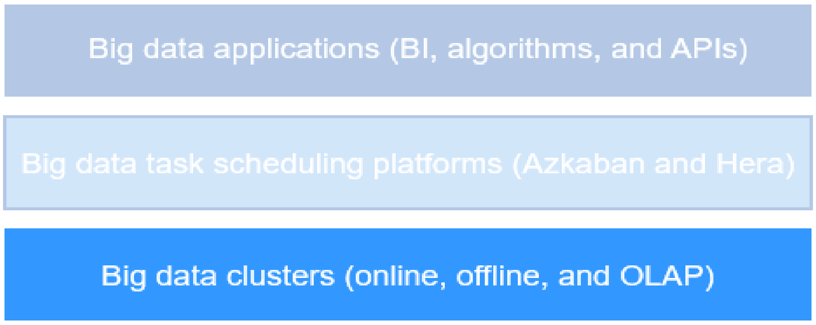
4.5 Big Data Survey

4.5.1 Platform Survey

Big Data Survey Overview

Big data migration is the process of migrating big data clusters, big data task scheduling platforms, and big data applications from one runtime environment to another.

Figure 4-11 Objects of big data survey


Big data applications (BI, algorithms, and APIs)
Big data task scheduling platforms (Azkaban and Hera)
Big data clusters (online, offline, and OLAP)

The following information needs to be surveyed for big data migration:

- Big data platforms, including big data clusters, task scheduling platforms, and data flows.
- Data information, including type, volume, metadata, permissions, and update frequency.
- Task information, including the type, number, and update period.

This section describes how to survey big data platforms, data, and tasks.

Platform Survey

The following describes the details.

- **Big data cluster survey**

  Survey the number and functions of big data clusters, services and data types processed by each cluster or component, components that process real-time or offline data and detailed version information, data format types and compression algorithms, data security and permission control, high availability (HA) and fault tolerance mechanisms, scalability, and elasticity.

  Survey the number and functions of big data clusters: For example, Hadoop, Spark, and Hive clusters serve as storage, computing, and query clusters based on service requirements.

  Survey the service scope of each cluster or component, types of data they process, and data transfer modes.

  Survey the components used to process real-time and offline data. For example, real-time data may be processed by Apache Kafka and Apache Flink, and offline data may be processed by Hadoop and Spark.

  Survey the data format types and compression algorithms:

  Survey the data security and permission control mechanisms of the platform, such as data encryption and user access permission management.

  Understand the HA and fault tolerance mechanisms of big data clusters, including fault recovery, backup policies, and disaster recovery (DR) solutions.

- **Big data task scheduling platform survey**

  Survey the big data task scheduling platform details for future selection and solution design, including the type, version, supported big data framework and technologies, scheduling task type, visualized and management GUI, scalability and integration, fault tolerance and recovery, security and permission control, community support, and documentation.

Survey the type of the platform, such as Azkaban. Understand its features and applications.

Survey the version of the platform and learn the function updates and improvements of the latest version.

Check whether the platform supports the current big data framework and technologies, such as Hadoop, Spark, Hive, Pig, and Flink.

Survey the task types supported by the platform, including Jar tasks, SQL tasks, and script tasks (Python and Shell).

Check whether a visualized and management GUI is provided by the platform for configuring, monitoring, and managing task scheduling.

Understand the fault tolerance mechanism of the platform, including the retry mechanism upon task failure and the fault recovery policy.

- **Data flow survey**

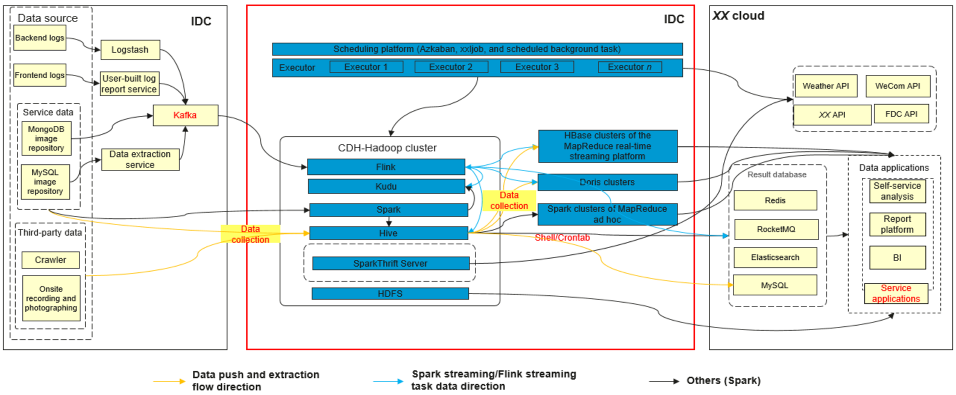
  The figure below shows the architecture and data flow diagram of the big data platform and services.

  Data access source of the platform.

  Data inflow mode, for example, real-time data reporting and batch data extraction.

  Analyze the data flow in the big data platform and between its components, such as data collection component types, the next layer after collection, data storage components, and data processing workflows.

Figure 4-12 Data flow example



4.5.2 Data Survey

The following describes the data survey.

Table 4-6 Data survey

Survey Content	Purpose	Example
Data type	Select a proper migration tool based on the data type.	HDFS, HBase, and MySQL
Data volume	Obtain historical data volume to evaluate the historical data migration period. Obtain daily incremental data to evaluate the daily incremental data synchronization period.	Historical data: *X* PB Daily incremental data: *Y* TB
Data layers	Survey the data layers to determine the migration priority and data verification standards.	Data access layer, intermediate layer, and result layer
Data permissions	Determine the permission data migration method based on the source data permission control component.	Sentry and Ranger
Data importance	Survey data helps identify core data from non-core data, setting migration priorities and data verification standards.	Core data: transaction data Non-core data: log data
Data update frequency	Determine the data migration plan and verification plan based on the update frequency.	Daily/weekly/monthly/real-time update
Task execution interval	Stagger peak hours of data migration, data verification, and services.	Offline task execution before and after work

The survey is performed through the current big data platform with surveys and interviews for supplement and confirmation.

4.5.3 Task Survey

The following describes the task survey.

Table 4-7 Task survey

Survey Content	Description
Task scheduling	For example, Azkaban, DolphinScheduler, Hera, and Crontab.

Survey Content	Description
Task type	The task is classified based on programming languages: ● Jar tasks: used in MRS, Flink, and Spark. ● SQL tasks: used in Hive, Spark, and UDF. ● Python tasks: used in Spark and algorithm scenarios. ● Others: used for script calling, such as Shell and Scala.
Task quantity	Survey the total number of all types of tasks to evaluate the task migration period and reconstruction workload. For example, there are 820 Jar tasks on the Azkaban task scheduling platform.
Task update period	Identify the task update periods of different scheduling platforms and task types. For example, the Jar tasks on the Azkaban scheduling platform update monthly, and Shell tasks on the XXL-Job platform update at 22:00 every day.
Task information	Obtain the details of all task for future task reconstruction and migration, including the task ID, name, responsible department, owner, execution time, and update period. Communicate with key personnel in a timely manner.
Task dependency	Identify key tasks and dependencies between tasks.

The survey is performed through the current big data platform with surveys and interviews for supplement and confirmation.

4.6 Survey Methods

There are various survey methods. Enterprises should select what is most appropriate for them in terms of survey efficiency, completeness, and authenticity. Generally, CMDB is most recommended. If there is missing information, use the cloud management platform or conduct surveys and interviews.

The following table lists the typical survey methods. Enterprises should conduct surveys in an easy-to-difficult manner. Some service providers may provide massive survey forms for enterprises, which is inefficient and error-prone. Use information systems such as CMDB if possible. For security purposes, enterprises can provide read-only accounts for investigators to obtain information from CMDB, or they can export the required data directly.

Table 4-8 Scenarios of different survey methods

No.	Method	Scenario	Advantages	Disadvantages
1	CMDB on the live network	The customer has CMDB with the application calling module.	Directly and efficiently obtain the survey information, including the detailed resource list and connections between data layers and application layers.	The system information in the old-version CMDB is not updated in time.
2	Cloud management platform on the live network	The customer has CMP or a virtualization management platform.	Accurately obtain resource details.	Information such as the application architecture and association cannot be obtained.
3	Existing documents	The customer has complete device archives, including design documents and implementation solutions.	Quickly obtain the live network information.	The timeliness and integrity of the documents cannot be guaranteed.
4	Installation tools (RDA/third-party)	The customer agrees to install the tool agent.	Quickly obtain the detailed resource list.	1. Installing the agent on the customer's live network is security-sensitive. 2. Only a small amount of information can be obtained from databases and middleware, and the application invocation relationship is not available.

No.	Method	Scenario	Advantages	Disadvantages
5	Surveys and interviews	The customer has sufficient manpower and time and is willing to cooperate.	The customer business team is responsible for the resource list and application invocation relationship.	1. The survey period cannot be controlled. 2. The accuracy and completeness of survey information is uncontrollable.

The efficiency, completeness, and authenticity of obtained information vary among survey methods. Generally, CMDB is the most efficient survey method.

Table 4-9 Comparison of different survey methods

Survey Channel	Application Architecture Survey					Technical Architecture Survey					Method Evaluation		
	Service panorama	Service domain	Service system	Application system module	Association between applications	Technical architecture (Overall)	Technical architecture (by service domain)	Technical architecture (by service system)	Technical architecture (by application system module)	Technical architecture (technical component details)	Efficiency	Completeness	Authenticity
CMDB on the live network	√ Sorting required	√ Sorting required	√	√	√	√ Sorting required	√ Sorting required	√ Sorting required	√	√	High	High	Medium-high or high

Cloud management platform on the live network	-	-	-	-	-	-	-	-	-	√	High	Low	High
Existing knowledge base	√ Possibly exist	√ Possibly exist	√ Possibly exist	√ Possibly exist	√ Possibly exist	√ Possibly exist	√ Possibly exist	√ Possibly exist	√ Possibly exist	√ Possibly exist	High	Low/Medium/High All possible	Low/Medium/High All possible
Collection using the RDA tool	-	-	-	-	-	-	-	-	-	√	Medium	Low	High
Surveys and interviews	√	√	√	√	√	√	√	√	√	√	Low	High	Medium-high

4.7 Cloud Service Selection

4.7.1 Compute Service Selection

Huawei Cloud provides the Elastic Cloud Server (ECS) and Cloud Container Engine (CCE) services. The following table lists the ECS types provided by Huawei Cloud to meet the requirements of diversified computing scenarios.

For details about the preceding ECS types, see **ECS Types**.

Table 4-10 ECS type

Architecture	ECS Type	Instance Family	Description	Scenario
x86	General computing-plus	c	A balance of compute, storage, and network performance, dedicated CPU, and stable performance	Applicable to most application scenarios
		ac	Compared with C series, smaller network bandwidth is allocated to different CPUs and the same specifications, ensuring stable performance and lower costs.	
	High-performance computing	h	Compared with the C series, higher CPU frequency delivers 20% higher compute performance.	HPC/ Gaming/ Scientific computing
	Memory-optimized	m	Compared with the C series, the memory-optimized ECSs with a CPU/ memory ratio of 1:8 provide higher memory performance.	Memory-intensive and database/ memory database
		am	Compared with the AC series, the memory-optimized ECSs with a CPU/ memory ratio of 1:8 provide higher memory performance.	
	Large-memory	e	Compared with the C series, the memory-optimized ECSs with a CPU/ memory ratio of 1:20 provide higher memory performance.	
	Disk-intensive	d	Compared with the C series, the large-capacity and low-cost SATA local disks are provided.	Big data/ Cached database
	Ultra-high I/O	i	Compared with the C series, large-capacity NVMe local disks with high IOPS and low latency are provided.	

Architecture	ECS Type	Instance Family	Description	Scenario
		ir	Compared with the C series, small-capacity NVMe local disks with high IOPS and low latency are provided.	
	General computing	s	Compared with the C series, General computing ECSs use a CPU-unbound scheduling scheme. When the host load is light, the same computing performance as the C series can be provided. The cost is lower. However, the stability of computing performance cannot be guaranteed. It is suitable for scenarios that can tolerate performance jitter.	General web/ Development environment/Small database
	General computing-basic	t	Burstable performance instances with low costs. The burstable duration is determined by CPU credits.	Personal use/ Maintenance terminal
	GPU-accelerated	g	T4 GPU for image acceleration	3D animation rendering, CAD, and more
		p	V100 GPU for computing acceleration	AI deep learning and scientific computing
		pi	T4 GPU for inference acceleration	Real-time inference + light-load training
	AI-accelerated	ai	Ascend 310 for computing or inference acceleration	Deep learning, scientific computing, and CAE

Architecture	ECS Type	Instance Family	Description	Scenario
ARM	Kunpeng general computing-plus	kc	Compared with the C series, the Kunpeng processor is used, and the price is lower.	Adapted to most Arm application scenarios
	Kunpeng memory-optimized ECSs	km	Compared with the M series, the Kunpeng processor is used, which is more cost-effective.	Database /Memory database
	Kunpeng Ultra-high I/O ECSs	ki	Compared with the i series, the Kunpeng processor is used, and the price is lower.	Big data/ Cached database
	Kunpeng AI Inference-accelerated ECSs	kai	Compared with the AI series, the Kunpeng processor is used, and the price is lower.	Deep learning, scientific computing, and CAE

The following describes the ECS selection principles:

- **Service applicability**: Meet service requirements is the first principle for selection. In addition to CPU and memory, pay attention to bandwidth requirements. Generally, the larger the instance specifications of the same series, the higher the bandwidth allowed.
- **Cost-effectiveness**: A cost-effective solution must be selected if service requirements can be met. For example, with the same specifications, the price of the S/AC series is lower than that of the C series. If there is no high performance requirement for O&M terminals, the T series is more cost-effective. For fluctuates services, you are advised to use multi-node cluster load sharing and AS. In this scenario, you are not advised to use high-specification instance nodes. Otherwise, performance will be wasted when the number of nodes is reduced to the minimum.
- **Reliability**: Consider how to reduce the failure rate and prevent single points of failure when selecting resources. Therefore, you are advised to select the new series (with a larger number in the specifications) and deploy the resources in a balanced manner across two AZs. Resource selection optimization and cost reduction cannot compromise service reliability. The failure of a single node in a cluster network should not cause overload of other nodes.
- **Consistency**: To ensure fast scale-out, fast recovery, and auto scaling based on images, hosts that carry the same type of services must have the same

specifications. Do not use too many instance types or specifications in the same service system unless otherwise specified.

- **Resource satisfaction**: Considering service development and scale-out requirements, you are advised to select mainstream models in mainstream AZs (for example, AZ 1 and AZ 7 in CN North-Beijing4 and AZ 1 and AZ 4 in CN East-Shanghai1) and avoid selecting old flavors.

BMSs are required in special scenarios such as AI. Generally, ECSs are used for general-purpose computing. The recommended models for typical scenarios are as follows:

Table 4-11 ECS selection for typical scenarios

Location	Typical Application		Selection Suggestion
Access layer	Load balancing/ Application proxy	Nginx	C/M series
	O&M terminal	Jump server	T series
Application layer	Common application	Web services	AC/AM series
	High-performance computing service	Transcoding	C/M series
Middleware	Self-built middleware	Self-built Redis/ RocketMQ	C/M series
Data layer	Self-built databases	Self-built MySQL/Oracle	C/M series

4.7.2 Storage Service Selection

Huawei Cloud provides Object Storage Service (OBS), Elastic Volume Service (EVS), and Scalable File Service (SFS). The following table compares the three types of storage services.

Table 4-12 Comparison between the three types of storage services

Dimension	EVS	SFS	OBS
Description	High reliability, high IOPS, and elastic scalability (equivalent to disks)	High bandwidth, on-demand expansion, and shared access (equivalent to NAS)	High reliability, low cost, massive scalability, and support for objects of any type and size

Dimension	EVS	SFS	OBS
Scenario	HPC, enterprise core cluster applications, enterprise application systems, and development and testing	High-performance computing (HPC), media processing, file sharing, content management, and web services	Big data analytics, static website hosting, online video on demand (VoD), gene sequencing, and intelligent video surveillance
Storage logic	Stores binary data and cannot directly store files. To store files, you need to format the disk with a file system first.	SFS stores files, and sorts and displays in the hierarchy of files and folders.	Stores objects. Files are saved directly to OBS. The files automatically generate corresponding system metadata. You can also customize the metadata if needed.
Access method	It can only be used and accessed from applications after being attached to ECSs or BMSs and initialized (OS layer, application reconstruction not involved).	It can be mounted to ECSs, BMSs, or CCEs using network protocols. NFS and CIFS are supported (CIFS is not supported by common file systems). A network address must be specified or mapped to a local directory for access. (OS layer, application reconstruction is not involved).	OBS buckets can be accessed through the Internet or Direct Connect. The bucket address must be specified for access using HTTP or HTTPS. (Application layer, the application needs to integrate the SDK or invoke APIs, which involves application reconstruction.)
Data sharing	Supported. It is controlled by the cluster management software installed on the ECS/BMS and cannot be shared across AZs.	Supported. You can directly access SFS using NFSv3 (SFS Turbo also supports CIFS). Cross-AZ sharing is supported.	Supported. Direct HTTP/HTTPS access is required, and unlimited sharing is supported.
Remote access	Not supported	Supported	Supported
Independent use	Not supported	Supported	Supported

Dimension	EVS	SFS	OBS
Capacity	TiB-level	PiB-level (SFS Trubo)/ EiB-level (general purpose file system)	EiB-level
Latency	Minimum	Medium	Maximum
Bandwidth (throughput)	MiB/s-level	GiB/s-level	TiB/s-level
Data redundancy	Single AZ	Single AZ (SFS Turbo)/Single or multiple AZs (general purpose file system)	Single AZ/Multiple AZs
Data reliability	9 nines	10 nines	11 nines in a single AZ or 12 nines in multiple AZs
Storage billing mode	Pay-for-capacity	SFS Turbo billed by capacity/SFS general file system billed by usage	Pay-as-you-go

The following describes the selection principles of storage services.

Service Applicability Principles

You need to select a proper storage type based on service scenarios. The following aspects must be considered:

1. **Available access mode**: After EVS disks or SFS file systems are attached to a host, they are displayed as file system paths in the OS and can be directly accessed by upper-layer applications. OBS needs to be accessed by service applications using dedicated SDKs or APIs. You need to know the acceptable access modes of services. For database applications that require direct raw disk mapping, only block storage (EVS) can be used.
2. **Shared or not**: EVS supports sharing. You need to select the sharing feature during purchase and use the dedicated cluster software to manage shared disks. SFS and OBS support sharing. Therefore, you need to analyze whether the content to be stored needs to be shared by multiple nodes based on service scenarios.
3. **Storage capacity**: Different storage types support different capacities. You need to estimate the required capacity level based on the current service volume and future development to select a proper storage type.

Table 4-13 Minimum and maximum capacity of the storage service

Storage Type		Minimum Capacity	Maximum Capacity
EVS		10 GB	32 TB
SFS Turbo	20 MB/s/TiB	3.6 TB	1 PB
	Others	1.2 TB	1 PB
SFS general-purpose capacity-oriented		0	Unlimited
OBS		0	Unlimited

Performance Matching

The performance metrics of storage services include transmission bandwidth, IOPS, and latency, as shown in the following table. You need to select the appropriate storage service and specifications based on the performance requirements and specifications of the service system.

In addition, EVS and OBS have no restrictions on the size of stored objects. SFS general-purpose Capacity-Oriented is not suitable for applications with massive small files smaller than 1 MB. SFS Turbo and subsequent SFS General Performance-Oriented can support massive small file applications.

Table 4-14 Performance metrics of storage services

Storage Type		Bandwidth Upper Limit (GB/s)	IOPS Upper Limit	Average Latency
EVS	High I/O-SAS	0.15	5K	1–3 ms
	General-purpose SSD-GPSSD	0.25	20K	1 ms
	Ultra-high I/O-SSD	0.35	50K	1 ms
	General Purpose SSD V2-GPSSD2	1	128K	1 ms
	Extreme SSD-ESSD	1	128K	Sub-milliseconds
	Extreme SSD V2-ESSD2	4	256K	Sub-milliseconds
SFS Turbo	20 MB/s/TiB	20	250K	2–5 ms
	40 MB/s/TiB	20	250K	2–5 ms
	125MB/s/TiB	100	Millions	1–3 ms

	250 MB/s/TiB	100	Millions	1–3 ms
	500 MB/s/TiB	200	Millions	1–3 ms
	1,000 MB/s/TiB	200	Millions	1–3 ms
SFS General-Purpose	Capacity	50	100K	7 ms
	Performance	200	2000K	5 ms
OBS		TB-level	Tens of millions	10 ms+

Cost Optimization

You need to consider costs when selecting a storage type to reduce storage costs while meeting service performance requirements.

1. Select the storage service with the lowest unit price while meeting service performance requirements.
2. For storage (EVS and SFS Turbo) billed by specifications, predict the service increment and monitor the capacity. You are advised to reserve 15% to 20% as the scale-out threshold to prevent resource waste caused by excessive capacity specifications.
3. For pay-per-use storage (general-purpose SFS and OBS), plan the usage and purchase resource packages to reduce costs.
4. You can plan lifecycle management policies for storage (general-purpose SFS and OBS) and move cold data to infrequently accessed storage in a timely manner to reduce costs.
5. For services that require large storage capacity and long data retention period, you can reconstruct the service application layer and combine different types of storage (for example, combine EVS/SFS Turbo and OBS) to optimize costs while ensuring service performance.

Reliability Assurance

EVS, SFS Turbo, general-purpose SFS, and OBS use three-replica storage. Data durability meets service requirements, but reliability varies.

- The three replicas of EVS and SFS Turbo are in the same AZ. If the AZ egress or equipment room is faulty, services will be unavailable.
- General-purpose SFS and OBS support both single-AZ and multi-AZ deployment. (Currently, general-purpose SFS supports only single-AZ deployment. Multi-AZ deployment will be available in the future.) For services that require high continuity, you can select multi-AZ instances.
- EVS supports quick data backup and restoration using images, snapshots, and cloud backup. SFS Turbo supports backup and restoration using cloud backup. General-purpose SFS and OBS are generally used in ultra-large-capacity service scenarios and the backup capability is not planned.

Based on the preceding selection principles, the following are some typical scenario suggestions:

- Shared disks are not recommended unless self-built databases are deployed in two-node cluster or cluster mode. Instead, SFS is used to implement file sharing among multiple hosts. (Shared disks cannot be attached to multiple ECSs across AZs, but SFS supports this function.)
- For applications that require frequent read and write of a large number of logs and log summary analysis, SFS is recommended as the unified log storage for multiple nodes (select the type based on performance requirements).
- For asynchronous interaction and latency-insensitive services, OBS is preferred to reduce costs. If services are difficult to adapt to reconstruction, SFS general-purpose capacity-oriented storage can be used.
- In AI scenarios, it is recommended that SFS Turbo and OBS be used together to reduce costs and improve performance.

4.7.3 Network Service Selection

Huawei Cloud provides the following network services: VPC, Enterprise Router (ER), Enterprise Switch (ESW), Direct Connect, Virtual Private Network (VPN), Global Accelerator (GA), Elastic Load Balance (ELB), NAT Gateway, and Elastic IP (EIP). The following are the network service selection suggestions:

- VPC peering connections are used for communication between a small number of VPCs in the same region, Cloud Connect is used for communication between VPCs in different regions, and Direct Connect or VPN is used for communication between on-premises and cloud. Enterprise Router is used to simplify the interconnection and route management between VPCs and between on-premises and cloud.
- The cloud and on-premises subnets overlap and IP addresses are separated. Layer 2 interconnection is required to connect to enterprise switches.
- The cloud and on-premises subnets overlap or routes between the two subnets cannot be directly enabled due to management reasons, but services need to communicate with each other. In this case, you need to use the private NAT gateway.
- You need to build an HA system on the cloud. It is recommended that the two ECSs be deployed in the same subnet and across AZs, and be bound with a virtual IP address and keep-alive mechanism.
- If you need to improve the cross-border resource experience of users in a specified region, you can use Cloud Connect and Global Accelerator to reduce the latency of traffic through the Huawei Cloud backbone network.
- In low-concurrency and heavy-traffic basic Layer 4/Layer 7 load distribution scenarios, you are advised to select a shared load balancer and enable the performance assurance mode (supporting 50,000 concurrent requests). Purchase two instances and use domain name resolution to support more concurrent requests.
- Dedicated load balancers are recommended when the number of concurrent users exceeds 100,000 and full-link HTTPS or advanced forwarding policies are required.

- When an ECS needs to access the Internet, you are not advised to bind an EIP to the ECS. Instead, you are advised to configure a public network NAT gateway to use SNAT for flexible management.
- If services need to be provided for the public network, you are advised to bind the public IP address to the ELB or NAT gateway instead of the ECS for flexible expansion and control adjustment.
- Unless otherwise specified, you are advised to use the default dynamic BGP for the EIP link type.

4.8 Anti-patterns in Survey and Evaluation

During cloud migration survey, there might be some anti-patterns. If not identified and avoided, these can affect the survey's effectiveness and accuracy. This can lead to poor decisions and cloud migration plans. The following are some typical anti-patterns in cloud survey and evaluation:

- **Incorrect survey method**

  Sending complex survey forms at the start can dampen the provider's willingness to cooperate, lowering survey efficiency and the quality of results.

  Suggestion: Use a scientific research method that starts simple and gets more detailed over time, with ongoing improvements. For example, collect data from CMDB, identify gaps using existing documents, and streamline the research process for efficiency.

- **Incomprehensive business survey**

  Only the technical survey and evaluation are concerned, while the business requirements and user scenarios are ignored.

  Suggestion: Conduct in-depth business survey and analyze user scenarios to ensure that cloud migration can meet business requirements and improve user experience.

- **Inadequate evaluation**

  Only one single metric (such as cost or performance) is considered, while other important factors are ignored.

  Suggestion: Establish a comprehensive evaluation system, covering various aspects, including cost, performance, security, scalability, and maintainability.

- **Underestimated migration complexity**

  Cloud migration was wrongly considered as just a simple technology migration, overlooking application structure, data dependencies, and their effect on business. This led to many issues after the migration.

  Suggestion: Thoroughly analyze internal and external associations, identify strong and weak connections, and evaluate risks and impacts. The results serve as the foundation for subsequent batch planning and solution switching, minimizing potential problems and impacts.

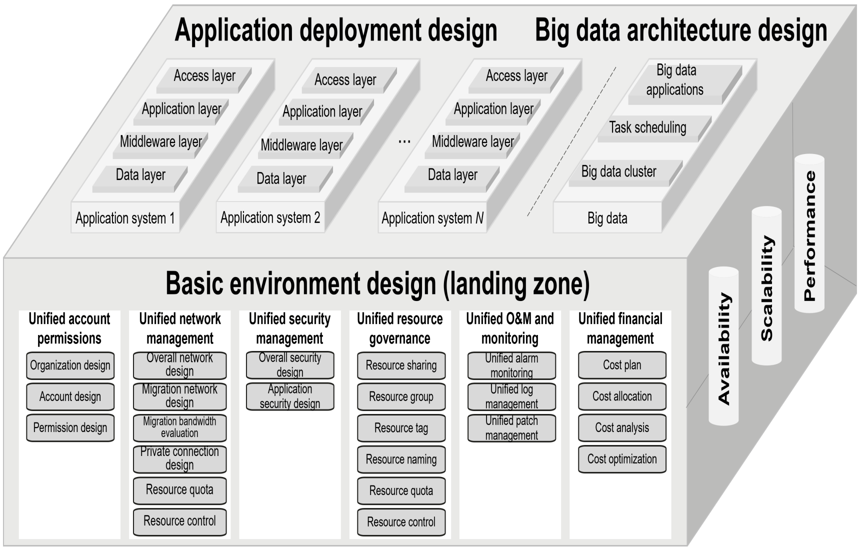
By identifying and avoiding these anti-patterns, cloud migration survey and evaluation can be performed more efficiently and accurately, providing powerful support for enterprises in making informed decisions and designing solutions.

5 Solution Design

5.1 Overview

The cloud architecture consists of the infrastructure, applications, and big data.

Figure 5-1 Cloud architecture



1. **Basic environment design**: Before migrating workloads to the cloud, you need to prepare the basic environment. The basic environment, often referred to as the landing zone, encompasses a comprehensive design framework that addresses six aspects: account and permissions, network, security strategy, resource governance, O&M monitoring, and financial management.
2. **Application deployment design**: Application deployment indicates the technical architecture of cloud applications. You need to design the access, application, middleware , and data layers as well as cloud service technologies for each layer. It is essential to consider six key elements: availability, performance, scalability, security, cost, and maintainability. Particular emphasis should be placed on availability, scalability, and performance, while

ensuring that security, cost, and maintainability align with basic requirements of the environment.

3. **Big data architecture design**: The design of big data architecture encompasses several key components: big data clusters, big data task scheduling platforms, and big data applications. You can refer to the application deployment design.

To design a more efficient, flexible, and scalable cloud architecture, you can refer to TOGAF ADM (a widely-used framework for developing architectures) across four domains, best practices, and tools.

5.2 Establishing a Solution Design Team

An efficient and professional solution design team is the key to ensuring the success of a cloud migration project. The team is tasked with developing a feasible cloud migration solution that encompasses the design of technology architectures, optimization of service frameworks, thorough cost-benefit analysis, and stringent adherence to security protocols. Cloud Center of Excellence (CCoE) can help enterprises establish such a team. Necessary team members are as follows:

1. **Cloud architect**: Experts from the IT architecture department or experts with profound cloud technology background are responsible for designing the cloud technology architecture, including selecting appropriate cloud services (IaaS, PaaS, and SaaS), designing the target cloud architecture based on the four architectures and six elements, ensuring technology rationality, optimizing resource configurations, and providing consultation for each technical decision.
2. **Data architect**: Experts from the big data team are assigned by the IT department director. They are responsible for designing the cloud data architecture, including data storage, processing, integration, and management.
3. **Application architect**: An application team designated by the business department director is responsible for designing and managing the cloud application architecture of the business system, for example, application architecture mode, technology selection, and deployment mode, to ensure the performance, scalability, security, and reliability of applications.
4. **Business expert**: Business experts are assigned by the business department director. They need to have an in-depth understanding of the current business process, ensure that the solution can meet actual business requirements, promote realization of business value, and improve operations efficiency and user experience.
5. **Financial expert**: Financial experts are assigned by the financial department director. They need to comprehensively evaluate the cost structure of the cloud migration solution, for example, the initial investment, operations cost, potential savings, and long-term benefit forecast. Quantitative analysis can provide key financial metrics such as cost-effectiveness ratio and ROI for decision-making.
6. **Cloud security expert**: Experts from the information security department or professionals with security compliance certification are responsible for evaluating the security compliance of cloud service providers, designing data protection policies, ensuring that cloud migration solutions comply with

industry security standards and legal requirements, and effectively preventing risks.

7. **Project manager**: A project manager comes from the project management office (PMO) or has rich project management experience. The manager is responsible for planning, executing, monitoring, and closing the entire cloud migration solution design project. The manager needs to ensure that the project is completed on time, with high quality, and within budget, coordinate resources across departments, and promote team collaboration and communication.
8. **Cloud service provider consultant**: The consultant is selected from a cloud service provider or professional service company. The consultant provides cloud migration solution suggestions based on best practices and assists enterprises in customizing cloud migration solutions, including technical implementation details and best practice sharing.

The solution design team can be set up in two modes. If an enterprise independently designs the cloud migration solution, the above roles are mainly played by internal personnel of the enterprise, and the cloud service provider provides consulting and technical support. If the enterprise purchases third-party professional services for cloud migration solution design, the third-party team takes the lead. The enterprise provides necessary business requirements and technical information. The enterprise team needs to closely cooperate with the third-party team to ensure that the solution design meets the enterprise's business requirements. The two parties work closely together to promote the solution design.

5.3 Designing the Basic Environment

The basic environment of an enterprise on the cloud is the landing zone. Before migrating any service system to the cloud, enterprises need to plan and design a cloud environment that is excellent, stable, reliable, scalable, and secure.

For details, see **Landing Zone Design**.

Enterprises need to design a comprehensive security protection solution for the cloud environment. For details, see **Security Architecture Design**.

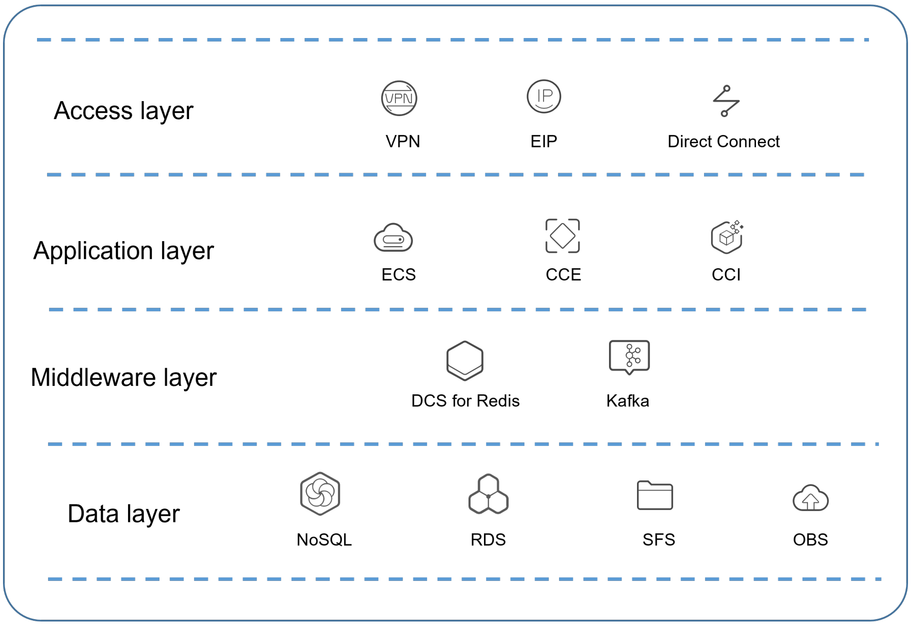
5.4 Designing the Application Architecture

5.4.1 Application Deployment Overview

The methodology of application deployment design comes from the practical experience of Huawei Cloud architects in various fields. The methodology can help enterprises migrate applications to the cloud seamlessly and use the cloud effectively.

Applications can be divided into four layers based on the functions of each component: access layer, application layer, middleware layer, and data layer.

Figure 5-2 Four-layer deployment of applications



- **Access layer**: Cloud services are deployed in a VPC, which is isolated from external networks. To access the cloud services from external networks, the following methods can be used:
  - **Direct Connect**: Direct Connect is used to establish a dedicated network connection between your or other cloud vendors' on-premises data center and a VPC on the cloud. This connection features high speed, stability, security, and low latency. You can access resources such as ECSs and load balancers on the cloud through intranet addresses, and implement service communication and data transmission between the cloud and on-premises data centers.
  - **EIP**: An elastic IP address (EIP) provides independent public IP addresses and bandwidth for Internet access. EIPs can be bound to or unbound from ECSs, BMSs, virtual IP addresses, load balancers, and NAT gateways, to access to or be accessed from the public network.
  - **VPN**: Virtual Private Network (VPN) establishes a convenient, flexible, and out-of-the-box IPsec encrypted communication channel between the local data center and Huawei Cloud VPC. This helps to build a flexible and scalable hybrid cloud computing environment.
- **Application layer**: The application layer controls workflows and implements service logic. The application layer processes requests from the access layer and returns request results. It also connects to the middleware layer or data layer to add, delete, modify, and query data. There are the following types of resources on cloud applications:
  - **VM**: On the cloud, a VM is also called an Elastic Cloud Server (ECS), which is a basic computing component consisting of CPUs, memory, OS, and Elastic Volume Service (EVS) disks. After creating an ECS, you can use it like you use your local PC or physical server.
  - **Container**: Container virtualization technology has become a widely recognized container technology for sharing server resources. Container technology provides system administrators with great flexibility in the process of creating container OS instances on demand.

- **Middleware layer**: Resources are shared between application software at this layer, which also manages compute resources and network communication. Middleware is mainly used for data transmission, data access, application scheduling, and process management in a distributed environment. On the cloud, the following service middleware is commonly used:
  - **Cache**: Huawei Cloud provides Distributed Cache Service (DCS), including DCS Redis and DCS Memcached.
  - **Message middleware**: Huawei Cloud provides distributed message middleware, including Kafka, RabbitMQ, and RocketMQ.
- **Data layer**: System service data is kept in persistent storage to implement upper-layer service logic. The data layer generally consists of databases and file systems.

Application deployment design is intended to ensure the performance, availability, and security of enterprise applications, while also considering scalability, cost, and maintainability. Therefore, the following six factors must be considered during application deployment design: availability, scalability, performance, security, cost, and maintainability. The reasons are as follows:

- **Availability**: Availability design aims to ensure the availability and reliability of cloud application systems and ensure that the system can run stably in the event of exceptions.
- **Scalability**: Scalability design ensures that applications can maintain availability and performance under different loads and can be scaled based on loads to meet requirements without causing system breakdown or performance deterioration.
- **Performance**: Performance design is to ensure that the deployment of cloud applications meets users' performance requirements, including response time, throughput, and number of concurrent requests.
- **Security**: Security design protects cloud applications and data against malicious attacks and data leakage.
- **Cost**: Cost design aims to reduce the deployment and O&M costs as much as possible while ensuring application performance, availability, and security.
- **Maintainability**: Maintainability design aims to improve the system maintainability (including automatic deployment, monitoring and alarm reporting, log analysis, capacity planning, and troubleshooting), display system status, and ensure quick restoration.

Security, cost, and maintainability are global design elements and have been designed in the basic environment. Therefore, you need to focus on the availability, scalability, and performance when designing application deployment.

5.4.2 Availability Design

5.4.2.1 Definition

Availability is the probability that a product or service is reliable and maintainable under defined conditions at a given time. Service availability is generally measured by Service-Level Agreement (SLA). Each type of cloud service provides services based on their SLA commitment. The following table lists the downtime acceptable for different SLA commitments.

Table 5-1 SLA levels

SLA	Weekly Downtime	Monthly Downtime	Annual Downtime
99%	1.68 hours	7.2 hours	3.65 days
99.90%	10.1 minutes	43.2 minutes	8.76 hours
99.95%	5 minutes	21.6 minutes	4.38 hours
99.99%	1.01 minutes	4.32 minutes	52.56 minutes

5.4.2.2 Availability Zone

An availability zone (AZ) is an independent fault domain in a public cloud environment with isolated physical data centers. Each AZ has independent power supply, network connections, and hardware facilities. Public cloud vendors usually deploy different AZs in different geographical locations to improve system availability and fault tolerance. AZs have the following advantages:

- High availability: Applications and data are deployed in multiple AZs. If one AZ is faulty, other AZs can still provide services.
- Fault tolerance: If an AZ is faulty, applications and data can be quickly taken over by normal AZs.
- Geo-redundancy: AZs are deployed in diverse geographical locations to safeguard against catastrophic events, such as natural disasters or power outages.

The following aspects can be considered for designing HA applications in multiple AZs:

- Cross-AZ deployment: Various components of an application are distributed across AZs to maintain functionality even if one AZ becomes unavailable. To streamline the management of these multi-AZ deployments, enterprises can leverage tools offered by cloud service providers or utilize container orchestration platforms.
- Load balancing: Load balancers can distribute traffic to application instances in different AZs. You can set proper forwarding policies to prevent a single AZ from being overwhelmed. In this way, even if one AZ is affected by heavy load or faults, other AZs can handle the traffic without disruption.
- Data redundancy and backup: Redundancy and backup policies in different AZs can ensure data availability and recoverability. If data in an AZ is lost or unavailable, the data can be restored from redundant data in other AZs.
- Automated fault recovery: The automated failover mechanism instantly redirects applications to available AZs during outages, ensuring swift service recovery. Enterprises can utilize container orchestration tools, automation scripts, or cloud providers' failover features to enable seamless fault recovery.
- Monitoring and alarms: The system can continuously monitor the health of applications and infrastructure across all AZs. In the event of any anomalies or faults, immediate alarms are generated, notifying relevant personnel to

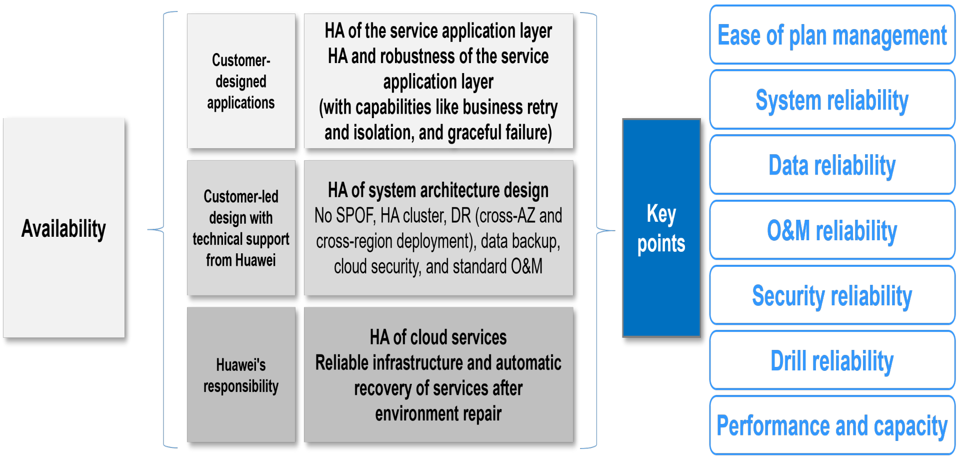
swiftly address and resolve faults, thereby minimizing downtime and ensuring seamless service continuity.

Deploying HA applications across AZs significantly enhances system availability and fault tolerance, reduces the risk of service disruptions and data loss, and ensures seamless service continuity.

5.4.2.3 HA Solution on the Cloud

The availability of workloads on the public cloud is determined by the availability of the application layer, architecture design, and cloud services. The availability of workloads on the public cloud is a systematic project. The reliability of workloads on the public cloud depends on the availability design of the overall service architecture and O&M specifications (such as the backup mechanism, routine drills, and personnel operation specifications).

Figure 5-3 Service availability solution


Availability
Customer-designed applications
HA of the service application layer
HA and robustness of the service application layer
(with capabilities like business retry and isolation, and graceful failure)
Customer-led design with technical support from Huawei
HA of system architecture design
No SPOF, HA cluster, DR (cross-AZ and cross-region deployment), data backup, cloud security, and standard O&M
Huawei's responsibility
HA of cloud services
Reliable infrastructure and automatic recovery of services after environment repair
Key points
Ease of plan management
System reliability
Data reliability
O&M reliability
Security reliability
Drill reliability
Performance and capacity

Most Huawei Cloud services can be deployed in HA mode. HA capabilities are implemented for data centers, hardware, data, and self-service functions. Huawei Cloud data centers are deployed around the globe to meet resource requirements for different regions. Each region is divided into multiple AZs. Each AZ has independent cooling, fire extinguishing, moisture-control, and electrical facilities, and the failure of one AZ does not affect others. There are four types of HA deployments:

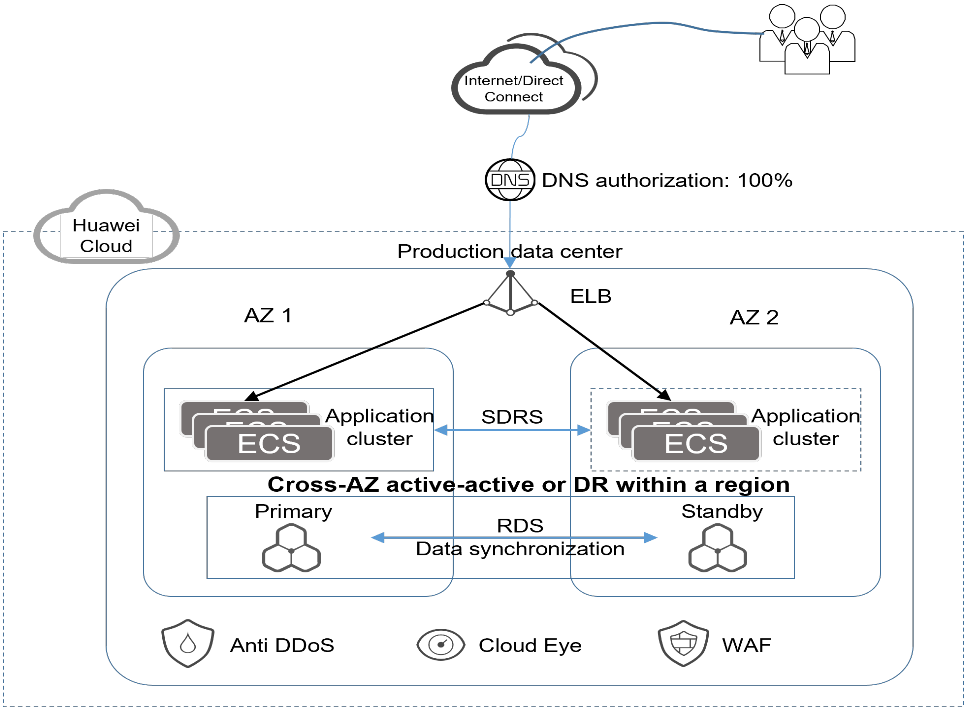
- **Single-AZ deployment**: Generally, single-AZ deployment is not recommended on the cloud, except for latency-sensitive services. You can deploy cloud services in primary/standby or cluster mode to quickly recover services from a fault. Automatic detecting and switching over faulty nodes in a cluster can eliminate single points of failure (SPOFs).
- **Dual-AZ (intra-city) HA**: For services that require high availability, you can deploy services in multiple equipment rooms in the same city. This way, service continuity is guaranteed even if the network, physical device, or power supply of one equipment room fails. Huawei Cloud users can deploy services across AZs. AZs are isolated from each other, so if one AZ fails, you can switch services to another AZ to quickly recover services. Most cloud service products have capabilities related to dual-AZ HA. You only need to select desired capabilities when purchasing the cloud service.

- **Geo-redundant DR**: For some ultra-large or commercial systems that demand extra robust protection, the dual-AZ (intra-city) HA solution cannot solve regional-level faults, such as earthquakes and floods. So, you need to add a remote DR equipment room on the basis of intra-city DR. That way, you are protected in the event of a regional disaster.
- **Cross-cloud HA**: To meet requirements for multi-cloud HA, enterprises can deploy a primary site on Huawei Cloud and deploy a standby site on other cloud vendor's platform.

5.4.2.4 Dual-AZ HA Design

The dual-AZ HA solution is the most commonly used solution on the public cloud. The four-layer architecture (access, application, middleware, and data layers) of applications are recommended for end-to-end dual-AZ deployment.

Figure 5-4 Dual-AZ HA design



Key design points:

- Service module: For services supporting cluster deployment, resources are deployed in two AZs, where loads are balanced by ELB. For single-node ECSs, SDRS is used for AZ-level DR.
- Cloud service HA: Primary and standby nodes are deployed in different AZs.
- Database synchronization: Primary/Standby RDS instances are deployed across AZs, where data can be synchronized.
- DR switchover: If an AZ fails, RDS database services automatically switch to the standby databases. Application services can be taken over by the DR servers automatically or in just a few clicks.
- DR drill: You can switch over applications or use the DR drill function of SDRS.

NOTE

For applications that are highly sensitive to latency, such as during the cloud migration of Telecom's NFV NEs, it is essential to conduct thorough testing, validation, and optimization. This ensures optimal performance and a seamless user experience of dual-AZ deployment on the public cloud.

- **Selecting appropriate AZs**

  Cloud service providers provide the physical locations and network latency information of each AZ. To implement HA across AZs, select AZs that are close to each other to reduce the network latency and improve the application response speed.

- **Verifying latency**

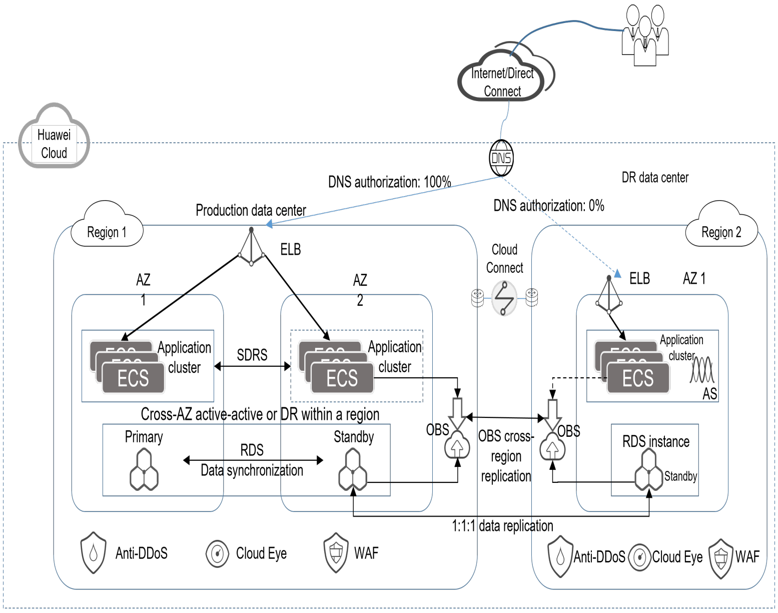
  Before implementing the dual-AZ HA solution, you need to fully test and verify the latency of applications across AZs. By conducting a simulated stress test, you can measure the latency of application calls across AZs to assess whether the service meets performance requirements and make informed decisions based on the results.

5.4.2.5 Geo-Redundant DR Design

Geo-redundant DR can ensure service continuity.

- This solution ensures unparalleled service continuity and maximum data availability, safeguarding both data and critical services even during regional natural disasters.
- The RPO is determined by the database replication interval. Servers at the DR center are always running, so an RTO of almost zero can be achieved. The time needed to complete a DR switchover depends on how long the DNS cache takes to update. It usually takes a few minutes and can be faster if GSLB is used.

Figure 5-5 Three-AZ HA design



Key design points:

- The production and DR centers are deployed in two different regions of Huawei Cloud.

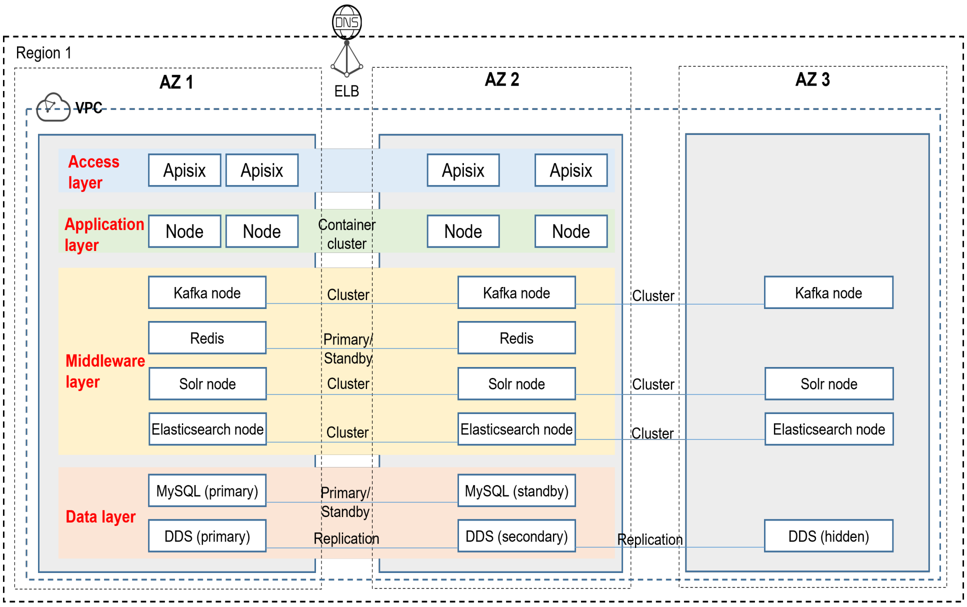
- The production centers are deployed in two AZs, and the DR center is deployed in one AZ.
- RDS instances are deployed at the production and DR centers. One primary and two standby databases are deployed.
- Configurations, logs, snapshots, and backups generated at the production and DR centers are replicated across regions using OBS.
- If an AZ at the production site is faulty, services are switched to the other AZ, and switchover is performed between the primary/standby databases.
- If both production centers fail, a database switchover is performed, and by modifying DNS configuration, all user traffic is directed to the DR center.
- After the production centers recover, a database switchover is performed, and DNS directs all user traffic to the production centers.
- To improve DR center utilization, some read-only and analysis services are distributed to the DR center.

Establishing HA DR capabilities is a complex project that includes ingress traffic control, service layer reconstruction, middleware and database control, and the collaboration of numerous systems. It requires professional skills to build HA DR systems. If customers lack related experience and want to quickly build an HA DR system, the Multi-cloud high Availability Service (MAS) is a good choice. This service is derived from the multi-cloud application HA solution of Huawei consumer services. It provides end-to-end service failover and DR drill capabilities that include everything from the traffic ingress and application layer, to the data layer. MAS ensures quick service recovery and improved service continuity.

5.4.2.6 Cross-AZ HA Design Example

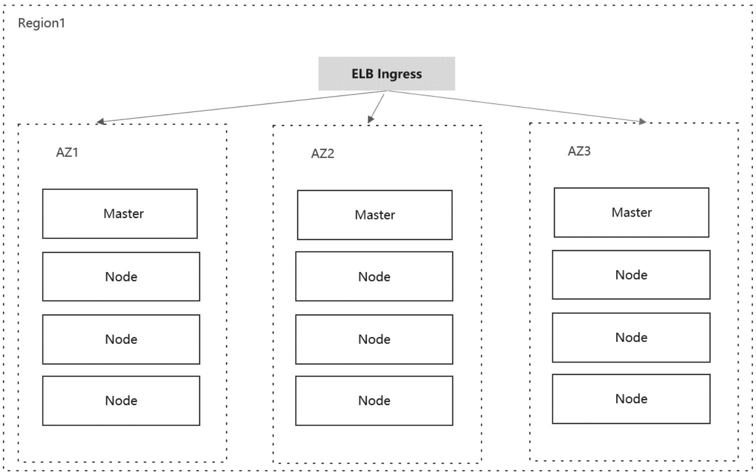
Cross-AZ HA is one of the most important benefits of migrating IDCs to the cloud. Cross-AZ HA is the best choice for enterprises after cloud migration. It is cost-effective and convenient. The following uses a large-scale retail e-commerce platform as an example to describe how to design cross-AZ HA after cloud migration.

Figure 5-6 HA design example



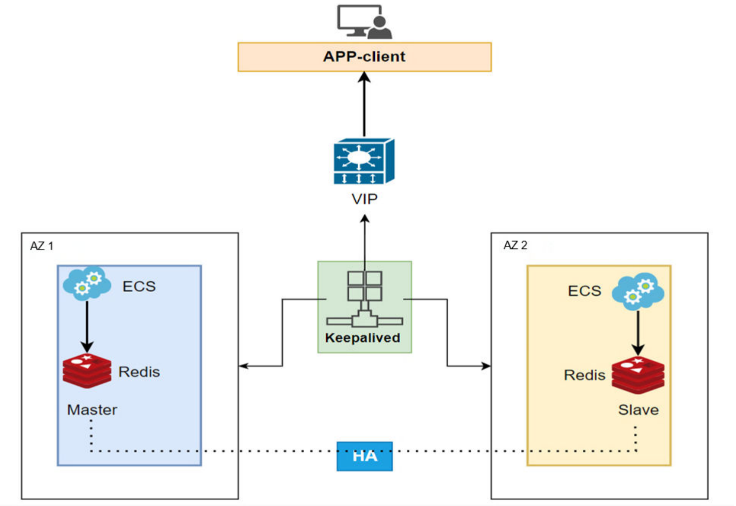
- **Access layer**: Apisix instances are evenly distributed in dual AZs. If an AZ is faulty, ELB can still forward traffic to the Apisix instances in the normal AZ.
- **Application layer**: Containers are deployed, and service nodes are distributed across AZs. Even if an AZ is abnormal, Apisix can forward traffic to normal servers.
- **Middleware layer**: Kafka, Solr, and Elasticsearch are deployed in a three-AZ cluster. If any AZ is faulty, services are still available. Redis is deployed in primary/standby mode across two AZs.
- **Data layer**: MySQL databases are deployed in primary/standby mode across two AZs to implement HA. MongoDB uses replica sets or clusters across three AZs. If an AZ is faulty, other AZs provide services properly.
- **Application layer - container cluster HA**
  - Master HA: Master nodes in the container cluster are evenly distributed in three AZs.
  - Ingress gateway HA: If the multi-AZ function is enabled for load balancers, ELB ingresses support cross-AZ HA.
  - Application HA: Kubernetes supports application HA. You can configure topologyKey to distribute pods across AZs.

Figure 5-7 Application HA design example



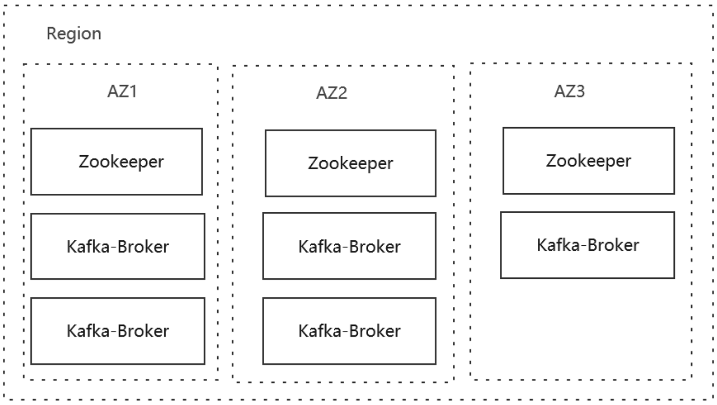
- **Middleware layer - Redis HA**
  - Data is persisted on the primary node and synchronized to the standby node in real time. The standby node also persists a copy of the data.
  - Primary/standby instances are deployed in different AZs with isolated power supplies and networks. If the equipment room where the primary node is located is faulty due to power or network faults, the standby node takes over services, and the client can connect to the standby node and read and write data.
  - Redis clusters work with Keepalived to generate VIPs, improving service availability.

Figure 5-8 Example of Redis HA design at the middleware layer



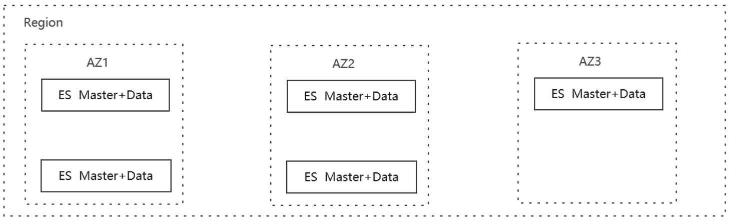
- **Middleware layer - Kafka HA**
  - ZooKeeper HA: ZooKeeper nodes are deployed in three AZs. Each node is in one AZ, or two nodes in AZ 1, two in AZ 2, and one in AZ 3. If an AZ is unavailable, the cluster still has more than half of the quorum nodes for election.
  - Kafka-Broker data node HA: Kafka-Broker nodes are deployed in three AZs. Two nodes in AZ 1, two in AZ 2, and one in AZ 3. At least three replicas are configured for a topic, and **unclean.leader.election.enable** is set to **true**. If any AZ is down, at least one replica is available in the cluster.

Figure 5-9 Example of Kafka HA design at the middleware layer

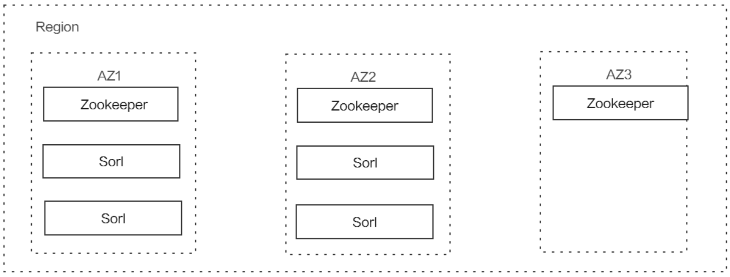


- **Middleware layer - Elasticsearch HA**
  - Master HA: Elasticsearch Master nodes are deployed in three AZs. Two nodes in AZ 1, two in AZ 2, and one in AZ 3. If an AZ is unavailable, the cluster still has more than half of the quorum nodes for election.
  - Data node: Elasticsearch Data nodes are deployed in three AZs. Two nodes in AZ 1, two in AZ 2, and one in AZ 3. At least two replicas are configured for index shards, and three replicas are configured for the primary shard. If any AZ breaks down, the cluster still has a complete replica to ensure HA.

- **Figure 5-10** Example of Elasticsearch HA design at the middleware layer

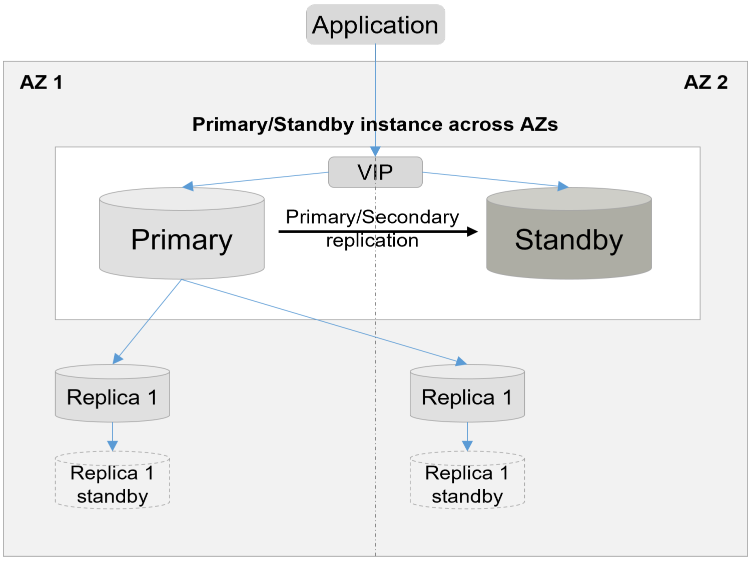


- **Middleware layer - Solr HA**
  - ZooKeeper HA: ZooKeeper nodes are deployed in three AZs. Each node is in one AZ, or two nodes in AZ 1, two in AZ 2, and one in AZ 3. If an AZ is unavailable, the cluster still has more than half of the quorum nodes for election.
  - Sorl data nodes: Sorl data nodes are evenly distributed in two AZs. At least (N/2) + 1 replicas are set for index shards. If an AZ is down, the cluster still has a complete replica to ensure HA.
  - **Figure 5-11** Example of Sorl HA design at the middleware layer



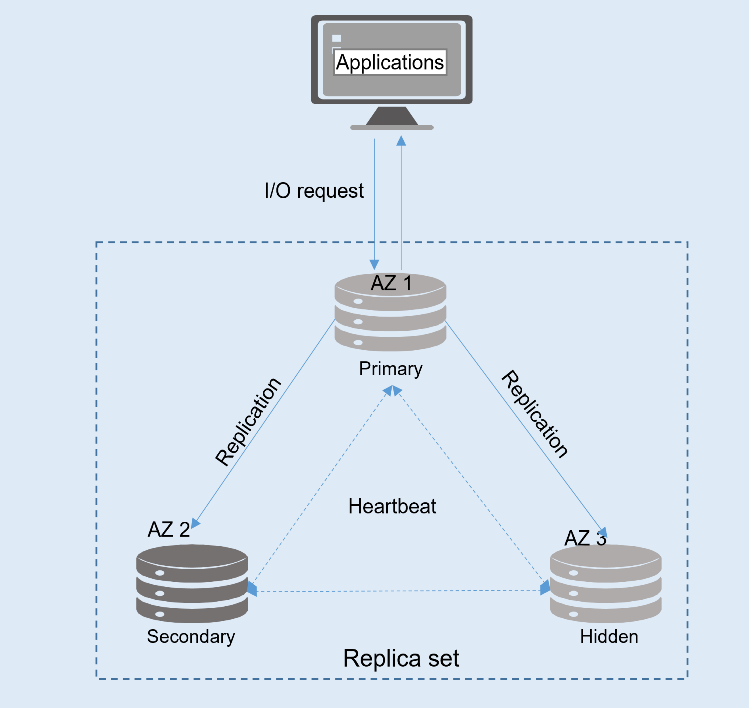
- **Data layer - MySQL HA**
- The primary and standby instances are deployed across AZs. Data is seamlessly synchronized between the primary and standby instances by leveraging the built-in replication and synchronization features of MySQL.
- The primary/standby instances provide services through a VIP, and their actual IP addresses are inaccessible to tenants.
- Primary/Standby switchover is performed within seconds. During the switchover, the VIP is switched to the new primary node, and applications are interrupted for only seconds.
- Read replicas can be attached to the instances and deployed across AZs.

Figure 5-12 Example of MySQL HA design



- **Data layer - MongoDB HA**
  - DDS replica sets can be deployed across three AZs. There are three nodes by default, and a maximum of seven nodes are supported. They are deployed in three AZs. Data is seamlessly synchronized by leveraging the built-in replication features of MongoDB.
  - The Mongo client supports multiple server addresses and availability detection.
  - If the AZ where the primary node is located is faulty, a new primary node will be elected. If the secondary node is unavailable, the hidden node takes over services to ensure HA. Currently, three, five, and seven replicas can be configured.

Figure 5-13 Example of MongoDB HA design



5.4.3 Scalability Design

5.4.3.1 Scalability on the Cloud

In contrast to traditional IDCs, cloud resources provide enhanced scalability. Based on different requirements, the scalability can be classified into the following types:

- **Vertical scaling** is applicable to a single application, independent applications, and stateful applications, particularly when rapid hardware upgrades are necessary to accommodate evolving demands. For instance, during promotional campaigns, resource demands frequently surge multiple times than usual. To address this, enterprises can swiftly upgrade their cloud resource configurations (for example, additional CPUs, increased memory, expanded bandwidth, and greater disk capacity) via the console or OpenAPI. After the promotion, they can scale down these resources back to the previous specifications to achieve cost efficiency.
- **Horizontal scaling** is applicable to distributed, stateless, and rapidly-changing applications when fixed data resources fail to handle the swift shifts in service requirements. In case of an unexpected surge in enterprise demand or promotional campaigns, you can swiftly scale resources through dynamic scaling policies. Moreover, you can fine-tune resource allocation to seamlessly accommodate business growth.

  The following scaling policies are supported:

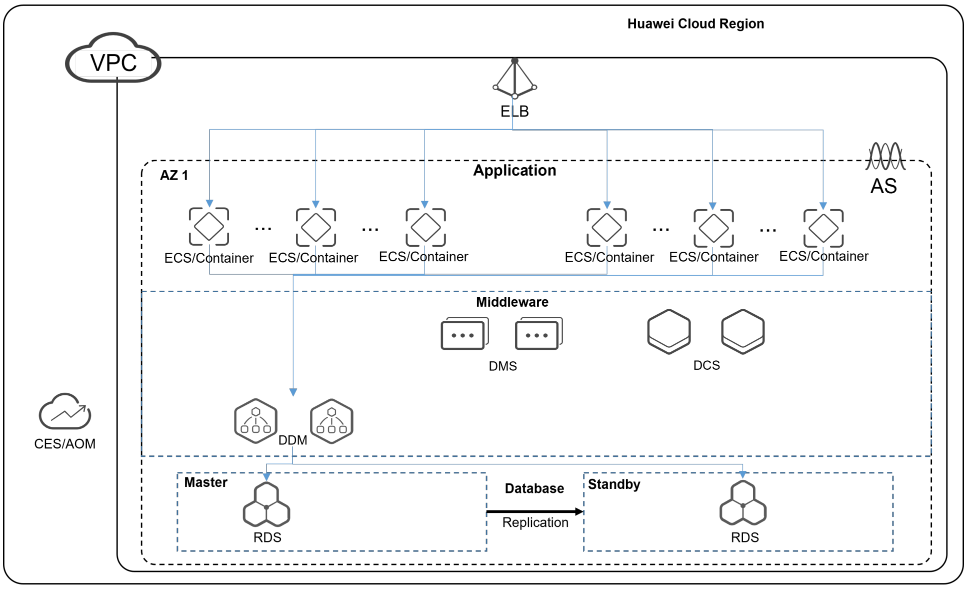
  - Scheduled mode: Create a scheduled task to scale resources at a specified time.
  - Metric mode: Create a threshold for triggering an alarm based on resource performance metrics (such as CPU usage and average network traffic). After an alarm is triggered, resources are automatically scaled.

  - Fixed quantity mode: Set the minimum and maximum expected resource quantities. If the quantity of instances is less than the minimum or greater than the maximum, the system automatically scales resources.
  - Manual mode: Manually add, remove, or delete resources.

5.4.3.2 Scalability Design

The following figure shows scalability of Huawei Cloud products at each layer.

Figure 5-14 Scalability design example



The key points for designing scalability at each layer after cloud migration are as follows:

- **Key points for scalability design at the application layer**
  - If applications are deployed in containers using Huawei Cloud CCE, they can be scaled based on scaling policies of CCE. Application pods can be automatically scaled using AOM to efficiently handle workload fluctuations based on alarm policies.
  - If the application layer is deployed on ECSs, you can use Huawei Cloud Auto Scaling to implement horizontal scaling based on scaling policies.
- **Key points for scalability design at the middleware layer**
  - Message middleware layer: The underlying layer of Huawei Cloud DMS for RabbitMQ is a cluster. As the message volume and load increase, the specifications can be smoothly expanded.
  - Cache middleware layer: Huawei Cloud DCS Redis master/standby can seamlessly increase node specifications as the hot data capacity increases.
- **Key points for scalability design at the data layer**
  - Database middleware layer: Huawei Cloud DDM is deployed in a cluster. The DDM cluster specifications can be smoothly expanded to cope with increased database processing requirements.
  - Database layer: Huawei Cloud RDS can smoothly expand read replicas to handle a large number of reads. DDM is used to horizontally scale

multiple instances. Large tables are horizontally split and evenly distributed to multiple database instances to improve the database capacity and performance. In addition, Huawei Cloud GaussDB adopts the decoupled storage-compute architecture, which supports scale-out within minutes and reduces interruption time.

5.4.4 Performance Design

Performance is key to architecture design. The previous section describes scalability design. Scalability is a prerequisite for high performance. The following factors affect the performance of applications on the cloud:

- In terms of compute resources, latency, or the wait time between operations, stands out as the most straightforward metric for evaluating cloud computing efficiency.
- For network resources, throughput is the rate at which data is processed.
- Data transmission is represented by byte/second or bit/second. The throughput limit is an important performance bottleneck.
- For storage resources, IOPS is a measurement method of data transmission, which refers to the number of input/output operations per second.
- For database resources, the concurrency capability refers to the number of programs running in a period.

The following aspects also need to be considered: solution selection, performance measurement, performance monitoring, and performance balancing.

- **Solution selection**

  Select solutions tailored to different scenarios and combine multiple methods to meet requirements.

  Methods are continuously iterative and optimized and the data-driven method is also used to optimize the selection of resource types and configuration options.
- **Performance measurement**

  Set performance measurement and monitoring metrics to capture key performance metrics.

  Use visualization technologies to display performance metrics and issues (such as abnormal status and low utilization).
- **Performance monitoring**

  Determine the monitoring scope, measurement, and threshold.

  Create a complete view from multiple dimensions.
- **Performance balancing**

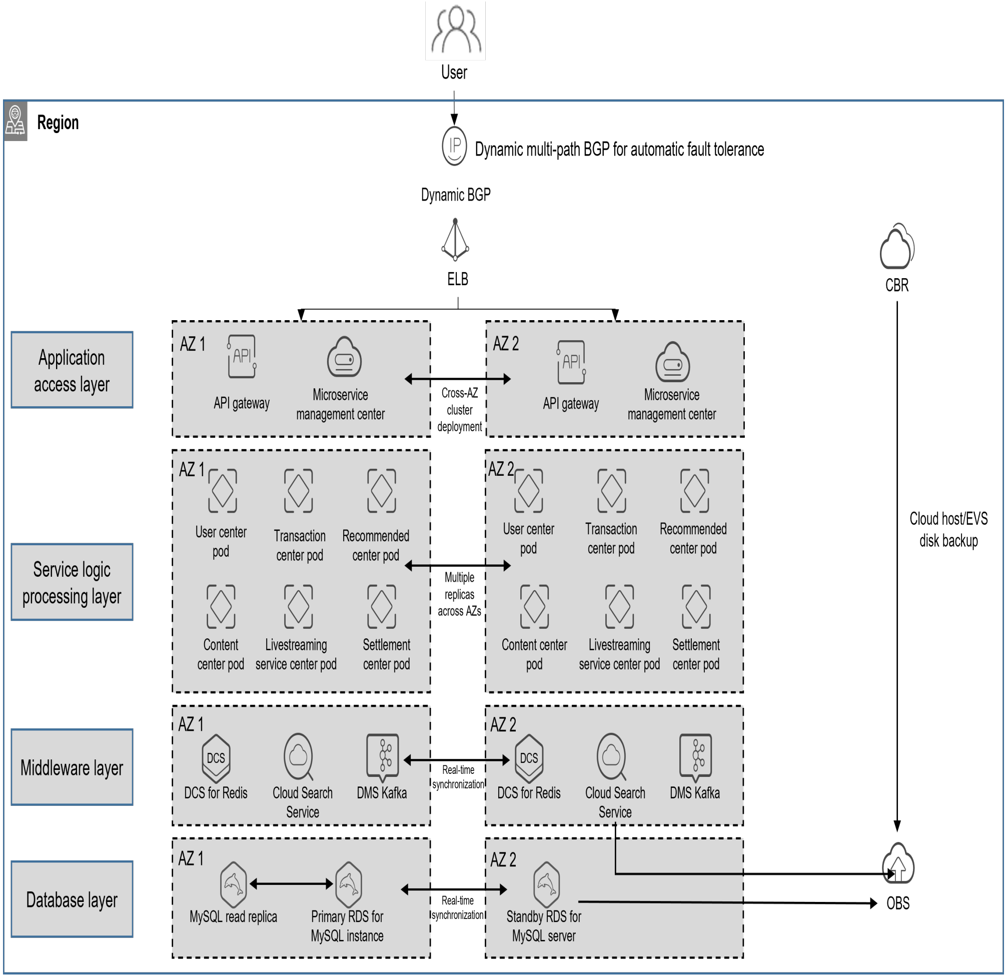
  Select a better solution considering the architecture to improve performance, such as using compression or caching.

5.4.5 Application Deployment Reference

5.4.5.1 Application Deployment Example

The following figure shows an example for deploying audio applications on the cloud.

Figure 5-15 Application deployment design example


User
Region
Dynamic multi-path BGP for automatic fault tolerance
Dynamic BGP
ELB
CBR
Application access layer
AZ 1
API gateway
Microservice management center
Cross-AZ cluster deployment
AZ 2
API gateway
Microservice management center
Service logic processing layer
User center pod
Transaction center pod
Recommended center pod
Content center pod
Livestreaming service center pod
Settlement center pod
Multiple replicas across AZs
Cloud host/EVS disk backup
Middleware layer
DCS for Redis
Cloud Search Service
DMS Kafka
Real-time synchronization
Database layer
MySQL read replica
Primary RDS for MySQL instance
Standby RDS for MySQL server
OBS

Key design points:

- Dynamic multi-path BGP can implement automatic fault tolerance of public network access and ensure HA.
- Huawei Cloud ELB is deployed in a cluster across AZs to ensure HA. A faulty data center does not affect services.
- The application access layer is deployed in a cluster across AZs. A fault AZ does not affect global services.
- The scheduling policy of Huawei Cloud CCE can implement cross-AZ deployment of multiple replicas of service container pods while ensuring HA across data centers.
- DCS for Redis is deployed in primary/standby mode across AZs to ensure HA. DMS for Kafka is deployed in a cluster across two or three AZs to ensure HA of messages. CSS can be deployed in a cluster across AZs to ensure that services are not affected by a faulty AZ.
- RDS for MySQL is deployed in primary/standby mode. Data is synchronized between the primary and standby instances in real time. If the primary instance is faulty, the standby instance can be quickly promoted to primary.
- Redis, Kafka, CSS, and RDS for MySQL can back up data to OBS buckets to prevent misoperations.

- CBR can be used to back up entire ECSs or EVS disks.

5.4.5.2 Reference Architecture Library

Haydn, a digital platform offered by Huawei Cloud, serves as a valuable resource for both partners and customers. To date, it has amassed over 700 reference architectures tailored to diverse business needs. Enterprises can explore these architectures aligned with specific service scenarios, leverage them to design robust solutions directly within Haydn, or even customize them to precisely align with their unique operational demands.

- **Searching for an architecture template**

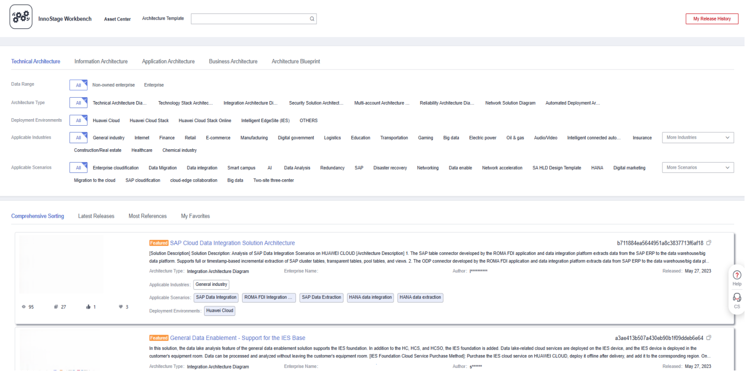
  Visit the Huawei Cloud official website, choose **Solutions** > **By Use Case** > **Haydn Solution Digital Platform**, and click **Try Now**. In the lower right corner, click **Architecture Templates** under **Solution Acceleration Field**.

  a. To query an architecture template, enter keywords such as the template name, applicable industry, or applicable scenario.

  b. Filter the templates by architecture type, deployment environment, applicable industry, and applicable scenario, or select more industries and scenarios from the **Applicable Industries** and **Applicable Scenarios** drop-down list boxes on the right.

  c. You can sort architecture templates by comprehensive ranking, latest update, popularity (most referenced), or your own favorite selection.

Figure 5-16 Haydn architecture design templates



Data Range: All | Non-owned enterprise | Enterprise

Architecture Type: All | Technical Architecture Dia... | Technology Stack Architec... | Integration Architecture Di... | Security Solution Architect... | Multi-account Architecture ... | Reliability Architecture Dia... | Network Solution Diagram | Automated Deployment Ar...

Deployment Environments: All | Huawei Cloud | Huawei Cloud Stack | Huawei Cloud Stack Online | Intelligent EdgeSite (IES) | OTHERS

Applicable Industries: All | General industry | Internet | Finance | Retail | E-commerce | Manufacturing | Digital government | Logistics | Education | Transportation | Gaming | Big data | Electric power | Oil & gas | Audio/Video | Intelligent connected auto... | Insurance | Construction/Real estate | Healthcare | Chemical industry

Applicable Scenarios: All | Enterprise cloudification | Data Migration | Data integration | Smart campus | AI | Data Analysis | Redundancy | SAP | Disaster recovery | Networking | Data enable | Network acceleration | SA HLD Design Template | HANA | Digital marketing | Migration to the cloud | SAP cloudification | cloud-edge collaboration | Big data | Two-site three-center

Featured SAP Cloud Data Integration Solution Architecture b711884ea5644951a8c3837713f6af18

[Solution Description] Solution Description: Analysis of SAP Data Integration Scenarios on HUAWEI CLOUD [Architecture Description] 1. The SAP table connector developed by the ROMA FDI application and data integration platform extracts data from the SAP ERP to the data warehouse/big data platform. Supports full or timestamp-based incremental extraction of SAP cluster tables, transparent tables, pool tables, and views. 2. The ODP connector developed by the ROMA FDI application and data integration platform extracts data from SAP ERP to the data warehouse/big data pl...

Architecture Type: Integration Architecture Diagram Enterprise Name: Author: l********** Released: May 27, 2023

Applicable Industries: General industry

Applicable Scenarios: SAP Data Integration | ROMA FDI Integration ... | SAP Data Extraction | HANA data integration | HANA data extraction

Deployment Environments: Huawei Cloud

95 27 1 3

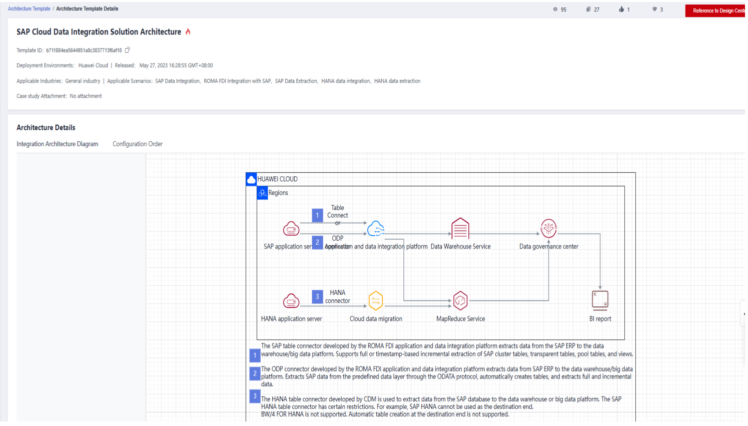
Featured General Data Enablement - Support for the IES Base a3ae413b507a430eb90b1f09ddeb6e64

In this solution, the data lake analysis feature of the general data enablement solution supports the IES foundation. In addition to the HC, HCS, and HCSO, the IES foundation is added. Data lake-related cloud services are deployed on the IES device, and the IES device is deployed in the customer's equipment room. Data can be processed and analyzed without leaving the customer's equipment room. [IES Foundation Cloud Service Purchase Method]: Purchase the IES cloud service on HUAWEI CLOUD, deploy it offline after delivery, and add it to the corresponding region. On...

Architecture Type: Integration Architecture Diagram Enterprise Name: Author: s****** Released: May 27, 2023

- **Architecture template details**

Figure 5-17 Haydn template details



SAP Cloud Data Integration Solution Architecture

Template ID: b711884ea5644951a8c3837713f6af18

Deployment Environments: Huawei Cloud | Released: May 27, 2023 16:28:55 GMT+08:00

Applicable Industries: General industry | Applicable Scenarios: SAP Data Integration、ROMA FDI Integration with SAP、SAP Data Extraction、HANA data integration、HANA data extraction

Case study Attachment: No attachment

1 The SAP table connector developed by the ROMA FDI application and data integration platform extracts data from the SAP ERP to the data warehouse/big data platform. Supports full or timestamp-based incremental extraction of SAP cluster tables, transparent tables, pool tables, and views.

2 The ODP connector developed by the ROMA FDI application and data integration platform extracts data from SAP ERP to the data warehouse/big data platform. Extracts SAP data from the predefined data layer through the ODATA protocol, automatically creates tables, and extracts full and incremental data.

3 The HANA table connector developed by CDM is used to extract data from the SAP database to the data warehouse or big data platform. The SAP HANA table connector has certain restrictions. For example, SAP HANA cannot be used as the destination end.
BW/4 FOR HANA is not supported. Automatic table creation at the destination end is not supported.

- **Referencing an architecture template**

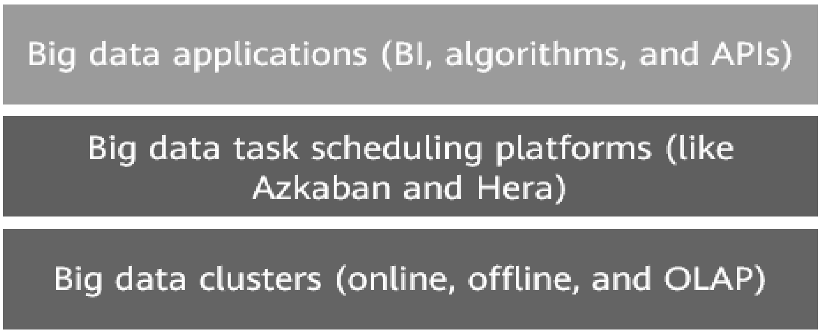
  In the upper right corner, click **Reference to Design Center** to reference the architecture template to a specified solution.

5.5 Designing a Big Data Architecture

5.5.1 Design Principles

The big data deployment architecture design involves big data clusters, task scheduling platforms, and applications. For details about the deployment architecture of big data applications, see **Designing the Application Architecture**.

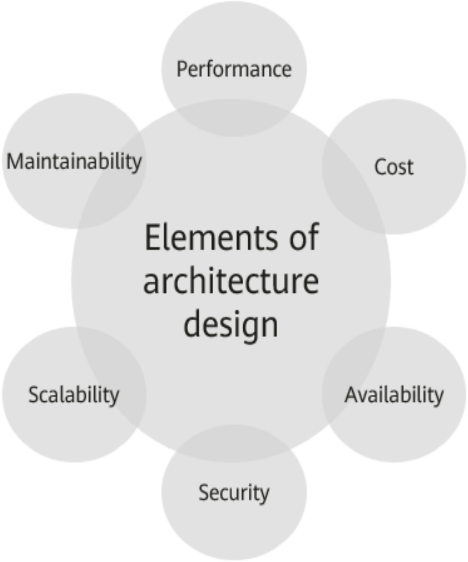
Figure 5-18 Big data architecture design classification


Big data applications (BI, algorithms, and APIs)
Big data task scheduling platforms (like Azkaban and Hera)
Big data clusters (online, offline, and OLAP)

Big data architecture design requires attention to these six key elements:

- Cost
- Availability
- Security
- Scalability
- Maintainability
- Performance

Figure 5-19 Six elements of architecture design



5.5.2 Big Data Cluster Design

When designing the big data cluster deployment architecture on the cloud, you are advised to comply with the following principles:

- **Big data cloud services preferred**: If the source is a self-built big data cluster, and there are equivalent cloud services available on the destination cloud platform which fulfill all functional, performance, and compatibility criteria with minimal assessed reconstruction workloads, you are advised to preferentially use big data cloud services when designing the deployment architecture for big data clusters. Use your existing setup if the destination cloud platform does not have the corresponding big data cluster components or if the destination cloud platform's big data cluster components have limited compatibility and require significant rework.
- **Minimum reconstruction**: If there is no special service requirement, avoid large-scale reconstruction. Design all components for the big data cluster in 1:1 benchmarking mode. Keep their versions consistent. Before a version upgrade, assess the adaptive reconstruction workloads.
- **Elastic and auto scaling**: When designing a big data cluster on the cloud, consider the elastic and auto scaling capabilities of the cluster. This means that compute and storage resources in the cluster can automatically increase or decrease based on workload requirements, thereby enhancing performance, boosting efficiency, and reducing costs.
- **Fault tolerance and high availability**: Big data clusters deployed on the cloud must have fault tolerance and high availability to ensure system reliability and stability. Use multiple replicas, redundant nodes, and failover mechanisms to ensure data and task persistence in the event of hardware or software failures.
- **Data security and compliance**: Big data clusters deployed on the cloud need strong data security and compliance measures. Implement appropriate data encryption, identity authentication, access control, and data isolation measures to safeguard sensitive data against potential security threats.

- **Cost-effectiveness**: When deploying big data clusters on the cloud, consider cost-effectiveness. Cloud service providers can provide elastic compute and storage resources, avoiding direct investment and maintenance costs on physical hardware. In addition, resources can be optimized and adjusted as required to minimize costs and improve resource utilization.

5.5.3 Big Data Task Scheduling Platform Design

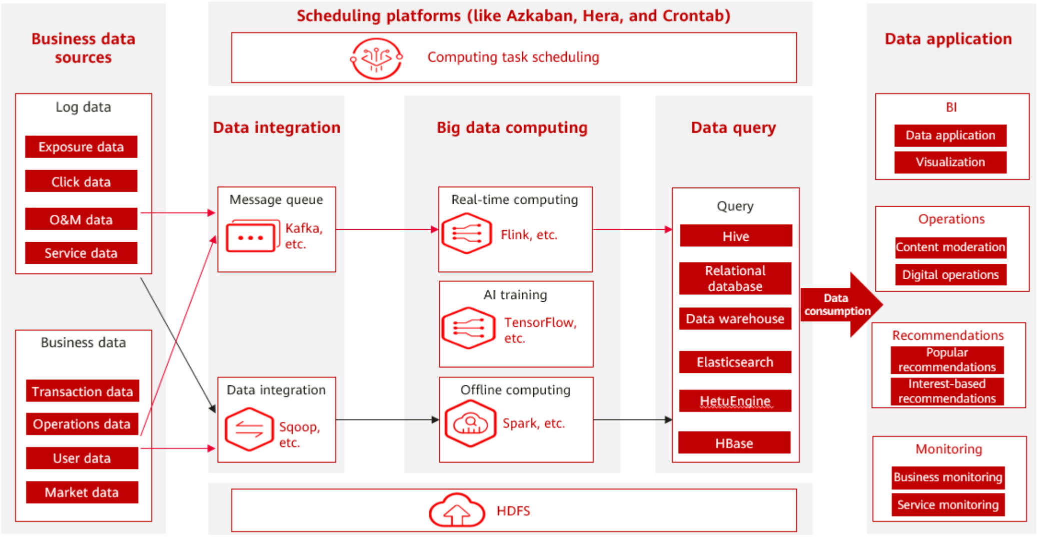
When designing the deployment architecture for the big data task scheduling platform on the cloud, you are advised to comply with the following principles:

- **Big data cloud services preferred**: If the source is a self-built big data task scheduling platform with required components, and there are equivalent cloud services available on the destination cloud platform which fulfill all functional, performance, and compatibility criteria with minimal assessed reconstruction workloads, you are advised to preferentially use big data cloud services when designing the deployment architecture. Use your existing setup if the destination cloud platform does not have the corresponding big data task scheduling components or if the destination cloud platform's big data task scheduling components have limited compatibility and require significant rework.
- **Minimum reconstruction**: If there is no special service requirement, avoid large-scale reconstruction. Design all components for the big data task scheduling platform in 1:1 benchmarking mode. Keep their versions consistent. Before a version upgrade, assess the adaptive reconstruction workloads.
- **Elasticity and scalability**: When deploying the big data task scheduling platform on the cloud, pay attention to the elasticity and scalability of the platform. Cloud environments offer flexible compute and storage resources that scale automatically to meet varying demands. Ensure that the task scheduling platform can quickly process increased workloads and support horizontal expansion to meet service requirements.
- **High availability and fault tolerance**: Ensure that the task scheduling platform on the cloud is highly available and fault-tolerant. Use the redundancy design and automatic fault recovery mechanism to ensure continuous system availability. For example, use multiple scheduling nodes and backup policies to prevent single points of failure (SPOFs) and ensure that tasks are not interrupted due to node failures.
- **Security and data protection**: The task scheduling platform on the cloud must have security and data protection mechanisms. Ensure that proper access control and encryption measures are provided for sensitive data and system components to prevent unauthorized access and data leakage.
- **Performance optimization**: When deploying the task scheduling platform on the cloud, pay attention to performance optimization. Optimize resource configurations, task scheduling algorithms, and data distribution policies to improve task execution efficiency and speed. You can also use services and functions provided by the cloud platform, such as cache and data prefetching, to optimize task execution performance.

5.5.4 Big Data Reference Architecture

The following figure shows a typical big data architecture. Data integration, storage, computing, scheduling, query, and application constitute a complete data flow.

Figure 5-20 Big data reference architecture



The big data architecture usually includes the following core components and processes. Enterprises can select cloud services or build big data components as needed.

- **Business data sources**

  Big data platforms collect information from various business data sources like sensors, website logs, mobile apps, and social media. Through data collection and extraction, raw data is collected and transferred to the big data platform for subsequent processing and analysis.

- **Data integration**

  Data integration is a process of integrating and converting data from different data sources. It includes operations such as data cleansing, data preprocessing, data format conversion, and data merging to ensure data consistency and accuracy.

- **Data storage**

  The big data platform must have efficient data storage capabilities to carry massive amounts of data. Distributed file systems like HDFS and columnar databases like HBase are common data storage solutions. These storage systems provide high reliability, scalability, and fault tolerance to support large-scale data storage and access.

- **Big data computing**

  Big data computing is a key step for processing massive amounts of data in a distributed, parallel, and real-time manner. Hadoop, Spark, and Flink are key computing frameworks that enable distributed computing and manage task scheduling. These computing frameworks can be used to perform complex computing and analysis tasks such as data processing, feature extraction, machine learning, and data mining.

- **Data query and analysis**

  A large amount of data stored on the big data platform requires flexible, high-performance query and analysis capabilities. This can be achieved by using an SQL query engine like Apache Hive or a distributed database like Elasticsearch. The tools and systems allow you to query, aggregate, and view massive datasets for better insights and informed decisions.

- **Task scheduling**

  Big data platforms often handle complex data tasks. Task scheduling systems (such as Azkaban) help manage and schedule data processing tasks. They allow you to set task dependencies, scheduling frequency, and retry policies for reliable and on-time task execution.

- **Data application**

  Big data platforms aim to deliver useful data applications for various businesses. Data applications can be real-time reports, visual dashboards, intelligent recommendation systems, and fraud detection systems based on big data analysis. Combining big data analysis results with business processes enables data-driven business decision-making and innovation.

5.5.5 Huawei Cloud Big Data Components

The following lists common Huawei Cloud big data service components. You can refer to these components when designing the big data deployment architecture.

- **MapReduce Service (MRS)**

  With MRS, you can deploy your Hadoop clusters with just a few clicks and manage them on Huawei Cloud. MRS is fully compatible with open-source APIs and can easily run big data components such as Hadoop, Spark, HBase, Kafka, and Storm. MRS can be customized based on business requirements, helping enterprises quickly build a massive data processing system.

  For details, see **MapReduce Service Documentation**.

- **Data Lake Insight (DLI)**

  DLI is a serverless big data compute and analysis service that is fully compatible with Apache Spark, Apache Flink, and Trino. It provides streaming, batch, and interactive data processing. DLI supports standard SQL and is compatible with Spark SQL and Flink SQL. It also supports multiple access modes, and is compatible with mainstream data formats. Data on cloud-based platforms such as CloudTable, RDS, GaussDB (DWS), CSS, OBS, ECS databases, as well as offline databases, can be explored using SQL or programs. This eliminates the need for complex ETL processes.

  For details, see **Data Lake Insight Documentation**.

- **Cloud Search Service (CSS)**

  CSS is a distributed search engine service based on Elasticsearch, fully hosted on Huawei Cloud. You can use it for structured and unstructured data search, and use AI vectors for combine search, statistics, and reports. Elasticsearch is an open-source distributed search engine that can be deployed in standalone or cluster mode. As the heart of the ELK Stack, Elasticsearch clusters support multi-condition search, statistical analysis, and create visualized reports of structured and unstructured text.

  For details, see **Cloud Search Service Documentation**.

- **GaussDB(DWS)**

  GaussDB(DWS) is a native cloud service based on Huawei converged data warehouse GaussDB. It is compatible with standard ANSI SQL-99, SQL:2003, PostgreSQL, and Oracle database ecosystems. GaussDB(DWS) comes in three types: standard data warehouse, stream data warehouse, and hybrid data warehouse.

  For details, see **GaussDB(DWS) Service Documentation**.

- **DataArts Studio**

  DataArts Studio can be interconnected with all Huawei Cloud data lake and database services which function as the data lake foundation, such as MRS Hive and GaussDB(DWS). It can also be interconnected with traditional data warehouses, such as Oracle and MySQL.

  For details, see **DataArts Studio Service Documentation**.

- **Data Ingestion Service (DIS)**

  DIS addresses the challenge of transmitting data from outside the cloud to inside the cloud. It builds data streams for custom applications capable of processing or analyzing streaming data. DIS continuously captures, transmits, and stores terabytes of data from hundreds of thousands of sources every hour, such as logs, Internet of Things (IoT) data, social media feeds, website clickstreams, and location-tracking events.

  For details, see **Data Ingestion Service Documentation**.

- **Cloud Data Migration (CDM)**

  CDM is an efficient and easy-to-use data migration service. Based on big data migration to the cloud and the intelligent data lake solution, CDM provides easy-to-use migration capabilities and can integrate a broad set of data sources into the data lake, making data migration and integration easier and more efficient.

  For details, see **Cloud Data Migration Service Documentation**.

- **Data Express Service (DES)**

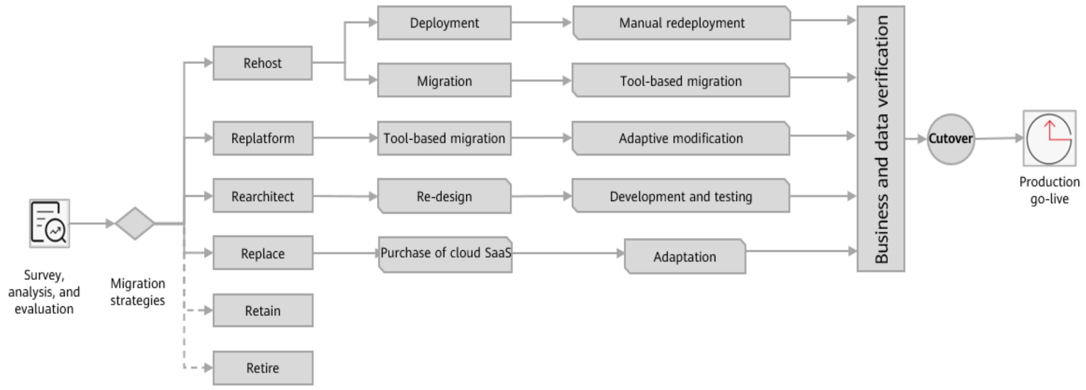
  DES is a transmission service for moving TB-level or hundreds of TB-level data to the cloud. DES allows you to transmit data by Teleport or by disk. Use disk mode for migrating data under 30 TB. Use Teleport mode for data ranging from 30 TB to 500 TB. For data over 500 TB, use Direct Connect.

  For details, see **Data Express Service Documentation**.

5.6 Formulating the 6 Rs Strategies

The 6 Rs strategies are six methods for migrating existing applications and data to the cloud. The following figure illustrates the strategies.

Figure 5-21 6 Rs strategies



The following table describes the meaning and application scenarios of the 6 Rs strategies.

Table 5-2 Descriptions and application scenarios of the 6 Rs strategies

Strategy	Description	Application Scenario
Retire	This strategy can stop applications or their components when they are unnecessary or there are more suitable alternatives. This is not a strict migration, but the elimination of existing applications.	● The applications are no longer used by businesses. ● The functions of the applications have been replaced by other systems. ● The applications require high maintenance yet provide minimal business benefits. ● The applications are outdated, and they are difficult to maintain and upgrade.
Retain	This strategy keeps applications in their current states without migrating them. This is usually a short-term strategy or temporary measure during ongoing IT strategy development.	● The applications depend on specific hardware or software and cannot be easily migrated. ● Application migration is risky, and there is no urgent need for migration in the short term.

Strategy	Description	Application Scenario
Rehost	Also known as "direct migration" or "lift and shift", this strategy is used to migrate applications from an on-premises data center to the cloud platform without making any change to the applications. This strategy usually employs specific tools to convert virtual machines or physical servers into cloud-based virtual machines.	● A quick migration to a cloud platform is required to reduce costs or improve availability. ● The migration must be expedited as time is critical. ● There is a lack of in-depth understanding of applications or resources to modify the code.
Replatform	This strategy involves making slight changes to applications during migration to fit the cloud platform. For example, migrate applications from an on-premises database to a cloud-based one. This does not involve modifying the core application code.	● PaaS services, such as databases and message queues, on the cloud platform are required to ease the O&M burden of self-built databases and message queues. ● The performance or scalability of applications needs to be improved. ● No large-scale code modification is required, but optimizing application performance on the cloud platform is required.
Rearchitect	This strategy involves rewriting or refactoring application code for them to fit the cloud-native architecture. For example, a monolithic application can be refactored to a microservice architecture or switched to a serverless, event-based model.	● The performance, scalability, and maintainability of applications need to be significantly improved. ● Cloud-native technologies, such as containers and serverless computing, are required. ● The application architecture is outdated and difficult to maintain and scale.

Strategy	Description	Application Scenario
Replace	This strategy involves replacing existing applications with new applications or services. It typically means purchasing Software as a Service (SaaS) products or other new applications.	● Existing applications cannot meet business requirements. ● Maintaining existing applications costs too much, and there are more suitable alternatives. ● Mature SaaS products are available in the market to meet business requirements. ● New functions and services need to be quickly deployed.

Among the 6 Rs strategies, only Rehost, Replatform, and Rearchitect are involved in migration to the cloud. The following table compares these three strategies. Enterprises can select the most suitable migration strategy based on their service requirements, application scenarios, and the migration risks, period, costs, difficulty, and benefits of each strategy.

Table 5-3 Migration strategy comparison

Migration Strategy	Risk	Migration Period	Migration Cost	Migration Difficulty	Benefit
Rehost	Low	Short	Low	Low	Low
Replatform	Medium	Medium	Medium	Medium	Medium
Rearchitect	High	Long	High	High	High

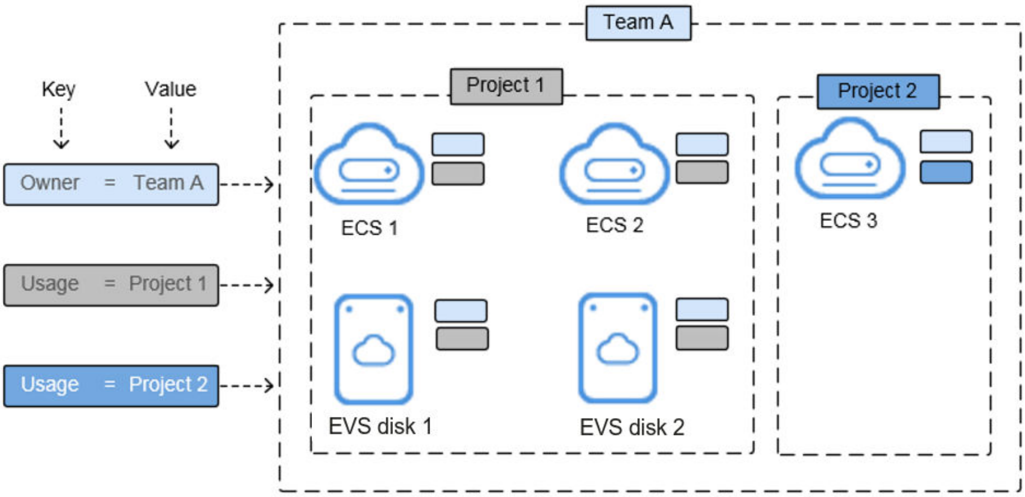
5.7 Designing a Tag Solution

5.7.1 Overview

Tags are key-value pairs that help identify and organize cloud resources. You can use tags to identify and classify cloud resources by dimensions like usage, owner, or environment, and then filter resources, classify costs, and set fine-grained permissions based on tags. This simplifies resource management and lowers costs.

In this example, you assign two tags to each cloud resource. Each tag is a key-value pair. The key of one tag is **Owner**, and the key of another tag is **Usage**.

Figure 5-22 Resource tag example



Huawei Cloud Tag Management Service (TMS) is a visualized service for fast, unified tag management. It allows you to centrally manage tags and categorize resources for different regions and services. Resource tags can also facilitate permission and billing management. You can access TMS through a unified console or APIs.

For details about the functions and usage of TMS, see the **official documentation**.

5.7.2 Best Practices for Tags

Tag Design Principles

Follow these tips when designing tags for Huawei Cloud resources:

1. Avoid keeping personally identifiable information (PII) or other sensitive information in tags.
2. Apply consistent case-sensitive formatting to tags on all resources.
3. Keep tags brief but detailed enough to convey specific information. Avoid making them unnecessarily long.
4. Identify tag application scenarios like cost management, O&M, automation, fine-grained permission control, data classification, and security operations. Formulate tag key and value specifications based on these scenarios.

Tag Key and Value Specifications

Formulate enterprise-wide tag key and value specifications based on tag design principles. Follow the specifications when tagging cloud resources. The following table provides examples for tag key and value specifications.

Table 5-4 Examples for tag key and value specifications

Scenario	Tag Key	Allowed Tag Value
Cost management	Department	Marketing, Engineering, Sales, Service, Research, etc.
Cost management	Application	CRM, ERP, HRM, Financial management system, etc.
Cost management	CostCenter	123, 456, 789, etc.
Permission management	Environment	Development, Test, Stage, Product, etc.
Permission management	Layer	DB, App, Web, etc.
Data classification	DataClass	Public, Private, Confidential, etc.
Security operations	Compliance	PCI-DSS, HIPPA, etc.
O&M and automation	Status	Active, Inactive, Deprecated, etc.

Tag Policies

Always follow the tag key and value specifications when tagging cloud resources. If not, tags might get mixed up, and their effectiveness could suffer. Huawei Cloud provides **tag policies** to help you manage tags added to cloud resources in your Huawei Cloud account.

For example, a tag policy can specify that a tag attached to a resource must use the case treatment and tag key and value specifications defined in the tag policy. If the case, key, and value of the tag do not comply with the tag policy, the resource will be marked as non-compliant.

Currently, tag policies can be used as preventive governance policies. Specifically, if enforcement is enabled for a tag policy, non-compliant tagging operations will be prevented from being performed on specified resource types.

You can attach tag policies to the root OU, other OUs, and accounts within your organization. When you attach a tag policy to the root OU and other OUs, all their child OUs and member accounts inherit that tag policy. The effective tag policy for an account specifies the tagging rules that apply to the account. It is the combination of tag policies that account inherits and tag policies directly attached to that account.

Tag Restrictions and Requirements

1. Each resource can have a maximum of 20 user-created tags. Note that system-created tags starting with **_sys_** are reserved for Huawei Cloud and are not counted in the 20 tags.

2. For each resource, each tag key must be unique and can have only one tag value.
3. A tag key must contain 1 to 128 Unicode characters in UTF-8 format.
4. A tag value must contain 0 to 255 Unicode characters in UTF-8 format.
5. Tag keys and values are case-sensitive. You are advised to use tag policies to apply consistent tag case treatment rules for all resource types. For example, choose one from **HuaweiCloud**, **huaweicloud**, or **Huaweicloud** for all resource types.

5.7.3 Typical Application Scenarios

Cost Management

You can tag each cloud resource to track and manage costs effectively, improving the efficiency and convenience of enterprise financial management. Tags are also an important tool for financial departments to implement refined management. Tags can enable financial personnel to better understand, manage, and optimize enterprises' IT investment. This includes but is not limited to the following aspects:

1. **More accurate cost control**: Reasonable and standard tags enable financial personnel to accurately track the actual cost of each resource. This helps identify and optimize high-cost resource usage, thereby achieving cost savings.
2. **More effective budget management**: Financial personnel allocate cloud expenditures to specific budget categories, making budget formulation and adjustment more scientific and reasonable. This simplifies tracking real costs.
3. **More transparent expense allocation**: For a company with multiple departments or projects, tags enable costs to be directly associated with corresponding departments or projects, ensuring fair and transparent expense allocation. This simplifies internal cost tracking and offers transparent breakdowns for customer and stakeholder reports.
4. **Simplified bill management and analysis**: Tags can significantly simplify bill parsing and help financial personnel quickly understand and review bill content. Tags can also be used to analyze expenses thoroughly and pinpoint ways to cut costs.
5. **More accurate decision-making**: Accurate tag information provides a factual data foundation for executives, helping them make more informed investment decisions.

The Cloud Center of Excellence (CCoE) team should work with someone skilled in finances to create effective resource tag policies for cost and finance management.

Process of Using Tags to Manage Costs

- **Create a tag rule.**

  The CCoE team works with enterprise finance personnel to develop standardized tag naming rules. For example, the team sets keys and values of tags used for cost management based on the tag design principles by subsidiary, operating environment, department, business unit, and application system.

- **Set a resource tag.**

  Create a tag for each cloud resource based on the preceding tag naming rules. To prevent cloud resource operators from adding tags randomly, you can use tag policies provided by Huawei Cloud to enforce tag naming rules.

- **Activate cost tags**.

  Activate cost tags before classifying your cloud resource costs by categories like usage, environment, or department.

  For details, see **Activating Cost Tags**.

- **Analyze and optimize costs.**

  After cost tags are activated, go to the **Cost Analysis** page to summarize and filter cost data by cost tag. The CCoE team can identify idle or overused resources using tags and optimize costs accordingly.

O&M Management

Tags are an important part of cloud O&M practices. They can improve work efficiency and help cloud O&M teams better control complex IT environments. Tags are important for cloud O&M in the following aspects:

- **Simplified resource management**: Tags allow O&M personnel to manage and organize cloud resources in a more flexible manner. For example, resources can be classified by project, environment (like development, test, or production), owner, or cost center for better visibility and management.
- **Lifecycle management**: Resource lifecycle information is tagged to simplify planning for starting or removing resources and help O&M personnel manage resources effectively.
- **Compliance and security**: Security operations personnel can add tags for resources based on application data sensitivity levels. This ensures applications and data comply with all relevant security and privacy regulations or internal/external audit requirements.
- **Troubleshooting assistance**: O&M personnel can use tags to quickly locate affected resources, accelerating problem diagnosis and resolution.
- **O&M automation assistance**: O&M personnel can compile scripts or configure rules based on standard tags to implement automation tasks. For example, instances with specific tags are automatically started or stopped, or resources with deletion tags are periodically released. This significantly reduces the need for manual intervention and reduces the risk of human errors.
- **Performance optimization assistance**: O&M personnel can add key metrics like CPU and memory usage to tags to remind related personnel to take actions for performance optimization.
- **Security hardening assistance**: O&M personnel can assign security policy implementation permissions based on tags to ensure that only authorized users can access or operate certain sensitive resources.
- **Disaster recovery assistance**: Tags can be created for applications that need to be quickly restored during disaster recovery to ensure that these applications can be restarted immediately after a disaster occurs.

5.8 Piloting Cloud Migration

5.8.1 Reasons for Pilot Cloud Migration

Pilot cloud migration is an important step before large-scale cloud migration. It helps enterprises fully understand and evaluate various factors before large-scale migration. By piloting the cloud migration process and related configurations, enterprises can identify risks in advance and provide experience for large-scale cloud migration.

- **Risk control**: Cloud migration is a complex process involving different systems and services. Pilot cloud migrations let businesses test the full process on a smaller scale. They help spot and fix issues early, ensuring successful larger migrations. In addition, pilot cloud migrations can help enterprises identify risks and challenges. For example, some firms might struggle with adapting to the cloud. Through pilot cloud migrations, these problems can be exposed in a small scope and corrective measures can be taken in a timely manner to reduce risks during a full-scale migration.
- **Feasibility verification**: Pilot cloud migrations can verify whether enterprise applications and data can be migrated to the cloud. By selecting a small number of applications or businesses for pilot cloud migrations, enterprises can evaluate whether the compatibility, performance, security, and reliability of businesses in the cloud environment meet requirements. If some applications are not suitable for migration to the cloud, enterprises can re-plan migration strategies or find alternative solutions based on the evaluation results to prevent applications that are not suitable or difficult to migrate from being directly put into the production environment.
- **Experience accumulation**: Pilot cloud migrations help technical teams and business personnel of enterprises build hands-on experience. During this process, they can learn and get familiar with the features, functions, and best practices of the cloud platform, understand the migration tools and processes, and accumulate related knowledge and skills to prepare for subsequent full-scale migrations.
- **Priority determination**: Enterprises can evaluate the migration priorities of different applications or businesses through pilot migrations. Based on the pilot migration results, enterprises can determine which applications or businesses have little impact on cloud migration and migrate them first to promote the entire migration process in an orderly manner.
- **Performance optimization**: Pilot migrations allow enterprises to spot and fix performance issues early. Testing on a smaller scale helps check key application metrics like latency, throughput, and response time in the cloud. This approach identifies bottlenecks beforehand, enabling adjustments for better performance and user experience before the formal migration.
- **Cost control**: Pilot migrations help enterprises assess cloud migration costs effectively. By testing during this phase, enterprises gain clear insights into pricing models, resource usage, and potential hidden expenses. This prevents budget overruns and improves resource efficiency, ensuring better cost management.

- **Team cooperation**: A successful cloud migration depends on effective teamwork among all involved parties, such as the O&M team, development team, test team, cloud service provider, and third-party system provider. Pilot migrations help teams practice working together, and enable them to spot potential issues, assess risks, and plan solutions to ensure seamless collaboration during full-scale migration.

5.8.2 Selection of Pilot Applications

While selecting applications for a pilot cloud migration, consider whether the following conditions for this process are met in a holistic manner:

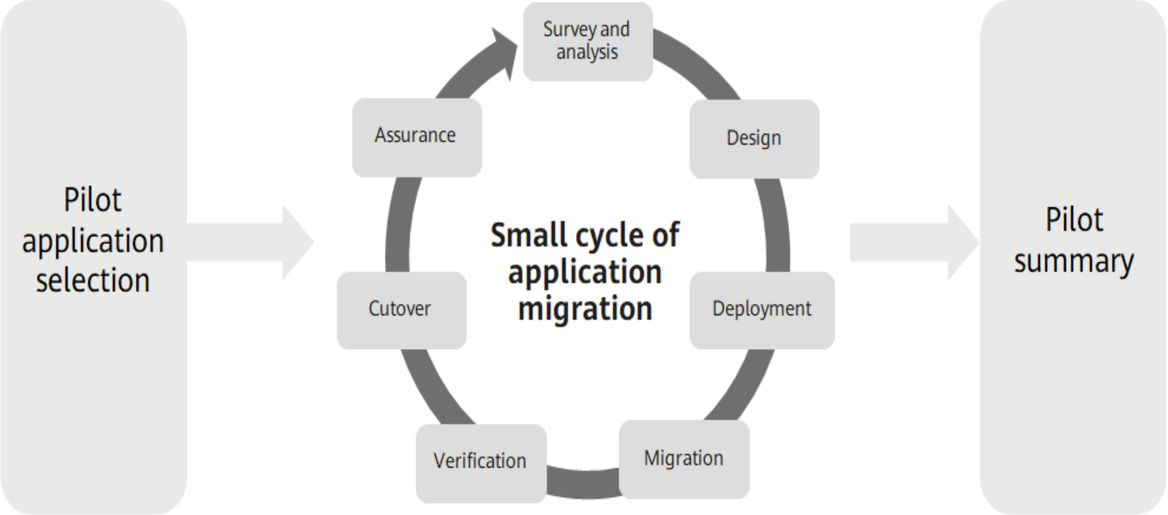
- **Willingness for cloud migration**: Different business departments have different willingness to migrate their businesses to the cloud. Preferentially select those who are highly motivated to migrate their businesses to the cloud, have sufficient workforce and time, and actively participate in the migration.
- **Business importance**: Based on the existing applications and businesses of the enterprises, select those with high importance but low impact on normal operations for pilot cloud migrations.
- **Cloud migration benefits**: Select applications for which cloud migration benefits can be easily quantified, such as reducing costs, improving availability, and quickly deploying businesses. Use these applications as pilots to show the benefits of cloud migration.
- **Implementation difficulty**: Select some applications with low implementation difficulty as pilots based on the implementation capability of the IT departments of enterprises.
- **Impact on businesses**: Consider how a migration affects other business processes and data flows to prevent disrupting normal operations.
- **Security**: Consider data security and compliance with related laws and regulations to avoid security risks or violations of related laws and regulations.
- **Testability**: Pilot cloud migrations can be used to verify migration solutions, identify potential problems, and continuously test and optimize methods to ensure successful cloud migration. Applications with high testability can be selected to fully verify the solutions and pave the way for subsequent large-scale cloud migrations.

5.8.3 Pilot Cloud Migration Execution and Summarization

Pilot Cloud Migration Execution

After selecting pilot applications, execute a pilot cloud migration based on the small cycle application migration process, and output a pilot summary.

Figure 5-23 Small-cycle migration process



Pilot Cloud Migration Summarization

Pilot cloud migration summarization aims to summarize the achievements, experience, and lessons learned from pilot projects, and provide guidance and improvement directions for subsequent large-scale migration. This is critical for enterprises to evaluate the benefits, risks, and challenges of cloud migration and to formulate effective migration strategies. You can summarize a pilot cloud migration from the following aspects:

1. **Objectives and scope**: Summarize the objectives and scope of the pilot cloud migration, specify the expected results, and describe the applications, systems, or business processes involved in the pilot, as well as the time, place, and participants of the pilot.
2. **Migration methods and strategies**: Summarize the used migration methods and strategies, and describe the used technologies, tools, and processes, as well as key decisions related to the migration.
3. **Achievement evaluation**: Evaluate the effect of the pilot cloud migration, including the number of applications that are successfully migrated, problems and challenges during the migration, solutions, and improvement measures, and summarize the impact and benefits of the pilot migration on enterprise businesses.
4. **Technology and performance evaluation**: Evaluate the performance and stability of the system and applications after the migration. Consider the scalability, response time, and data transmission speed of applications, and summarize the impact of the migration on system performance and user experience.
5. **Cost-benefit analysis**: Analyze the impact of the cloud migration on enterprise costs, including cost savings, resource utilization optimization, and changes in maintenance and support costs.
6. **Security and compliance evaluation**: Evaluate the security and compliance of the pilot cloud migration, and summarize the compliance level and security risk control level of the pilot migration in terms of post-migration data security, access control, and compliance requirements.

7. **Learning and lessons:** Summarize the learning and lessons learned during the pilot migration, including success factors, failure causes, and identified best practices. Document technical and management findings to help better plan and execute subsequent large-scale migrations.
8. **Suggestion and improvement measure:** Based on the pilot migration results, provide further suggestions and improvement measures to guide the future large-scale migration planning, including suggestions on optimizing the migration process, strengthening training and communication, and further measures in terms of security and compliance.
9. **Follow-up plan and risk management**: Create follow-up plans for subsequent stages of the pilot migration, covering large-scale migration, resource allocation, and risk and problem resolution. Highlight how the pilot migration guides future efforts and its overall influence. Outline risk management approaches and solutions.

5.9 Planning Migration Waves

5.9.1 Overview

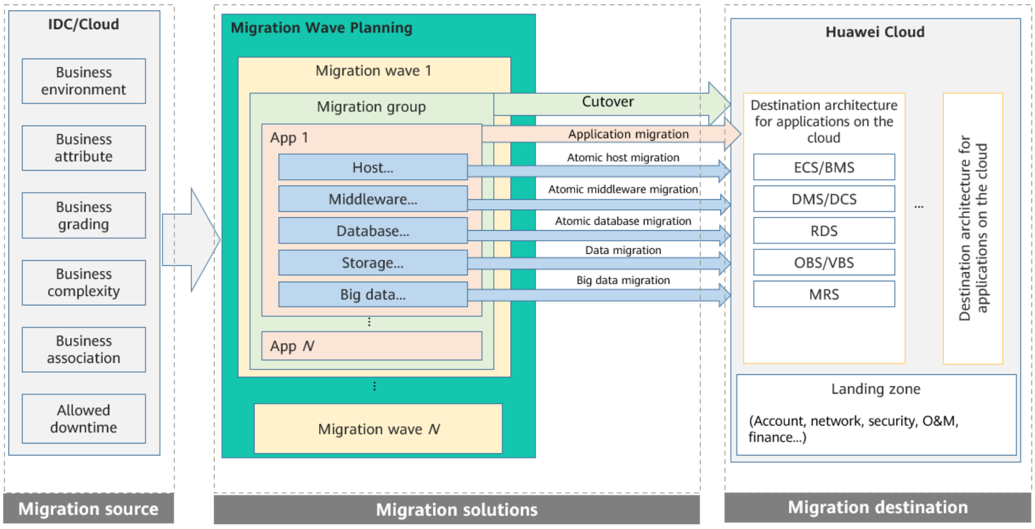
Why Is Wave Planning Required?

Cloud migration wave planning is used to divide the migration process into multiple stages for more effective transfer of enterprise applications to the cloud. It aims to lower migration complexity, conduct more effective migration in proper time windows, streamline resources, and analyze risks in each migration phase, thereby minimizing the cloud migration time.

Wave planning is an important task for enterprise cloud migration. Before wave planning, you need to understand the following concepts:

1. **Migration group**: It is a collection of applications and infrastructure that have dependencies (including environment dependencies), and includes apps, hosts, storage devices, databases, and middleware.
2. **Migration wave**: It is a combination of one or more migration groups that are executed on the same expected start date and end date. One migration wave may include multiple migration groups.
3. **Migration priority**: It indicates the migration sequence of applications.
4. **Migration wave planning**: It outlines the migration groups, waves, priorities, objects, and timelines for each migration phase.

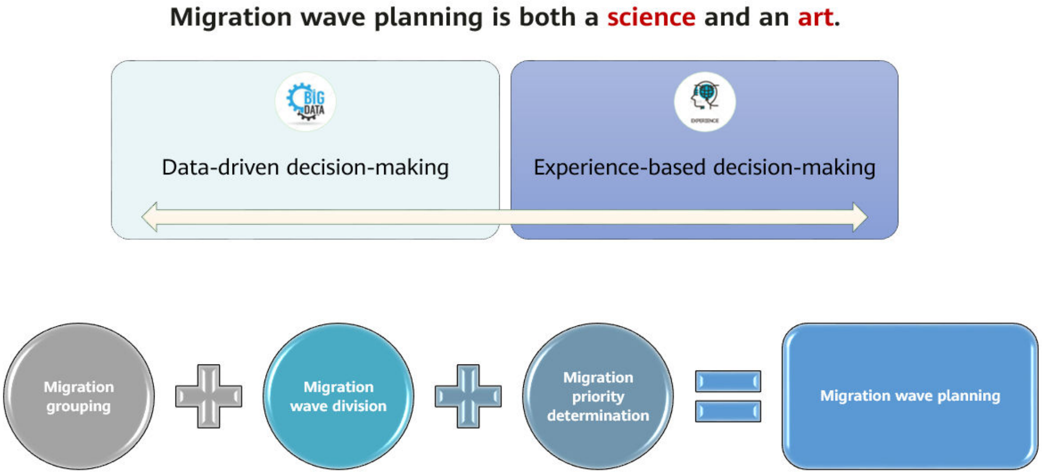
Figure 5-24 Migration overview



5.9.2 Migration Wave Planning Method

Migration wave planning is both a science and an art. It often hinges on a blend of empirical data and the seasoned insights of experts. It includes migration grouping, wave division, and priority determination.

Figure 5-25 Wave planning


Migration wave planning is both a science and an art.
BIG DATA
Data-driven decision-making
EXPERIENCE
Experience-based decision-making
Migration grouping
Migration wave division
Migration priority determination
Migration wave planning

1. Migration grouping

Migration grouping is used to group migration objects based on their dependencies. A group of applications and infrastructure (including applications, hosts, storage devices, databases, and middleware) that have strong dependencies are divided into a migration group. A group of applications and infrastructure must be migrated in the same wave and cut over together.

There are three types of dependencies: shared data dependency, shared server dependency, and inter-application communication dependency.

There are strong dependencies and weak dependencies. Take the shared data dependency as an example. Applications A, B, and C are connected to DB 01. A and B perform many read and write operations per second. C performs batch processing jobs during off-peak hours every night. A and B are tightly coupled with DB 01, while C is loosely coupled with DB 01. A and B must be deployed in the same migration group and migrated together with DB 01. C can be migrated independently and can be deployed in another migration group if necessary.

2. Migration wave division

Enterprise cloud migration is usually performed in waves. A wave can contain one or more migration groups. Migration waves are divided based on the following rules:

- **Applications with strong association should be placed in one wave based on the association analysis result.**

  Applications with strong association should be migrated together in one wave to avoid the impact of on- and off-cloud access latency on businesses.

- **The most appropriate time span for a wave is four to eight weeks.**

  Each wave needs to be sized appropriately. In this way, sufficient workforce and technical resources can be provided to perform the migration, thereby minimizing risks. The best practice in the industry is that a wave spans four to eight weeks, during which deployment, migration, and cutover are performed. This time span excludes the preparation period.

- **The size of a wave cannot be too large, or migration risks may increase.**

  According to the best practice in the industry, the number of migration objects in a wave cannot exceed 20 applications, 150 servers, and 30 databases, or there may be great challenges and high risks (such as task failures and rollbacks). If this rule is not obeyed, you are advised to perform a rigorous check.

  If a wave is large, identify strong and weak associations, disconnect weak associations, and divide them into smaller waves to reduce risks.

- **Systems of the same supplier should be migrated to the cloud in the same wave or adjacent waves.**

  Multiple systems of the same supplier are highly coupled. If these systems are migrated to the cloud at the same time, the supplier can pool human resources to support the migration in a short period of time, ensuring collaboration between project teams and facilitating smooth cloud migration.

- **Different migration environments are placed in different waves.**

  The test environment is migrated before the production environment, which can significantly reduce the migration risks of the latter.

3. Migration priority determination

The following table describes factors that affect the cloud migration priority.

Table 5-5 Factors affecting the cloud migration priority

Factor	Result
Willingness for cloud migration	Migrate workloads with high willingness first.

Business environment	Migrate workloads for the test environment first.
Business importance	Migrate general workloads prior to critical workloads.
Business association	Migrate workloads with simple associations first.
Infrastructure complexity	Migrate workloads with simple underlying infrastructure and a small number of instances first.
Allowed downtime	Migrate workloads with longer allowed downtime first.
Migration strategy	Migrate applications that can use the rehosting approach ("lift-and-shift") before those requiring re-architecting.

Business departments' willingness for cloud migration is crucial and should be prioritized first, followed by other factors. You can score each factor for further determination of priorities.

Table 5-6 Reference scoring table for migration priority determination

Category	Factor	Reference Score
Business environment	Development	5
	Test	3
	Production	1
Business importance	Minor	5
	Major	3
	Core	1
Association	Simple	5
	Complex	3
	Very complex	1
Infrastructure complexity	Simple (1–3 instances)	5
	Complex (4–10 instances)	3
	Very complex (11 instances and more)	1
Allowed downtime	More than 120 minutes	5
	60 to 120 minutes	3
	Less than 60 minutes	1
Migration strategy	Rehost	5
	Replatform	3

Category	Factor	Reference Score
	Rearchitect	1

Application Migration Wave Planning Example

Table 5-7 Example table for wave planning

App	Strateg y	Wave	First Wave	Second Wave	Third Wave	Fourth Wave
N/A	N/A	N/A	2025.01.01 – 2025.03.31	2025.04.01 – 2025.06.30	2025.07.01 – 2025.09.30	2025.10.01 – 2025.12.31
The table header information provided here is for reference only. Fill the table with real business data.						

5.9.3 Big Data Migration Wave Planning

When migrating big data to the cloud, determine whether to migrate the data in waves or in whole. The guidelines are as follows:

- **Scenarios suitable for migration in whole:**
  - Small scale: For a big data platform that has a small amount of data (TB-level) and a small number of computing tasks, you can deploy the platform on the cloud and then migrate all metadata, data, and tasks.
  - Complex association: Big data tasks are associated with each other and are difficult to split.
- **Scenarios suitable for migration in waves**: Large-scale big data with clear associations

  Big data platforms handle massive amounts of data (PB-level to EB-level), along with numerous computing tasks. Although the data scale is large, the association between computing tasks is clear. For example, you can sort out tasks and split big data by service domain and classify associated data, tasks, and applications into one wave for migration. Migration in waves effectively lowers risks, simplifies the process, and boosts efficiency.

Big data migration is usually performed in waves by subject area. Business functions determine how you classify subject areas. You can group related data that shares similar business logic, such as sales, supply chain, and log processing, into a subject area. Each subject area has a dedicated data processing process, analysis model, and related business logic to meet specific business requirements and analysis objectives. The reference principles for planning big data migration waves are as follows:

- **Wave by subject area**: Data correlation and task correlation need to be considered. Data correlation refers to the process of placing data with similar business logic, mutual dependency, or close relationship in the same wave to ensure consistency and integrity. Task correlation refers to the process of placing dependent tasks and data in the same wave. This ensures that tasks operate with accurate data while maintaining proper order and consistency. Based on the two correlations, subject areas can be divided into multiple migration waves, and related tasks and data flows are centralized in the same waves, improving migration efficiency and reducing risks.

- **Minimized number of waves**: During big data migration, data extraction, conversion, and loading are performed. Each operation increases the complexity and risk and affects data consistency. Therefore, the number of waves should be minimized.
- **Independent waves**: Ensure that different waves are independent and loosely coupled, and there are few dependent tasks and data flows. Independent wave division reduces the impact on other business domains during migration.
- **Intra-wave tight coupling**: Ensure that each wave contains highly correlated subject areas and interdependent tasks and data flows, including data sharing scenarios.
- **Business continuity**: Business interruption must be avoided during migration. Big data application systems closely related to a subject area must be deployed in the same wave to reduce business interruption risks.
- **Migration priority sorting**: Prioritize subject areas based on the business priority, migration complexity, and data volume. Generally, start by migrating smaller or simpler subject areas before moving on to larger or more complex ones.

5.10 Cost Budget Plan

Enterprises can plan and control their budget using Huawei Cloud's Cost Center during cloud migration. With Huawei Cloud Cost Center, enterprises can manage and monitor their cloud expenditures, improving resource utilization and reducing unnecessary expenditures. Proper budget planning and continuous cost optimization will help enterprises enhance return on investment (ROI) in the cloud environment and achieve successful digital transformation.

You can create a budget plan based on the following information:

- Forecasting based on historical cost data: Use the cost analysis function of Cost Center to forecast future costs based on historical expenditures.
- Forecasting based on business drivers: Forecast cost changes based on future business requirements (such as scale-out, promotion, and new service rollout).
- Create a budget plan based on historical data and future business requirements.

Huawei Cloud Cost Center provides you with budget templates to simplify your budget creation process. You can quickly create a cost budget and create a budget report for a budget task. Huawei Cloud will send you a budget report on the scheduled report date.

For details, see **Cost Center Documentation**.

5.11 Anti-patterns in Solution Design

During cloud migration solution design, you may encounter some anti-patterns. If these anti-patterns are not identified and avoided, the system performance and security may be reduced, unnecessary costs may be increased, maintenance may be difficult, and even the project may fail. The following are some common anti-patterns for cloud migration solution design:

- **Improper resource configuration**

  Resources are not properly configured during destination architecture design. This results in resource over-allocation or under-allocation, increasing costs or affecting performance.

  Optimization suggestion: Select proper cloud resource specifications based on business requirements and application characteristics. You can use the auto scaling policy, properly set the proportion of yearly/monthly and pay-per-use resources, periodically monitor resource usage, and continuously optimize and adjust resources.

- **Single point of failure**

  High availability is not considered during architecture design. As a result, key components become single points of failure. If there is a fault, the entire system cannot work properly.

  Optimization suggestion: Use redundant designs and load balancing to distribute critical application services across nodes, enhancing system reliability and availability.

- **Geographical distribution of businesses not considered**

  The geographical distribution of businesses is not considered during the cloud deployment architecture design, causing slow access and bad user experience.

  Optimization suggestion: Select proper regions and AZs based on user distribution to ensure that users can access the nearest application instances.

  If the traditional architecture is directly migrated to the cloud without necessary adaptation and optimization, problems such as stability and poor experience may occur.

  Optimization suggestion: Evaluate the application architecture and refactor the architecture to adapt to the cloud environment. Leverage cloud-native features to improve the availability, scalability, and performance of the architecture.

- **Non-compliance with laws and regulations**

  Compliance requirements for data storage and access are not considered during the cloud deployment architecture design, which may cause legal risks and data leakage.

  Optimization suggestion: Ensure that you understand the relevant laws, regulations, and compliance requirements when designing a cloud migration solution. Establish data governance and security policies to ensure that data storage, access, and processing comply with local laws and regulations.

By identifying and avoiding these anti-patterns and referring to industry best practices and success cases, you can design cloud migration solutions more scientifically, thereby better leveraging the advantages of the cloud and unleashing the power of cloud migration.

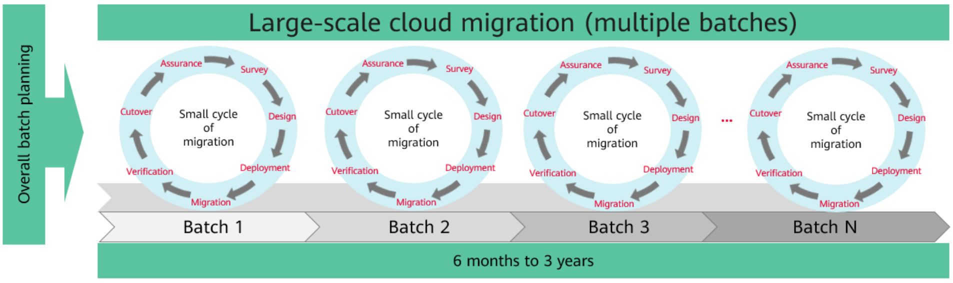
6 Adoption Implementation

6.1 Overview

Enterprises usually start the large-scale migration to the cloud after a pilot project is successful. The large-scale cloud migration can be completed in two ways:

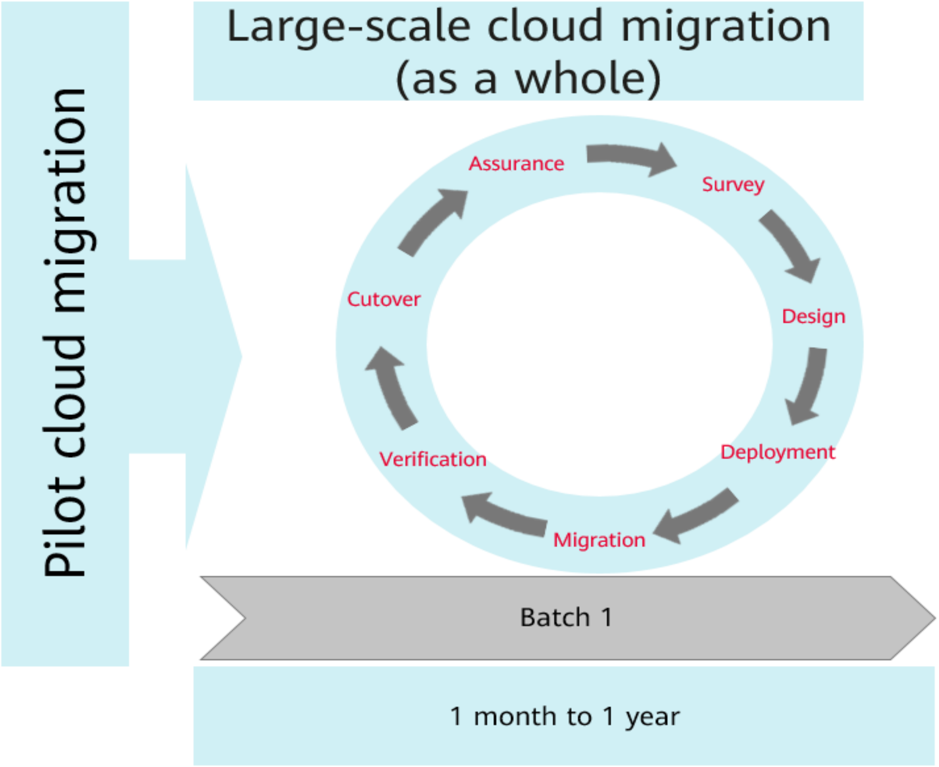
- **Batch migration**: This approach is best for divisible workloads. Enterprises typically divide the migration process into multiple batches, and complete them one by one. The following figure shows the process of batch migration.

Figure 6-1 Batch migration



- **Integrated migration**: This approach applies to systems where applications are associated with each other in a complex manner and cannot be migrated in batches. The following figure shows the integrated migration process.

Figure 6-2 Integrated migration



6.2 Implementation Team Establishment

An efficient and professional implementation team is the key to ensuring the success of the cloud migration project. The team is responsible for executing the cloud migration plan, performing the migration, and ensuring the success of each task in the cloud migration. The cloud migration implementation team should consist of professionals from different domains. Enterprises can set up a comprehensive and professional cloud migration implementation team based on the preceding CCoE organizational structure and role responsibilities. The following are the necessary members of the team.

- **Project manager**: a member of the Project Management Office (PMO) or an IT department member with extensive experience in project management. The project manager manages the entire cloud migration project and coordinates resources to resolve problems and ensure that the migration project is implemented as planned.
- **Migration implementation engineer**: a member of the IT department or an IT professional experienced in cloud migration. The engineer is responsible for migration implementation, including data migration, application migration, system configuration, and service cutover, and ensures data consistency and security as well as good system performance during migration. Migration implementation is a one-off task and is often outsourced to a cloud service provider or cloud implementation service provider.
- **Cloud architect**: an expert from the IT architecture department or one with profound cloud technology background. The cloud architect deploys and

optimizes the cloud architecture and provides technical support and guidance for the implementation team.

- **Cloud infrastructure administrator**: a member of the O&M team from the IT department. The administrator manages the cloud infrastructure and establishes and implements the cloud O&M process to support the management, monitoring, and optimization of the cloud environment.
- **Cloud security expert**: a professional from the information security department or one with security compliance certification. The expert focuses on cloud security and compliance, ensures that the migration complies with industry standards, laws and regulations, and enterprises' management requirements. The expert also formulates and implements security policies to identify potential risks, and provides solutions.
- **Application development engineer**: a member of the application team from a business department. The engineer adapts services to the cloud and modernizes applications to ensure that applications can run efficiently, stably, and securely in the cloud environment.
- **Application test engineer**: a member of the test team. The engineer tests service functions, performance, availability, and scalability, and assists in the mock and real cutover.

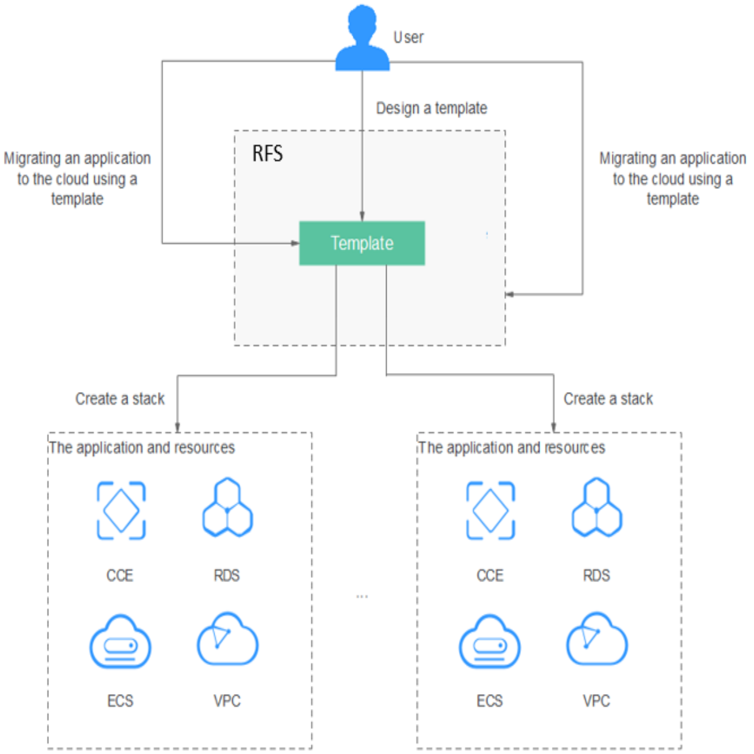
6.3 Infrastructure Deployment

A landing zone is the main infrastructure. It can be deployed in three ways.

- The implementation personnel manually deploy a landing zone on Huawei Cloud. This method is flexible and not restricted by the functions of automation tools, but it is time-consuming.
- Use the **Resource Governance Center (RGC)** for automatic deployment of a landing zone. For details, see the **official documentation**.

  However, the landing zone deployed using RGC may not meet enterprises' requirements, so manual or automatic configuration is required for the landing zone.
- Use Huawei Cloud's Resource Formation Service (RFS) or a third-party automation tool (such as Terraform) for automatic deployment and management of a landing zone. RFS is a new final-state cloud resource orchestration engine that fully supports Terraform (HCL and Provider), the industry's de facto standard for infrastructure as code. It automatically builds cloud resources in batches based on open ecosystem templates that use the HashiCorp Configuration Language (HCL) syntax. With RFS, you can create, manage, and upgrade cloud resources efficiently, securely, and consistently. For details, see the **official documentation**.

Figure 6-3 Orchestration using Huawei Cloud RFS (example)


User
Design a template
Migrating an application to the cloud using a template
RFS
Migrating an application to the cloud using a template
Template
Create a stack
Create a stack
The application and resources
CCE
RDS
ECS
VPC
...
The application and resources
CCE
RDS
ECS
VPC

6.4 Application Cloud Migration

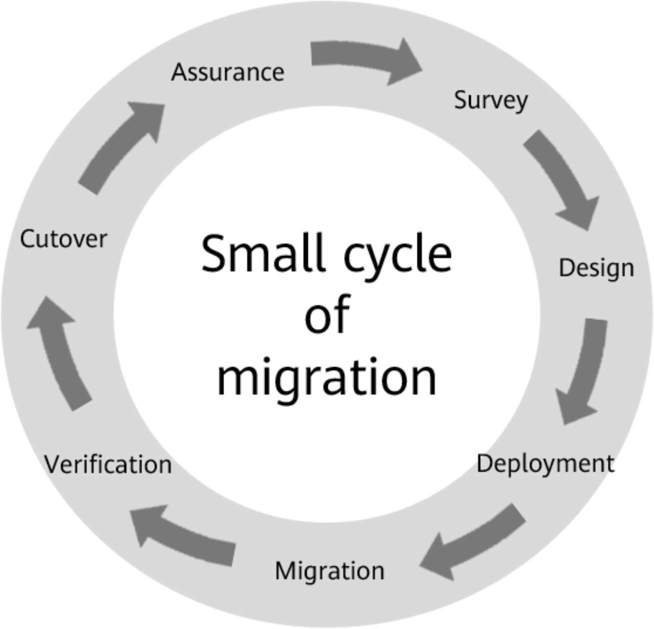
6.4.1 Overview

What Is Application Migration to the Cloud?

Application migration to the cloud is to migrate the access layer, application layer, middleware layer, and data layer of applications to the cloud by using the rehost or replatform strategy without refactor (or re-architect). The data layer includes object storage, block storage, file storage, relational databases, and non-relational databases.

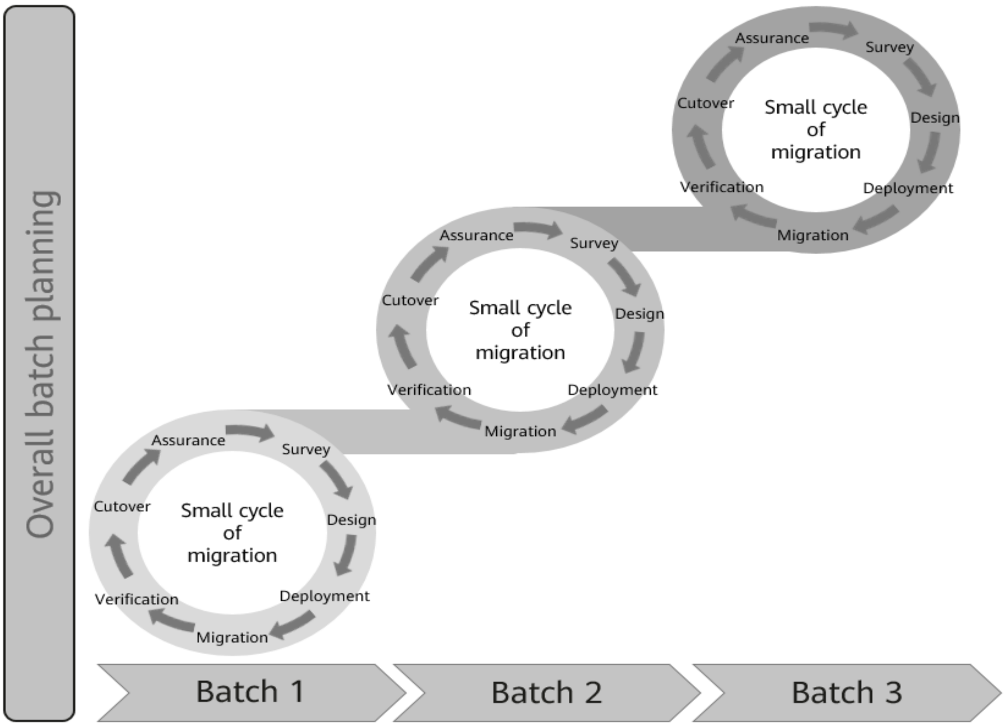
The following figure shows the process of application migration to the cloud.

Figure 6-4 Small cycle of application migration



In the preceding process, each migration batch contains one or more application migration groups. You need to repeat the preceding process to migrate all applications in a migration batch to the cloud.

Figure 6-5 Migration in batches



A small cycle consists of the following phases:

- **Survey**: Conduct a detailed survey on the technical architecture of applications, including their components and versions.
- **Design**: Provide the technical architecture and specifications on the cloud based on the survey result, and offer a detailed migration and cutover solution.
- **Deployment**: Create cloud resources, perform cloud adaptation and reconstruction (if required), and test the target environment.
- **Migration**: Migrate applications and data to the target environment on the cloud.
- **Verification**: Verify data and services.
- **Cutover**: Perform mock cutovers, update the runbook, and perform the real cutover.
- **Assurance**: Perform real-time monitoring and special O&M assurance for a certain period of time after the service cutover.

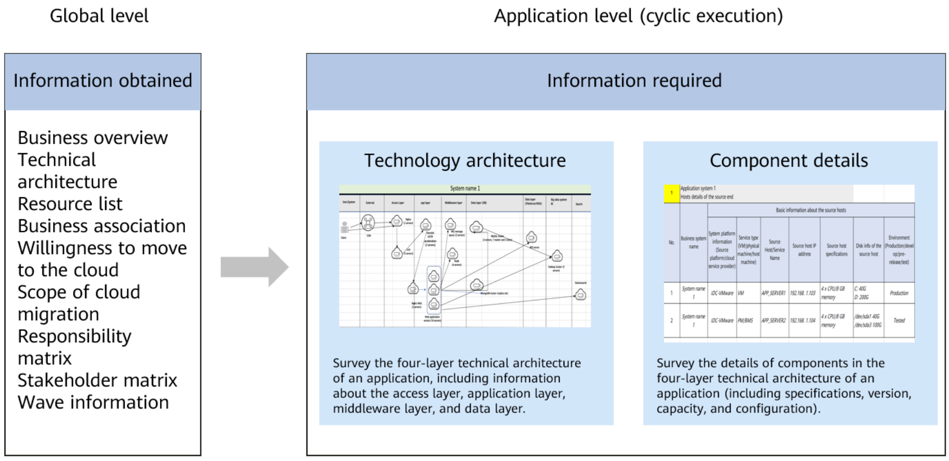
Survey

In the small cycle of application migration, information of an application needs to be surveyed. The survey results of previous phases can be reused.

The obtained information should be detailed enough to guide migration implementation.

For details about the survey method, see **Application System Survey**. The technical architecture and component details of an application need to be obtained.

Figure 6-6 Application survey



Survey the four-layer technical architecture of an application, including information about the access layer, application layer, middleware layer, and data layer.

Survey the details of components in the four-layer technical architecture of an application (including specifications, version, capacity, and configuration).

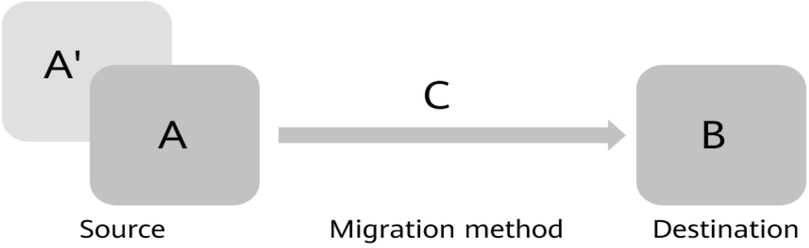
6.4.2 Migration Solution Design

6.4.2.1 Migration Solution Overview

Overview

Application migration to the cloud is like a "house moving." It includes a series of activities centering on the migration source, target, and process.

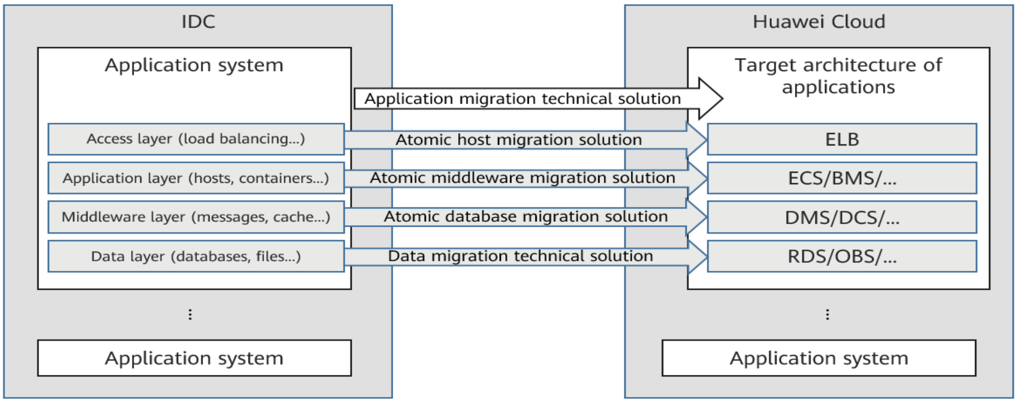
Figure 6-7 Application migration to the cloud



Before designing the migration solution (C), you need to design the target architecture of the application on the cloud (B) by referring to chapter 5.

You need to design a migration solution for each of the four layers of application architecture.

Figure 6-8 Application migration process


IDC
Application system
Access layer (load balancing...)
Application layer (hosts, containers...)
Middleware layer (messages, cache...)
Data layer (databases, files...)
⋮
Application system
Application migration technical solution
Atomic host migration solution
Atomic middleware migration solution
Atomic database migration solution
Data migration technical solution
Huawei Cloud
Target architecture of applications
ELB
ECS/BMS/...
DMS/DCS/...
RDS/OBS/...
⋮
Application system

- The **access layer** has components such as the load balancer and gateway proxy, which are migrated through reconfiguration.
- The **application layer** is usually deployed on servers or containers. Applications deployed on servers are migrated using SMS, and those deployed on containers are re-released through the CI/CD systems of enterprises.
- The **middleware layer** has cache middleware and message middleware. Generally, the cache middleware is migrated using the Huawei Cloud DCS migration tool. The message middleware is not migrated, but directly cut over to Huawei Cloud after messages in the channels are used up by consumer services.
- The **data layer** consists of databases, object storage, and file systems, which are migrated using DRS, OMS, and Rsync, respectively.

Migration Tool Compatibility

The compatibility of each migration tool is subject to the official documentation on their official websites.

- **OSs supported by SMS**
- **Databases supported by DRS**
- **Source object storage supported by OMS**

6.4.2.2 Access Layer Migration Solutions

The access layer is composed of Nginx/OpenResty, hardware or software load balancer, microservice gateway Kong/Zuul, and DNS, providing an entry for external access to an application. Generally, the components are migrated by reconfiguration.

Table 6-1 Access layer migration methods

Technical Component	Function	Migration Method
Nginx/ OpenResty	Forwards traffic.	Solution 1: Use SMS to migrate the server where the Nginx or OpenResty service is running to Huawei Cloud and modify the corresponding forwarding policies. Solution 2: Redeploy Nginx or OpenResty on a Huawei Cloud ECS, copy the configuration file from the source server to the destination server, and modify the forwarding policies in the configuration file.
Load balancer	Forwards Layer 4 or Layer 7 requests.	Copy the load balancing policies from the source to Huawei Cloud ELB.
Kong/Zuul gateway	Microservice gateway	Solution 1: Use SMS to migrate the server where the Kong/Zuul gateway service is running to Huawei Cloud. Solution 2: Redeploy the Kong/Zuul gateway on a Huawei Cloud ECS, copy the source configuration file to the ECS, and modify the forwarding policies.

DNS	Resolves the internal and external domain names of an application to IP addresses.	Solution 1: Replace the DNS used on the source server with Huawei Cloud DNS and reconfigure the mapped IP addresses. Solution 2: Deploy DNS on a Huawei Cloud ECS and reconfigure the mapped IP addresses. Solution 3: Use SMS to migrate the source DNS server to Huawei Cloud and modify the DNS configurations.

6.4.2.3 Application Layer Migration Solutions

The application layer is usually deployed on physical machines, VMs, or containers. Applications are classified into stateful and stateless applications. You need to consider the application deployment mode and application status when designing application-layer migration solutions. The migration solutions vary depending on the application deployment mode and application status.

- **Migrating applications deployed on servers**

  Applications of traditional architectures are usually deployed on physical machines or VMs. You are advised to use the Huawei Cloud SMS to migrate the applications. If SMS cannot be used for migration, you can redeploy the applications on the cloud. For applications that can be migrated during service interruption, you can also migrate them by exporting and importing images. The detailed solution is as follows.

Table 6-2 Migration solutions for applications deployed on servers

Migration Solution	Migration Method	Description	Scenario
(Recommended) Using SMS	Full and incremental migration	1. The downtime is short, and incremental synchronization can be performed continuously. 2. The solution depends on network transmission, and the source OS version must be supported by Huawei Cloud.	Migration of all x86 physical machines or VMs with incremental data
Redeploying an application on Huawei Cloud ECS	N/A	1. The solution does not depend on network transmission. 2. The migration workload is heavier than that using a migration tool.	All

Exporting and importing images	Full migration	1. The solution does not depend on network transmission. 2. The downtime is long, and complete images are created only after the source physical machine or VM is shut down.	A long downtime window (at least four hours)

- **Migrating applications deployed in containers**

  Containerized applications are usually cloud-native applications of microservice architecture, which may be migrated by re-release or image migration.

  Cloud-native application systems of enterprises are usually deployed in containers. Most enterprises have their own development pipeline CI/CD systems. Applications can be migrated to the cloud through container images or re-release of CI/CD pipelines.

Table 6-3 Container migration solutions

1. Migration Solution	1. Description	1. Scenario
1. (Recommended) Re-release through the CI/CD system	1. Simple operation and controllable configuration	1. The source end has a CI/CD pipeline.
1. Container image migration	1. Manual operations and heavy workload	1. All
1. Container migration tool (Velero or e-backup)	1. Simple operation and quick restoration of source configurations	1. All

- **Containerizing applications on the servers and migrating them to the cloud**

  To migrate the traditional applications deployed on servers to containers, you need to reconstruct the traditional applications into container images and deploy them on Kubernetes or Huawei Cloud CCE clusters. This process involves application modernization. For details, see **Application Modernization**.

6.4.2.4 Middleware Layer Migration Solutions

Cache middleware and message middleware are commonly used in enterprises' business. The middleware temporarily stores data, and generally the data does not need to be migrated. The cutover can be performed only after messages in the middleware message queues are consumed. This ensures consistency between the source and destination data. If the middleware cache data is stored persistently in databases, the data also needs to be migrated. This section describes how to design migration solutions for different types of middleware.

- **Redis migration solution**
- **Application scenarios of Redis**

  A Redis instance can be used as a cache or database. The following table lists the migration solutions for Redis in the two scenarios.

Table 6-4 Migration solutions for Redis in different scenarios

Redis Application Scenario	Migration Method
Redis caches data.	During service cutover, you can determine which migration solution to use based on the database performance to prevent the Redis database from breaking down. Solution 1: Do not migrate the Redis cache data. Preload it. Solution 2: Migrate the cache data using the Redis migration solution.
Redis stores data persistently.	Migrate persistent data using the Redis migration solution.

- **Redis migration solution**

Table 6-5 Redis migration solution

Migration Solution	Migration Method	Description	Scenario
(Recommended) DCS migration tool	Full and incremental migration	Short service downtime at the source, simple operations, and online real-time incremental synchronization	Self-hosted Redis is available at the source or Redis instances from other cloud vendors are migrated to Huawei Cloud DCS.
RDB/AOF file backup and restoration	Full migration	Complex offline migration, long service downtime at the source for RDB/AOF file creation, and unavailability of incremental data synchronization	All

- **Message middleware migration solution**
- **Cutover scenarios for the message middleware**

Table 6-6 Cutover scenarios for the message middleware

Applicable Product	Cutover Window	Migration Method
Kafka RabbitMQ RocketMQ ActiveMQ	Sufficient	The cutover time is sufficient. Messages can be used up within the time as estimated and do not need to be migrated.
	Limited	The cutover time is limited. Messages cannot be used up within the time as estimated and need to be migrated based on the message middleware migration solution.

- **Message middleware migration solution**

Table 6-7 Message middleware migration solution

Migration Solution	Migration Method	Description	Scenario
Open-source MirrorMaker 2.0	Full + Incremental	1. Complex deployment and operations 2. Synchronization of offset of a consumption queue	All
Huawei Cloud SmratConnect	Full + Incremental	1. Simple operations on the GUI 2. Synchronization of offset of a consumption queue	Migration of on-premises Kafka or cloud services to Huawei Cloud Kafka

6.4.2.5 Data Layer Migration Solutions

The data layer stores service data persistently and provides data for implementing upper-layer service logic. This layer stores structured and unstructured data. Structured data includes data in various types of databases, such as MySQL and MongoDB databases. Unstructured data includes data in object storage and file storage.

- **Structured data migration solution**

  Structured data provides instant support for services, including data query, computing, analysis, and modification. Services requiring continuity depend on the real-time synchronization capability of database migration tools. When developing the structured data migration solution, you need to consider the service continuity, migration network, and service architecture to strike a balance between data migration complexity, data migration timeliness, and service continuity.

- **MySQL migration solution**

Table 6-8 MySQL migration solution

Migration Solution	Migration Method	Description	Scenario
Huawei Cloud Data Replication Service (DRS) (Recommended)	Full and incremental migration	Simple configuration, one-click migration, and real-time synchronization of incremental data	1. Migration from on-premises or cloud service MySQL instances to Huawei Cloud MySQL instances 2. Migration from on-premises or cloud service MySQL instances to Huawei Cloud self-managed MySQL instances (migration to cloud service and then to self-managed MySQL instances)
Export and import using mysqldump	Full migration	The solution does not depend on networks, and the operations are complex. Only full migration is supported. Incremental data synchronization is not supported.	A long downtime window
Primary/ secondary replication	Full and incremental migration	Complex operations	The source and destination databases are both self-managed MySQL databases, but their versions are incompatible with each other, so Huawei Cloud DRS cannot be used for the migration.

- **SQL Server migration solution**

Table 6-9 MySQL migration solution

Migration Solution	Migration Method	Description	Scenario
Huawei Cloud DRS (Recommended)	Full and incremental backup import	Simple operations on a GUI	Migration from on-premises SQL Server instances or cloud service SQL Server instances to Huawei Cloud SQL Server instances
Backup and restoration	Full and incremental backup import	Complex operations	Migration from on-premises SQL Server instances or cloud service SQL Server instances to Huawei Cloud RDS for SQL Server

- **PostgreSQL migration solution**

Table 6-10 PostgreSQL migration solution

Migration Solution	Migration Method	Description	Scenario
Huawei Cloud DRS (Recommended)	Full and incremental migration	Simple configuration, one-click migration, and real-time synchronization of incremental data	Migration from on-premises PostgreSQL instances or cloud service PostgreSQL instances to Huawei Cloud self-managed PostgreSQL instances or RDS for PostgreSQL
Export and import using pg_dump	Full migration	Advantages: This solution does not depend on the network. Disadvantages: The operations are complex. Only full migration is supported.	Huawei Cloud DRS is not applicable.

- **MongoDB migration solution**

Table 6-11 MongoDB migration solution

Migration Solution	Migration Method	Description	Scenario
Huawei Cloud DRS (Recommended)	Full and incremental migration	Simple configuration, one-click migration, and real-time synchronization of incremental data	Migration from on-premises MongoDB instances or cloud service MongoDB instances to Huawei Cloud self-managed MongoDB instances or cloud service MongoDB instances
Export and import	Full migration	The solution does not depend on networks, and the operations are complex. Only full migration is supported. Incremental data synchronization is not supported.	Huawei Cloud DRS is not applicable.

- **Unstructured data migration solution**

  Unstructured data is in irregular or incomplete data structures and does not have a predefined data model. It cannot be represented in a two-dimensional logical table of a database. Unstructured data includes office documents, texts, images, XML, HTML, reports, audios, and videos. Enterprises' unstructured data is mainly stored in file storage and object storage.

- **NAS migration solution**

Table 6-12 NAS migration solution

Migration Solution	Migration Method	Description	Scenario
Huawei Cloud Cloud Data Migration (CDM) (recommended for migrating massive amounts of data)	Full and incremental migration	Simple operations and incremental migration, more suitable for massive amounts of data	Migration of data from object storage, network file storage, and big data storage to Huawei Cloud OBS, SFS, and big data storage.

Migration Solution	Migration Method	Description	Scenario
Open-source tools Rclone and Rsync	Full migration	Rclone is complex. Rsync offers low migration efficiency.	Huawei Cloud FMS and CDM are not applicable.

- **Object storage migration solution**

Table 6-13 Object storage migration

Migration Solution	Migration Method	Description	Scenario
Huawei Cloud OMS (recommended)	Full and incremental migration	Simple operations, high concurrency, data verification, and visualized reports	Migration of data from the source object storage to Huawei Cloud OBS
Open-source tools Rclone and Rsync	Full migration	Rclone is complex. Rsync offers low migration efficiency.	Huawei Cloud OMS is not applicable.

6.4.3 Cutover Solution Design

6.4.3.1 How Do I Choose a Cutover Solution (With or Without Downtime)?

Service cutover is a crucial step in the entire cloud migration process. Unexpected situations that occur during a cutover, if not handled promptly, can directly impact business continuity. Different services have different requirements on service continuity. Some services cannot tolerate even minimal downtime, as any interruption leads to significant losses. Some can be stopped during cutover. For example, OA systems may be stopped and upgraded during non-working nighttime hours without major disruptions. For some others, only browsing services must remain available during cutover to ensure a smooth user experience, while write operations can be temporarily halted. When designing a cutover solution, it is essential to assess workloads and their tolerance for downtime.

Three cutover solutions are available: cutover with downtime, cutover with writes stopped but reads continued, and zero-downtime cutover. The table below compares different cutover solutions.

Table 6-14 Comparison of cutover solutions

Cutover Solution	Description	Data Consistency Risk	Service Change Cost	Downtime (Hours)	
				Read	Write
Cutover with downtime	Commonly used solution that can ensure data consistency	Low	Low	0.5~3.5	
Cutover with writes stopped but reads continued	Less used solution, which requires service changes to implement cutover with writes stopped but reads continued. This solution can ensure data consistency.	Low	Medium	Continued	0.5~3.5
Zero-downtime cutover	Rarely used solution, which requires service changes to implement dual-write and bidirectional data synchronization. This solution requires service changes to ensure data consistency, which can be fairly complex.	High	High	Continued	Continued

Each solution has its advantages and disadvantages. There is no perfect option that offers low risk, low cost, and minimal downtime. Businesses need to select their cutover solution based on service scenarios, tolerance of downtime, and cost-effectiveness. Consider the following when choosing your cutover solution:

1. **By industry**

Different industries have their own standards and requirements. For example, retail e-commerce platforms typically have significantly fewer transactions or even zero transactions during early morning hours. For them, performing a cutover with services completely stopped during early-morning hours is acceptable. However, for some other industries, such as ride hailing apps that must be available 24/7, stopping services is simply not an option. For them, a zero-downtime cutover is the only viable option.

1. **By service importance**

Mission-critical services, such as online games and financial services, must be available 24/7. For them, a zero-downtime cutover is the only viable option. Other less critical service, such as OA and operations systems, can tolerate brief interruptions. For them, you may choose cutover with downtime.

1. **By project period**

If a full-service stoppage is not an option, consider using a zero-downtime cutover solution, if there is sufficient time. In this case, some service changes may be required. For example, an active-active design may be needed. Conversely, if the cloud migration project is on a tight schedule and the customer cannot devote enough manpower to the project, choose cutover with downtime, as it requires minimal changes to the migrated service systems and minimal manpower from the customer side.

1. **By cost-benefit ratio**

A zero-downtime cutover often requires significant R&D efforts to adapt service systems, while a cutover with downtime minimizes such investment. Businesses must carefully weigh the costs against the expected benefits.

6.4.3.2 Cutover with Downtime

Service Downtime Evaluation

According to statistics from Huawei Cloud's past projects, the downtime for most applications during a cutover is 0.5 to 3.5 hours, as analyzed in the following table.

Table 6-15 Service downtime breakdown

Service downtime evaluation for cloud migration								
Total Downtime (36–211 min)								
Stopping the source system (12–75 min)			**Incremental data synchronization and verification (6–40 min)**			**Starting the destination system (18 to 96 minutes)**		
Stopping the source system	**Duration (min)**	**How to reduce the duration**	**Incremental data synchronization**	**Duration (min)**	**How to reduce the duration**	**Starting the destination system**	**Duration (min)**	**How to reduce the duration**
Block inbound traffic at the access layer (gateway/ELB)	1~5	1. Call APIs or perform batch operations using scripts to reduce operation time.	Last incremental synchronization	1~10	1. Perform the cutover during off-peak hours to reduce incremental data.	Allow writes to databases	1	1. Run scripts to allow write operations.

Service downtime evaluation for cloud migration								
Stop applications	1~30	1. Stop non-critical services in advance to reduce the workload. 2. Shut down services in waves to reduce operation time. 3. Use a unified log platform to reduce the time needed to check application logs.	Data verification	5~30	1. Enable dynamic verification for tools to reduce the verification time.	Start applications	1~30	1. Use a unified O&M platform to stop and start services in batches. 2. Use a unified log platform to reduce the time needed to check application logs.
Stop messaging middleware (after all messages are consumed)	5~30	1. Stop non-critical services in advance to reduce the number of messages. 2. Use a unified monitoring platform to reduce inspection time.	-	-	-	Test applications	15~60	1. Use automated test cases. 2. Execute only critical test cases.

Service downtime evaluation for cloud migration								
Stop the data layer (writes blocked for check)	5~10	1. Use a unified monitoring platform to reduce the check time.	-	-	-	Cut over inbound traffic	1~5	1. Call APIs or perform batch operations using scripts to reduce operation time.

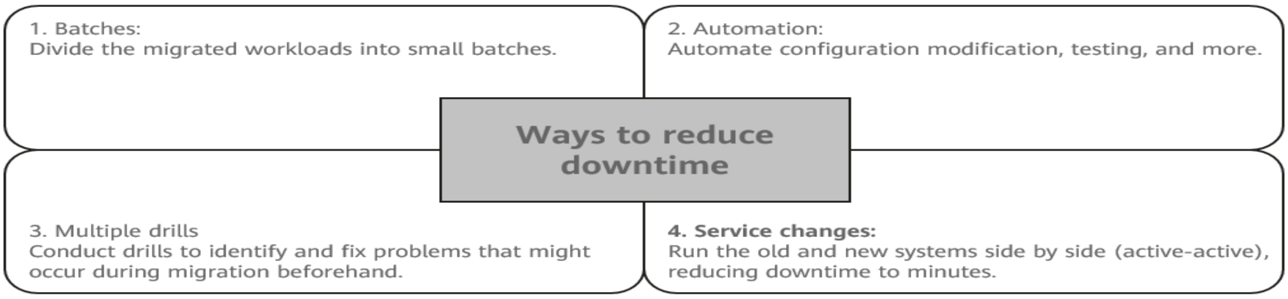
Exceptions of Service Duration

1. **Downtime < 30 min**: If the total service downtime is required to be less than 30 minutes, you can migrate data in waves (dividing data or workloads into different waves) or perform the cutover layer by layer. For example, cut over the application layer and then the data layer. Additionally, if scripts and tools can be used to automate most of the operations, the total service downtime may be reduced to less than 30 minutes.
2. **Downtime > 3.5 hours**: When there are large amounts of data and workloads to migrate; batching is impossible due to complex relationships between different components; or the runbook cutover is complex and not sufficiently automated, the total service downtime may be longer than 3.5 hours, or even up to 8–10 hours. For example, with a large organization that has over 600 microservices, over 100 middleware services, over 80 databases, over 1000 batch processing tasks, and over 4000 test cases, the total downtime is around 8 hours.

Ways to Reduce Downtime (in Minutes)

The service downtime depends on multiple factors. Downtime can be reduced during cutover through batch operations, automation, multiple drills, and service adaptation and changes. The figure below shows four ways to reduce downtime during cutover.

Figure 6-9 Ways to reduce downtime during cutover



1. Batches:
Divide the migrated workloads into small batches.

2. Automation:
Automate configuration modification, testing, and more.

Ways to reduce downtime

3. Multiple drills
Conduct drills to identify and fix problems that might occur during migration beforehand.

4. Service changes:
Run the old and new systems side by side (active-active), reducing downtime to minutes.

Four Ways of Cutover with Downtime

Table 6-16 Four ways of cutover with downtime

Cutover Method	Scenarios	Downtime	Number of Service Interruptions	Impact Scope
One-off cutover (stopping the application and data layers all at once)	The application system can tolerate long downtime; there are complex relationships between applications and data and between different applications.	Long	1	All services
Cutover in 3–4 waves: Perform gray cutover of the application layer (1%, 30%…100%) before stopping services. Then, cut over the entire data layer. Then switch internal and external domain names to the destination system.	The downtime is short. Cross-cloud access is allowed in a short period of time, and cross-cloud bandwidth and latency meet the service requirements.	Medium	1	All services
Cutover in 5–10 waves: Perform gray cutover of the application layer (1%, 30%, …100%). Then cut over the data layer in waves (for example, cache + database in wave 1, cache + database in wave 2, and middleware message queues in wave 3).	The downtime is short. Cross-cloud access is allowed in a short period of time, cross-cloud bandwidth and latency meet the service requirements, and the impact of the cutover on internal and external systems is controllable.	Short	Multiple times	Some services
Cutover by service domain	Service domains are relatively independent, with simple relationships between them. They can be migrated to the cloud separately.	Short	Multiple times	Some services

One-Off Cutover (Stopping the Application and Data Layers All at Once)

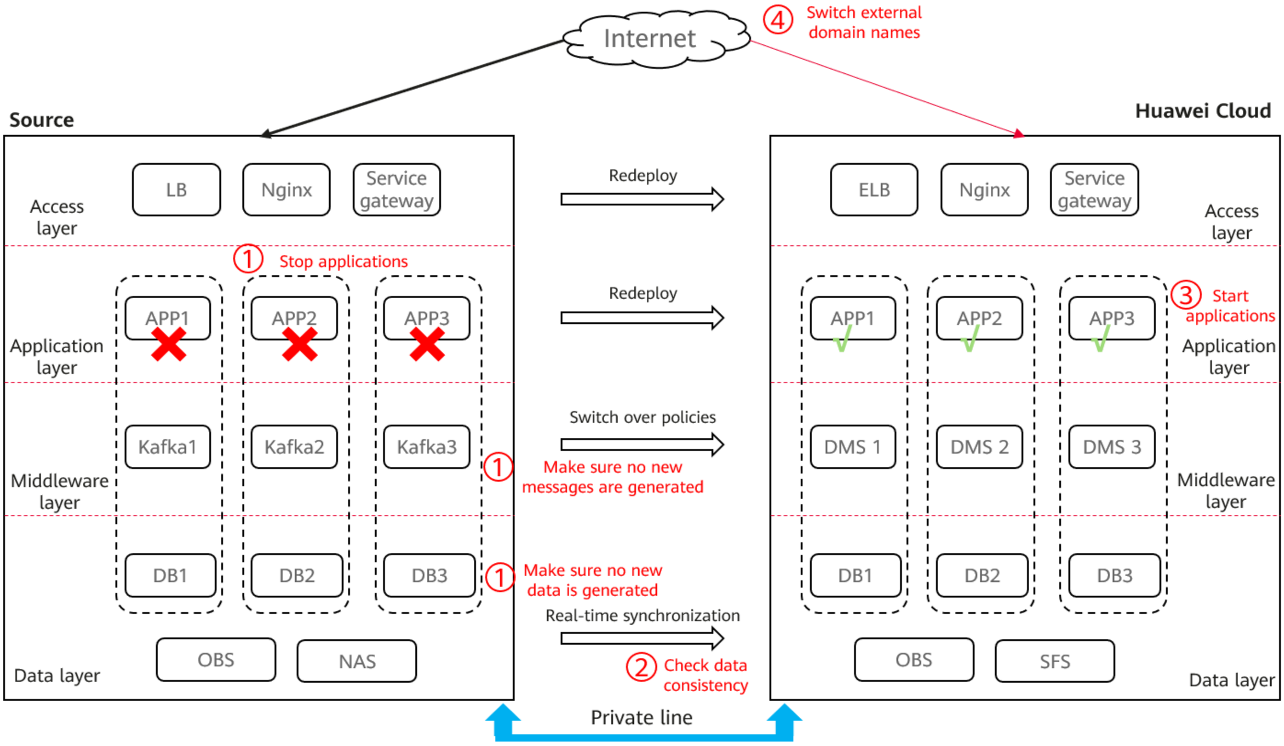
Preparations

1. The applications and data of the source system have been migrated to Huawei Cloud.
2. The applications and data have been verified to function properly in the destination system on Huawei Cloud.

Service cutover:

1. Stop the applications and batch processing tasks at the source end to prevent new data from being generated. Check that no new messages are generated in message queues and no new data is generated at the data layer.
2. Complete incremental data synchronization between the data layers at the source end and Huawei Cloud, compare the data for consistency, and then disconnect the data synchronization link.
3. Change the internal domain names used at the application and data layers on Huawei Cloud, and restart applications and services on Huawei Cloud.
4. Switch the addresses corresponding to external domain names from the source system to Huawei Cloud to redirect access traffic to Huawei Cloud.

Figure 6-10 One-off cutover



Gray Cutover of the Application Layer and One-Off Cutover of the Data Layer

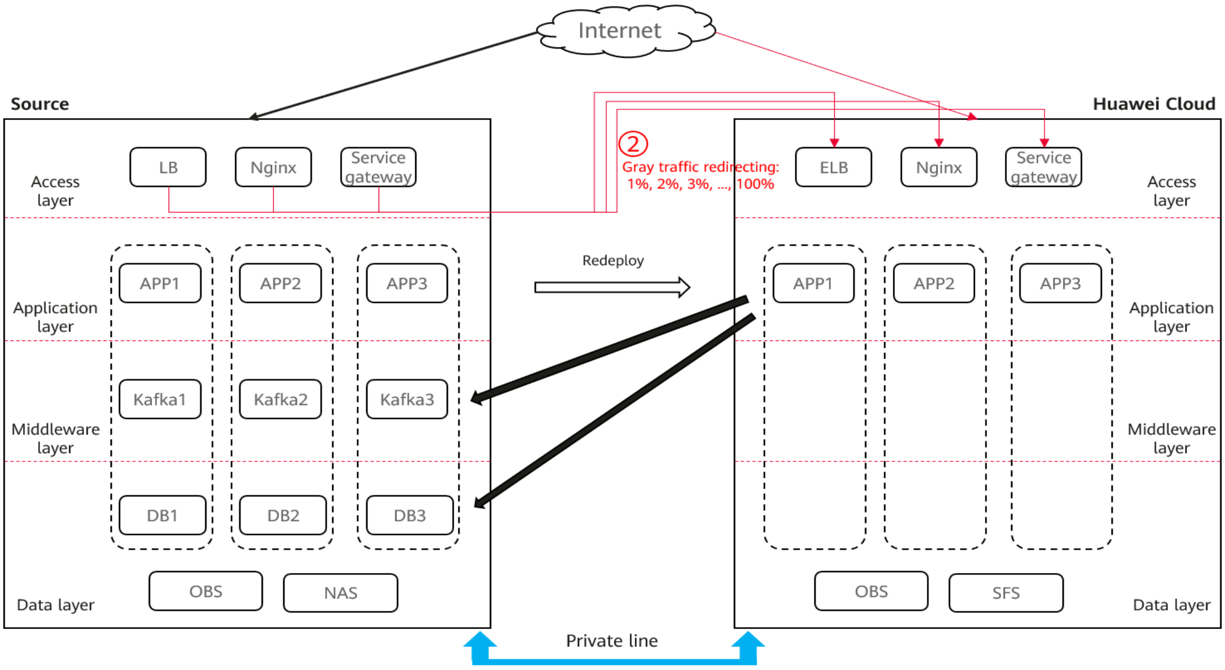
Before gray cutover of the application layer, ensure that:

1. The applications of the source system have been migrated or deployed to Huawei Cloud.

2. The applications on Huawei Cloud can access the databases at the source end. Functionality and performance have been verified.

After the preparations are complete, redirect traffic from the source-end access layer to Huawei Cloud gradually (1%~30%, ...100%).

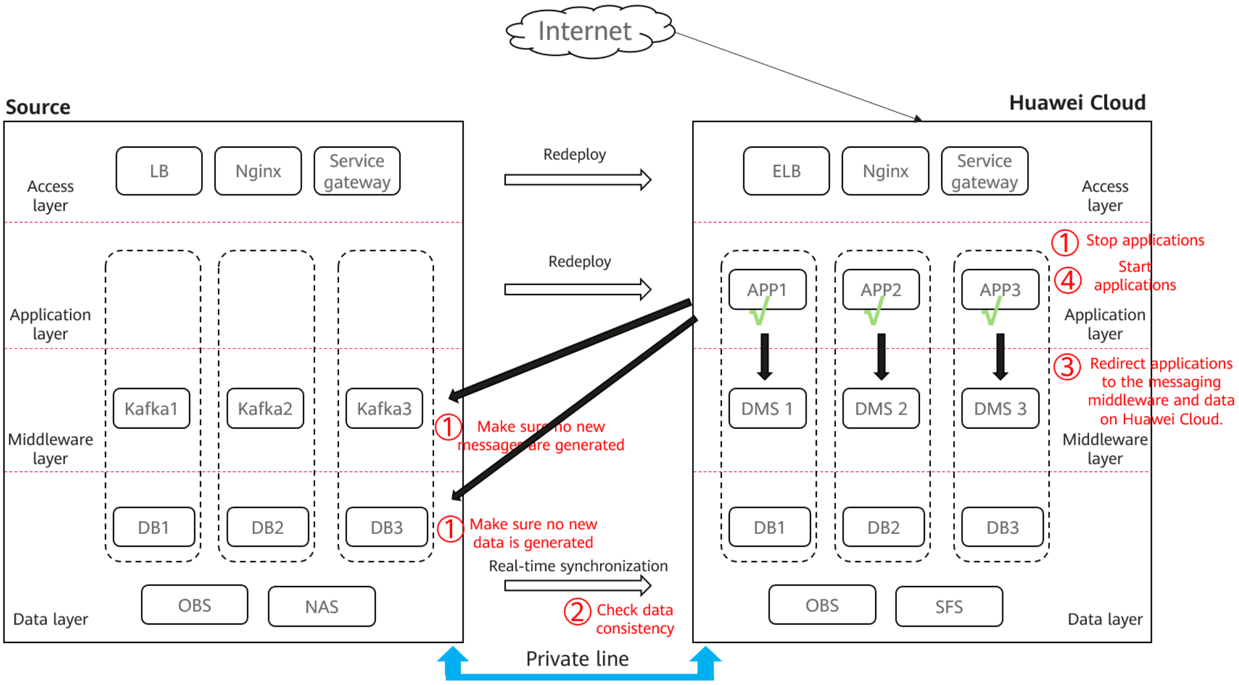
Figure 6-11 Gray cutover of the application layer



Then, cut over the data layer one-off by performing the following steps:

1. Stop applications and batch processing tasks in both the source and destination systems to prevent new data from being written in. This is when services become unavailable.
2. After all messages in the middleware message queue are consumed, synchronize the incremental data to Huawei Cloud. Verify data consistency between the source and destination systems.
3. Modify configuration to redirect the applications on Huawei Cloud to the data that is also on Huawei Cloud, and start the applications.
4. Switch the addresses corresponding to external domain names from the source system to Huawei Cloud to redirect access traffic to Huawei Cloud.

Figure 6-12 One-off cutover of the data layer



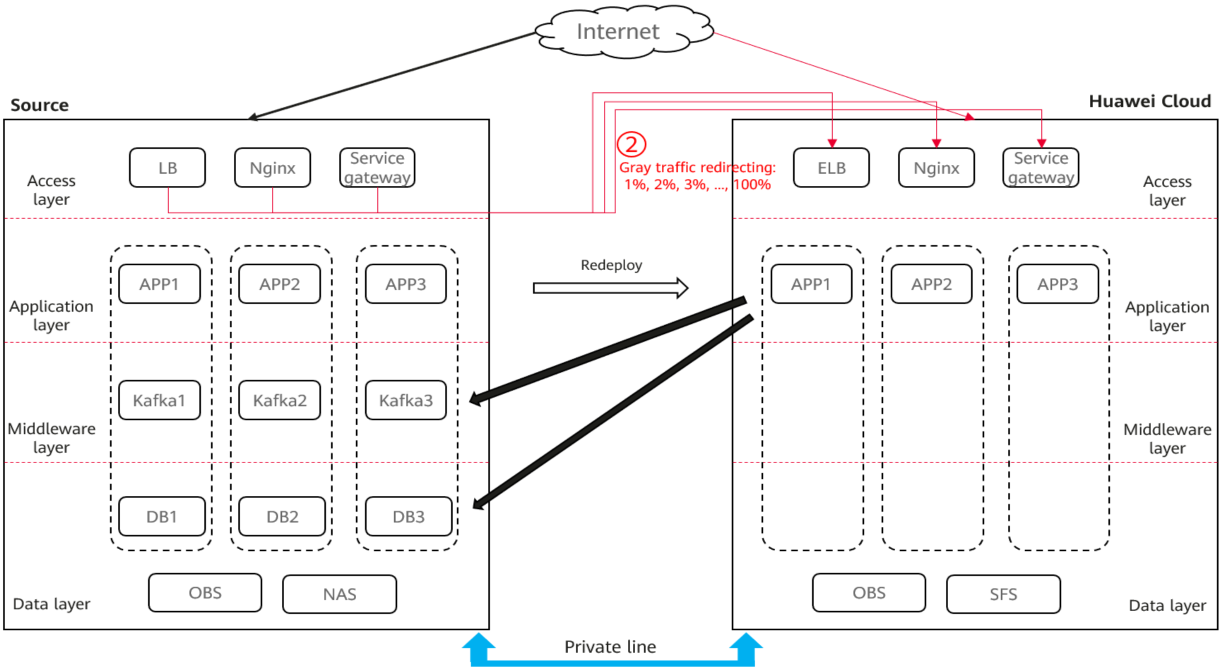
Gray Cutover of the Application Layer and Wave-by-Wave Cutover of the Data Layer

Before gray cutover of the application layer, ensure that:

1. The applications of the source system have been migrated or deployed to Huawei Cloud.
2. The applications on Huawei Cloud can access the databases at the source end. Functionality and performance have been verified.

After the preparations are complete, redirect traffic from the source-end access layer to Huawei Cloud gradually (1%~30%, ...100%).

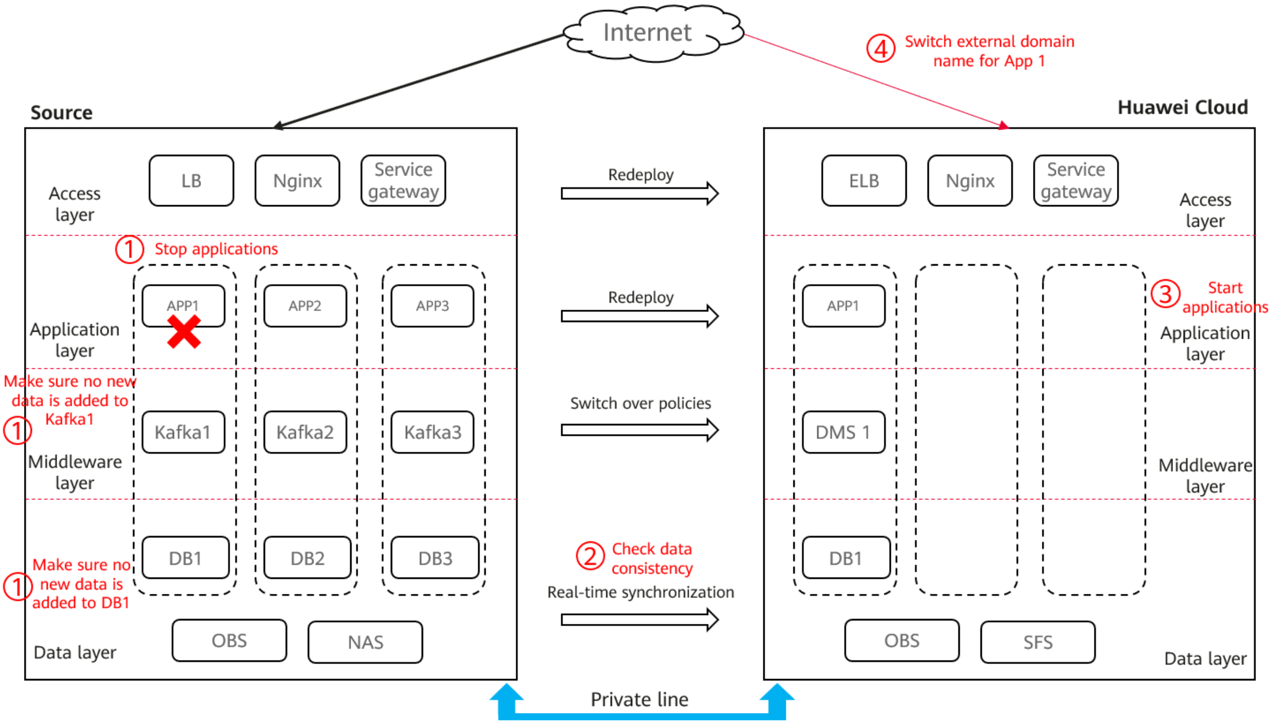
Figure 6-13 Gray cutover of the application layer



Then, cut over the data layer in waves by performing the following steps:

1. Stop the applications and batch processing tasks related to each wave of the data to be migrated to freeze the data. (All messages in message queues have been consumed, and no new data is written into databases.)
2. Verify data consistency and then cut over the data layer.
3. Modify related configurations and start the applications and batch processing tasks related each wave of data. Verify performance and functionality.
4. Switch the addresses corresponding to external domain names from the source system to Huawei Cloud to redirect access traffic to Huawei Cloud for each application.

Figure 6-14 Cutover of the data layer in waves



This solution has the following advantages:

1. Application cutover is simple. When redirecting traffic, you only need to adjust traffic weights on gateways and load balancers.
2. Traffic redirecting on gateways and load balancers occurs gradually. In case of a problem, the traffic can be switched back to the source system.
3. During application switchover, applications are still available. During data switchover, writing was disabled only for some of the applications, but the data can still be read.
4. Data nodes are switched over in waves, and the operations are simple. While some nodes are being switched over, other nodes can still provide services.

   This solution has the following disadvantages:
5. Significant time and manpower are required to analyze and map the dependencies between the application and data layers.
6. After cross-cloud migration, it is difficult to roll back the data layer.
7. Cross-cloud migration requires dedicated connections for accessing and transmitting large amounts of data between clouds, which may lead to increased latency.

Cutover by Service Domain

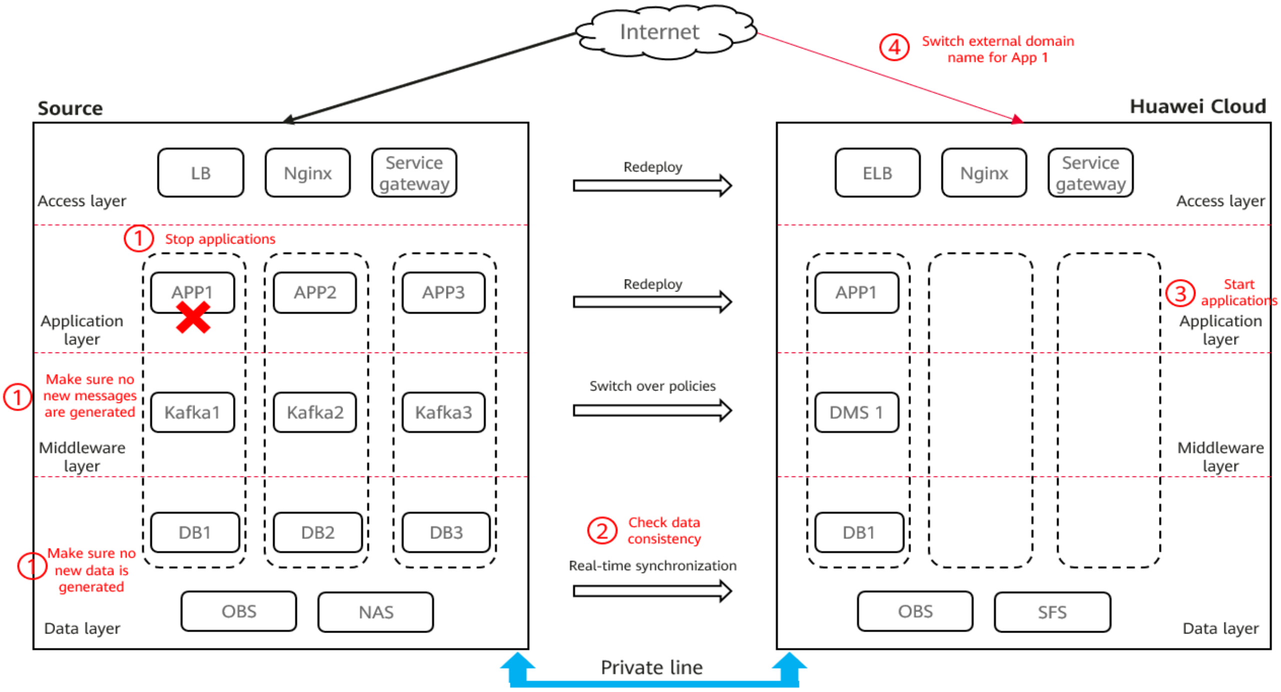
Preparations

1. The applications and data of the service domain to be cut over have been deployed and migrated to Huawei Cloud.
2. The applications and data have been verified for functionality and performance.

Per-service domain cutover:

1. Stop applications and batch processing tasks for one service domain at the source end to prevent new data from being generated. Also, stop applications and batch processing tasks that share middleware and databases with this service domain to prevent new data from being generated.
2. Check that no new data is generated. Verify data consistency, and then disconnect the data synchronization link.
3. Modify configurations on Huawei Cloud to start applications and batch processing tasks for the migrated service domain. Then, start applications and batch processing tasks that share middleware and databases with this service domain at the source end.
4. Cut over DNS domain names to Huawei Cloud and verify functionality and performance.

Figure 6-15 Cutover by service domain



6.4.3.3 Cutover with Writes Stopped but Reads Continued

For some applications, browsing services may remain available during cutover to ensure a smooth user experience, while write operations are temporarily halted to ensure data consistency. This is useful, for example, for some e-commerce apps. During such a cutover, customers can still browse the website but cannot place orders. If they do play an order, a friendly reminder is displayed telling them the system is upgrading and will recover at 4:00 a.m., for instance.

1. **Comparing different ways of cutover with writes stopped but reads continued**

Four options are available:

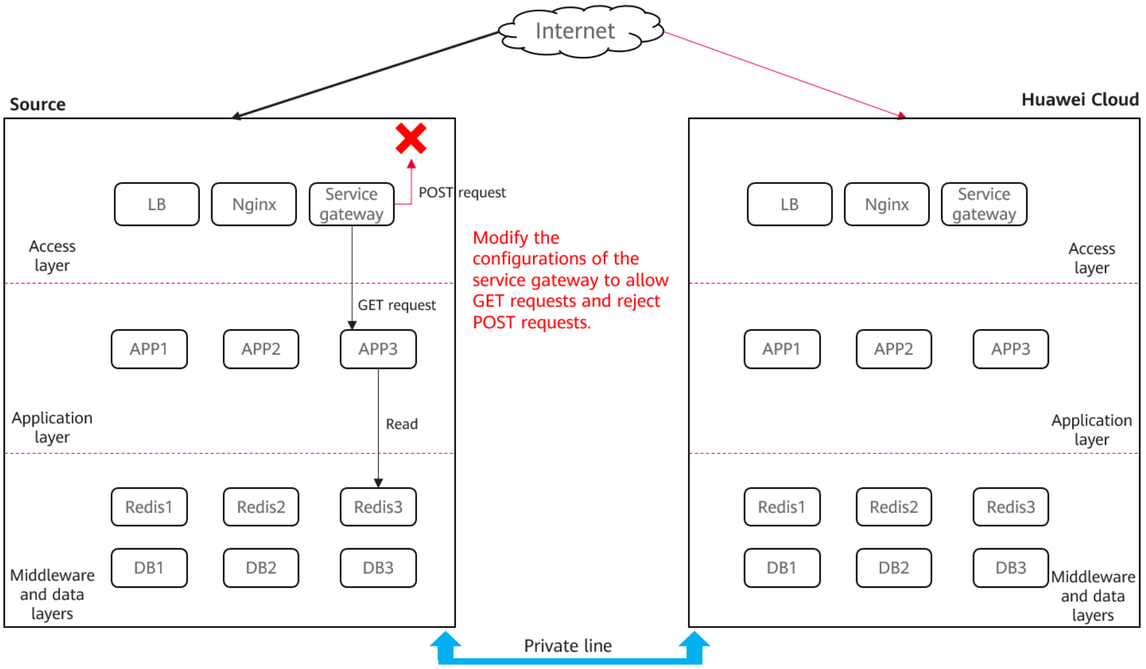
Table 6-17 Four ways of cutting over a system with writes stopped but reads continued

Solution	Description	Scenarios	Operation Complexity	Changes Required
Gateway blocking	The service gateway blocks write requests but allow read requests at the access layer.	A unified gateway blocks write requests and returns friendly reminders for write requests.	Easy to use	No changes required
Blocking writes while allowing read requests	Shut down the write service or the corresponding interface, but keep the read service or the corresponding interface alive.	Read/write splitting is enabled for applications, so that each service allows only read or write operations.	Easy to use	No changes required
Enable read/write splitting for applications first	Modify the application code to enable read/write splitting.	Read/write splitting is not enabled for applications.	Complex	Heavy workload
Revoke write permissions at the middleware and data layers	Revoke the write permissions for related accounts at the middleware and data layers.	When write permissions are revoked, the application system will report errors. Minor changes are needed to handle the error.	Easy to use	Minor changes required

- **Gateway blocking**

  The service gateway (such as Gatekeeper, Zuul, and Kong) blocks write requests while allowing read requests. For example, the Gatekeeper gateway can block POST requests and allow only GET requests to pass through. This can be achieved by setting packet handling rules on Gatekeeper.

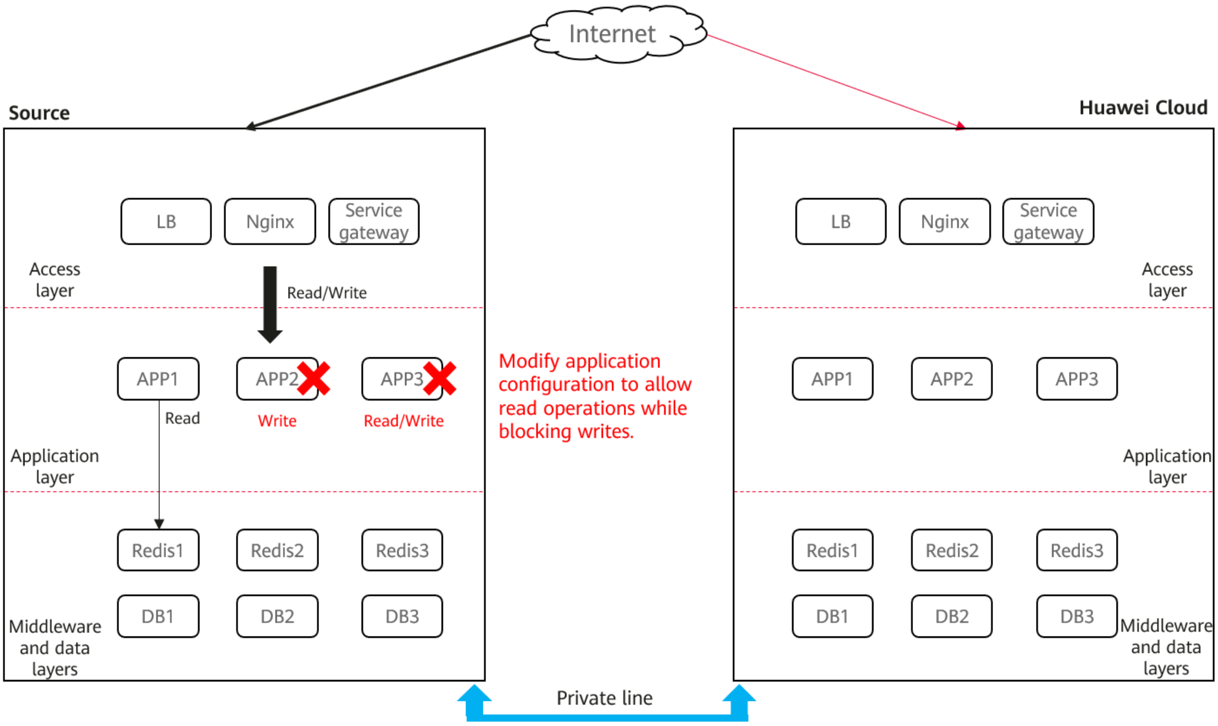
Figure 6-16 Gateway blocking



- **Disabling the write service**

  If read/write splitting is enabled, shut down the write service or the corresponding interface, but keep the read service or the corresponding interface alive. This makes applications read-only.

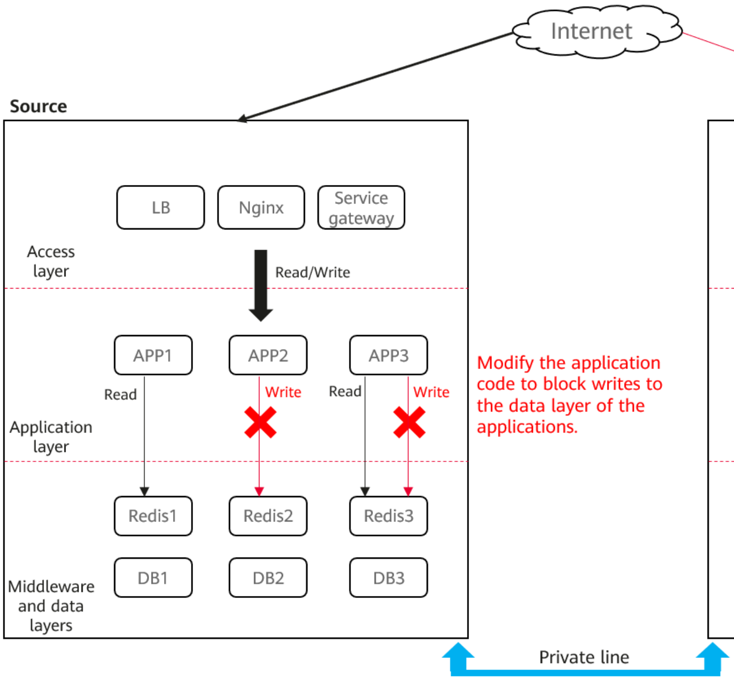
Figure 6-17 Shutting down the write service



- **App refactoring**

  Modify the application code to enable read/write splitting. Then use the solution above to enable read-only service.

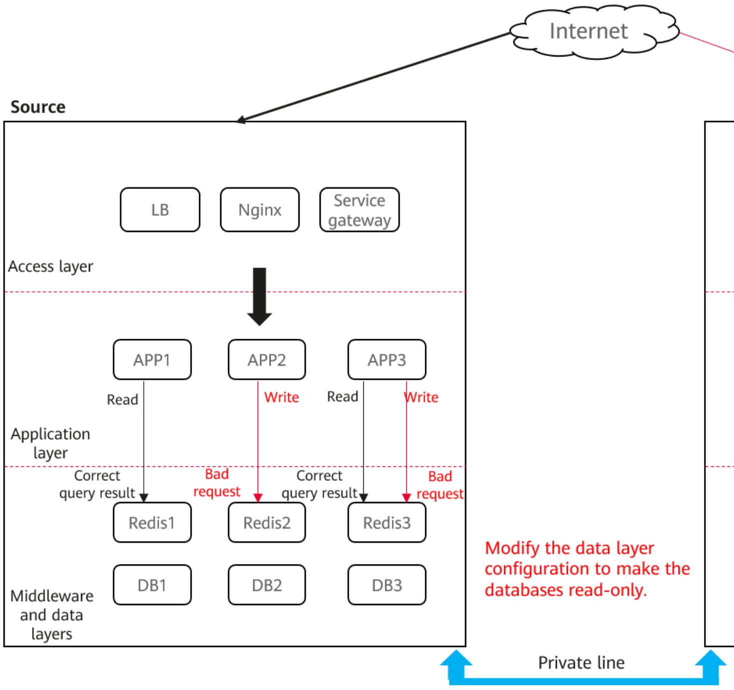
Figure 6-18 Application refactoring



- **Configure read-only at the middleware and data layers**

  Revoke the write permissions for related accounts at the middleware and data layers.

Figure 6-19 Configuring read-only at the middleware and data layers



6.4.3.4 Zero-Downtime Cutover

- **Zero-downtime cutover of the application layer**

If only the cutover of the application layer is involved, use the application gray cutover solution described in **Cutover with Downtime**.

- **Zero-downtime cutover of applications and data one-off**

  Preparations:

  a. Applications and data have been migrated to Huawei Cloud.

  b. The applications and data have been verified to function properly on Huawei Cloud.

  Cutover procedure:

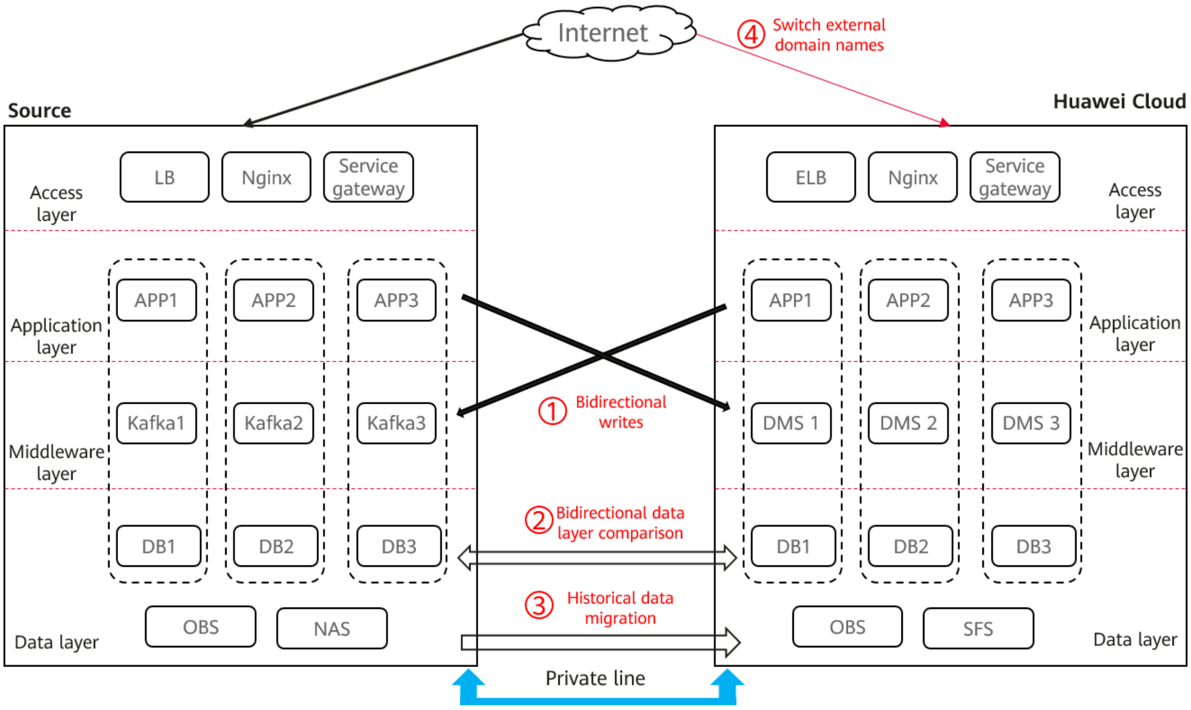
  a. Modify configurations on the source and destination ends to point the applications at the source end to messaging middleware and data at both the source and destination ends. Then, do the same to the applications at the destination end. Note: Data consistency in the case of a dual-write is guaranteed by applications.

  b. Compare data consistency between the source and destination ends in real time.

  c. Migrate historical data to Huawei Cloud.

  d. Change the addresses corresponding to external domain names to redirect access traffic from the source end to Huawei Cloud.

Figure 6-20 Zero-downtime cutover



6.4.4 Designing a Runbook

6.4.4.1 Runbook Design Principles

A runbook is an important document for cloud migration. It outlines the process and steps for service cutover, guiding team collaboration. A runbook has two parts: a checklist and an operation procedure. This section explains how to design the cutover runbook in several ways.

Runbook design principles:

- A runbook is created for a specific cutover.
- A runbook should provide details of the cutover procedure, operators, confirmers, and estimated start time, end time, and execution time.
- Each step should be clear for one operator to do and one confirmer to check. This avoids having multiple people confirm the same step.
- A runbook should list specific commands. Use scripts or tools for the runbook so operators can run it directly without changes, reducing the risk of errors.
- The runbook can have both parallel and serial operations. Clearly mark the order of these operations to avoid errors that could affect the execution time and results.
- The runbook procedure might fail. Predict possible issues and decide in advance whether to roll back or continue the cutover. This prevents delays in decision-making.

Design principles of rollback conditions:

- Each cutover phase has a set completion time. If it is not done by then, decide if you need to roll back.
- If the core table data comparison shows inconsistencies, decide if you need to roll back.
- If the core P0 test fails, decide if you need to roll back.
- If the performance does not meet the expectation, decide if you need to roll back.

6.4.4.2 Runbook Role Design

This table shows runbook roles.

Table 6-18 Runbook roles and responsibilities

Role	Responsibility
Operator	Performs operations and handles exceptions according to the runbook.
Confirmer	Checks if the procedure is done and reports the results to the facilitator. If issues arise, regularly updates the facilitator on progress according to the runbook.
Facilitator	Facilitates the runbook execution and notifies the exception handling progress.
Recorder	Updates the runbook execution status and records the completion time for each step and exception handling time.
Decision-making team	Makes decisions on key items under management of the team leader.

Role	Responsibility
Meeting affairs team	Maintains the site order and provides meeting affairs assurance during cutover.

NOTE

- Each step involves an operator and a confirmer. If there are multiple confirmers, they can update the progress online in a shared document.
- The facilitator, usually one or two people, leads the cutover. For large-scale cutovers with many participants and longer operations, two or three facilitators can work as backups. Facilitators must know the entire runbook, including the order of steps and how multiple steps run together.

6.4.4.3 Runbook Checklist Design

The runbook checklist lists the tasks to prepare and do before the cutover. It usually covers these aspects:

- **Personnel and site preparations**
  a. Check and notify internal personnel and third-party personnel participating in the cutover, and check the time.
  b. Create a cutover assurance group for notifications during the cutover.
  c. Determine the cutover date, sign-in time, and operation start time.
  d. Prepare the cutover site with a meeting room and necessary items like computers, power strips, and projection screens.
  e. Make ready the required tools, terminals, and login platforms. The related personnel should check whether the tools and platforms are available, including whether the bastion host account permissions are normal and whether test terminals (test mobile phones and computers) are available.
  f. Notify related personnel to release an announcement on the official website, and remind third parties to release the announcement on their official websites if required.
- **Application list check and script update**
  a. Software version development and iterative release of enterprises are still in progress during cloud migration. Therefore, you need to check the detailed environment lists before the cutover, including the application list and job list.
  b. After the lists are checked, notify the relevant people that the version is frozen to prevent differences between the test and live environments.
  c. Update the cutover scripts in the runbook based on the latest application list and job list.
- **Environment check**

  Before the cutover, check that the source end, destination end, and migration task are working properly and the script is ready.

  a. Source end check: Check if there is an alarm for exceeding the bandwidth limit on the Direct Connect synchronization bandwidth. If so, assess if you need to expand the bandwidth. Additionally, monitor alarms for applications and databases at the source to ensure they are clear and the status is normal.

b. Destination end check: Notify the cloud vendor to perform routine resource status inspection and high availability check. After the switching to the destination, the production environment is used. Make sure alarms, monitoring, logs, and security policies are set up and checked.

c. Forward migration task status check: The migration task is usually created before the system cutover. The task must be in the incremental synchronization state with no errors or alarms.

d. Reverse migration task status check: Before the cutover, check the status of the reverse migration task to ensure there are no errors or alarms. The data layer or middleware should manage rollback paths.

e. Parameter consistency check: Check the consistency between source and destination parameters, such as the character set consistency of databases and the consistency of database usernames.

6.4.4.4 Runbook Procedure Design

Each step in the runbook includes a clear procedure, operation command or script, serial or parallel marking, operator, confirmer, estimated start and end times, and expected duration. The runbook's steps depend on the cutover method. There are two types: service interruption and no service interruption. No service interruption requires major changes to the application structure. Because of this, service interruption is more common. The following uses service suspension as an example to list key points for designing a runbook.

- **Designing forward cutover procedure**

  Refine the forward cutover steps in the document based on the cutover solution. Consider these aspects:

  a. Announce the service interruption ahead of time to consider user experience.

  b. Consider the system's availability mechanism during service interruptions. Some systems automatically restart applications upon detecting they have stopped. Therefore, you should first disable the availability mechanism to prevent the risk of applications failing to stop.

  c. Consider data consistency during database cutover. Keep the data at the source end static and disconnect the incremental synchronization task to ensure consistency. Plan the data consistency check carefully. Decide if you need to compare row counts or content based on the table's importance and the cutover time.

  d. To make the source end data static, stop applications and consider message consumption in batch processing tasks and message queues.

  e. Applications and scheduled tasks usually start and stop in order. To prevent service issues, organize the startup and stop sequence of these applications and tasks.

  f. Public network DNS caches domain names. After switching systems, some traffic may still go to the source end. The runbook should address this DNS cache issue. Keep the forwarding path from the source to the destination end for a while. Watch the traffic, then cut off the forwarding path.

- **Designing rollback procedure**

  If a serious problem happens during the cutover and cannot be fixed quickly, you must roll back to restore the system to its previous state. This stops any lasting damage to services. Here are the main points for creating service rollback steps:

  a. If a forward operation fails, you might need to roll back. So, consider all rollback scenarios.

  b. Rollback types include lossy and lossless. Lossless rollback happens when no new data is added at the destination, allowing data to return to its original state. If new data appears at the destination and cannot be synced back to the source, the rollback is lossy, causing data loss. To avoid data loss with new destination data, create a sync task from the destination to the source.

  c. In large-scale service applications, rollback can be full or partial, depending on the service impact. For example, if 10 application systems and 10 databases switch on the same day, the service team must check if the cross-cloud access latency between applications and databases meets the requirements. If a database fails to switch, decide whether to roll back all databases or just that one.

When designing a cutover runbook, consider the rollback process. Create a clear rollback plan and procedure. Assign specific operators and follow the steps closely. This ensures that if issues arise during the cutover, the rollback can happen smoothly, avoiding service disruptions.

6.4.4.5 Runbook Reference Template

- Runbook checklist reference

Table 6-19 Runbook checklist example

Category	Preparations	Responsible Department	Activity	Involved	Completed	Planned Completion Time	Owner
Organizational and assurance preparation	N/A	Project manager	N/A	Yes	N/A	N/A	N/A
	N/A	Project manager	N/A	Yes	N/A	N/A	N/A
Third party/ Business model	N/A	Service-related departments	N/A	Yes	N/A	N/A	N/A

Category	Preparations	Responsible Department	Activity	Involved	Completed	Planned Completion Time	Owner
Environment checklist verification	Application list checking and application startup and stop sequence updating	R&D-related departments	N/A	Yes	N/A	N/A	N/A
	N/A	R&D-related departments	N/A	Yes	N/A	N/A	N/A
Environment (including source and destination ends, migration tasks, and execution scripts) check	Basic cloud service check	O&M-related departments	N/A	Yes	N/A	N/A	N/A
	Database check	O&M-related departments	N/A	Yes	N/A	N/A	N/A
		O&M-related departments	N/A	Yes	N/A	N/A	N/A
	Big data check items	Big data-related departments	N/A	Yes	N/A	N/A	N/A
		Big data-related departments	N/A	Yes	N/A	N/A	N/A

Category	Preparations	Responsible Department	Activity	Involved	Completed	Planned Completion Time	Owner
	Application check	O&M-related departments	N/A	Yes	N/A	N/A	N/A
	Script check	O&M-related departments	N/A	Yes	N/A	N/A	N/A
		O&M-related departments	N/A	Yes	N/A	N/A	N/A
	Log system check	O&M-related departments	N/A	Yes	N/A	N/A	N/A

- Runbook procedure reference

  In addition, the following items may be included, such as the actual start time, actual end time, and actual execution time.

  Table 6-20 shows an example. Fill out the table based on actual service needs.

Table 6-20 Runbook procedure example

No.	Task	Step	Subtask	Procedure	Instruction	Decision	Details	Operator	Confirmer	Planned Start Time	Planned End Time
1	XXX	1.1	N/A	N/A	N/A	Must be resolved	N/A	N/A	N/A	N/A	N/A

No.	Task	Step	Subtask	Procedure	Instruction	Decision	Details	Operator	Confirmer	Planned Start Time	Planned End Time
2		1.1	N/A	N/A	N/A	Must be resolved	N/A	N/A	N/A	N/A	N/A
3		1.2	N/A	N/A	N/A	Must be resolved	N/A	N/A	N/A	N/A	N/A
4		1.3	N/A	N/A	N/A	Must be resolved	N/A	N/A	N/A	N/A	N/A
5		1.3	N/A	N/A	N/A	Must be resolved	N/A	N/A	N/A	N/A	N/A
6		1.3	N/A	N/A	N/A	Must be resolved	N/A	N/A	N/A	N/A	N/A
7	XXX	1.4	N/A	N/A	N/A	Non-blocking	N/A	N/A	N/A	N/A	N/A
8		1.4	N/A	N/A	N/A	Non-blocking	N/A	N/A	N/A	N/A	N/A
9	Rollback decision 1: Complete the preceding steps before XX (a time point). Otherwise, decide whether to perform a rollback.					N/A	Rollback - decision point 1			N/A	N/A
10	XXX	21.1	N/A	N/A	N/A	Blocking	N/A	N/A	N/A	N/A	N/A

No.	Task	Step	Subtask	Procedure	Instruction	Decision	Details	Operator	Confirmer	Planned Start Time	Planned End Time
11		21.2	N/A	N/A	N/A	Blocking	N/A	N/A	N/A	N/A	N/A

6.4.5 Deployment

6.4.5.1 Cloud Resource Provisioning and Configuration

Deployment includes setting up and testing the cloud environment and getting the migration ready.

Set up cloud resources based on your application's design. You can set up cloud resources in these three ways:

- Create them manually on the cloud platform.
- Write scripts or connect to an automation platform to use the cloud platform's APIs to set up resources in bulk. Each cloud service has an API for resource lifecycle management. Check the service's help document for API details.

  To create an ECS using an API, see **Creating an ECS**.
- Use Huawei Cloud Resource Formation Service (RFS) to manage and set up resources in bulk. See **RFS documentation** for details.

A comparison table of the three methods is below. Choose the method that best fits your needs.

Table 6-21 Comparison of the three methods

Method	Scenario	Advantage	Disadvantage
Manual setup on the console	Setting up a few cloud resources	Easy operation	It takes a lot of time with many cloud resources.

Method	Scenario	Advantage	Disadvantage
Using scripts to call APIs for resource setup	- Setting up many cloud resources - Customizing services (The program runs as needed to create and delete resources automatically.)	- Automation: It reduces resource management and labor. - Flexibility: It allows you to quickly create, configure, start, and stop cloud resources for flexible deployment. - Programmability: You can use APIs with rich functions and parameters to perform secondary development using programming languages to meet specific needs. - Speed: It eliminates manual steps. - Consistency: It ensures uniform resource deployment and reduces misoperation risks.	- Learning curve: You need to learn APIs, programming languages, and tools. - Complexity: Scripts become harder to manage and maintain as services grow. - Security risks: Improper security measures can expose critical data or lead to resource abuse.

Method	Scenario	Advantage	Disadvantage
AOS resource orchestration	Setting up many cloud resources	- Automation: It automates cloud resource deployment and management, boosting efficiency. - Visualization: Templates show resource dependencies and configurations, reducing errors and improving management. - Reusability: Templates can be reused and edited, saving time and effort. - Traceability: Audit features help trace faults and backtrack issues. - Consistency: It ensures consistent resource configurations, minimizing human errors.	- Learning Curve: You need to learn the template language and cloud services. - Complex debugging: Errors involving multiple resources take longer to find. - Security risks: You need to securely manage sensitive information and keys. - Risk management: Template errors can make resources unavailable. - Not suitable for special scenarios: It does not fit cases needing complex interactions or manual steps. - Possible problems: Missing or misconfigured dependent resources can cause failures. - Restrictions: Direct OS control is limited, and not all resource types are supported.

6.4.5.2 Migration Tool Deployment

Huawei Cloud provides the following migration tools: Migration Center (MgC), Resource Discovery and Assessment (RDA), Server Migration Service (SMS), Data Replication Service (DRS), Redis data migration tool, Cloud Data Migration (CDM), Object Storage Migration Service (OMS), and the like.

- **MgC**: a complete migration platform. It combines several Huawei Cloud migration tools and offers ready-to-use migration workflows based on best practices. Choose a suitable template for your specific migration needs.

  For more information, see the **Migration Center Documentation**.

- **RDA**: a tool for Windows that helps assess your readiness for cloud migration. It discovers your application's infrastructure, including VM specifications, CPU, memory usage, and network topology. RDA also provides recommendations for moving applications to Huawei Cloud and offers full server migration support.

- SMS is a P2V or V2V migration service that allows you to migrate x86 physical servers or VMs on private or public clouds to ECSs on Huawei Cloud.

  See the **Server Migration Service Documentation** for more information.

- **DRS**: a cloud service for real-time database migration and synchronization. It offers real-time migration, backup migration, real-time synchronization, data subscription, and real-time DR.

  See the **Data Replication Service Documentation** for more information.

- **Redis**: migrates data from your self-hosted Redis or other Redis services to Redis on Huawei Cloud. The source Redis must support the **SYNC** and **PSYNC** commands.

  See the **Distributed Cache Service Documentation** for more information.

- **CDM**: supports nearly 20 common data sources to meet different data migration scenarios on and off the cloud.

  See the **Cloud Data Migration Documentation** for more information.

- **OMS**: migrates object storage data from other cloud vendors to Huawei Cloud OBS.

  See the **Object Storage Migration Service Documentation** for more information.

Deploy RDA on the Windows ECS in the Huawei Cloud VPC. OMS is a public service and does not use the VPC's internal IP addresses. If you use Direct Connect for data migration, deploy the offline OMS tool on an ECS in the Huawei Cloud VPC. Other tools will temporarily use the VPC's internal IP addresses.

6.4.6 Migration

6.4.6.1 Access Layer Migration

- **EIP**

  EIPs do not need to be moved. You usually need to buy new EIPs in your destination Huawei Cloud environment. If an EIP will serve areas outside Huawei Cloud, register its domain name and IP with the local information security office. Once approved, the EIP can connect to external systems.

  For how to purchase and use an EIP, see **Elastic IP Documentation**.

- **Load balancing**

  You can replace your current hardware or software load balancer with Huawei Cloud ELB. Set up ELB at the destination and configure it based on your current settings.

  For details about the deployment process, see **Elastic Load Balance Documentation**.

  If your load balancer is self-built on a server, you can use SMS to move it to Huawei Cloud.

- **VPN**

  The VPN needs to be re-deployed on the destination Huawei Cloud. For details, see the **Virtual Private Network Documentation**.

6.4.6.2 Application Layer Migration

- **Server migration**

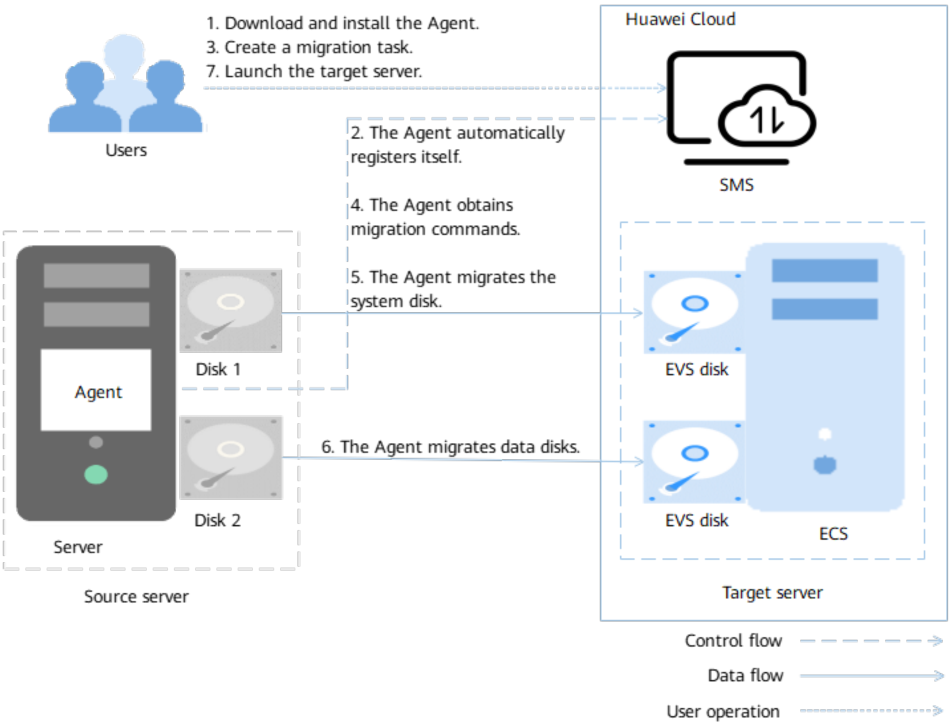
  Server migration is a common type of rehosting. While servers can run many applications like Nginx, databases, containers, middleware, and big data tools, some applications focus on data. For these, we often use separate data migration instead of server migration. Server migration moves both the applications and the OS. There are three types of server migration:

Table 6-22 Server migration methods

Server	Method	Scenario	Remarks
VM/PM	Redeployment	The customer does not need to keep the OS and can accept long downtime.	-
	Migration using SMS (free of charge)	The customer wants to keep the OS but can accept only a short downtime.	Recommended. Huawei Cloud offers the technical support.
	Image import and export	The customer wants to keep the OS and can accept long downtime.	-

- **Redeployment**: For public cloud migration, use the CI/CD pipeline to automate and redeploy your applications.
- **Migration using SMS**: SMS helps you move applications and data from on-premises x86 servers or VMs on other clouds to Huawei Cloud's ECSs.

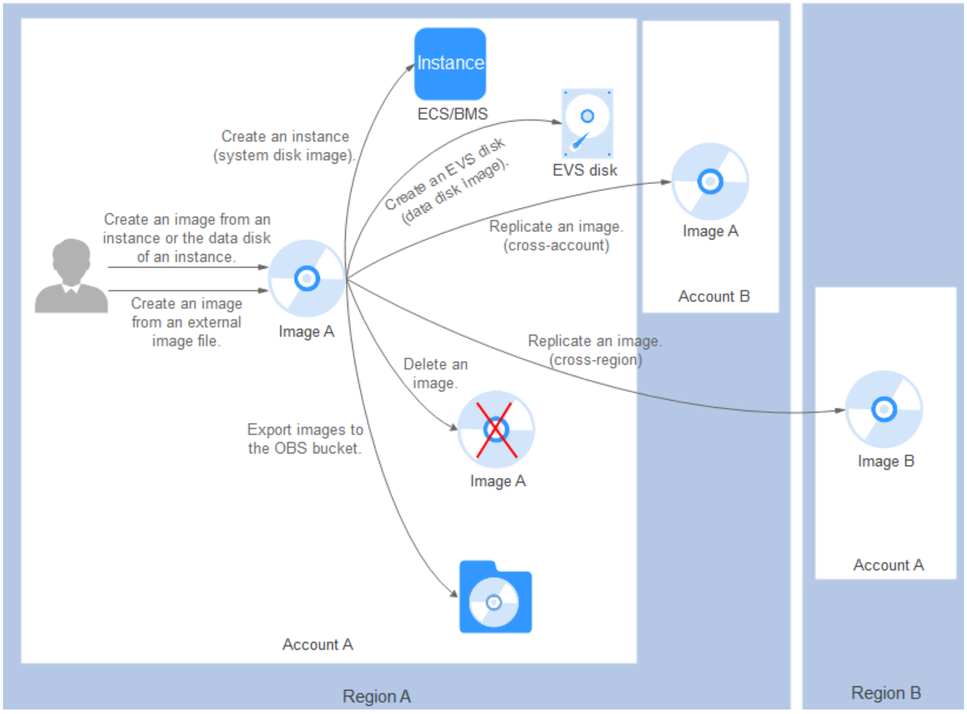
Figure 6-21 Process of migration using SMS


1. Download and install the Agent.
3. Create a migration task.
7. Launch the target server.
Users
Huawei Cloud
SMS
2. The Agent automatically registers itself.
4. The Agent obtains migration commands.
5. The Agent migrates the system disk.
6. The Agent migrates data disks.
Agent
Disk 1
Disk 2
Server
Source server
EVS disk
EVS disk
ECS
Target server
Control flow
Data flow
User operation

The Agent on the source server gets the migration instruction from SMS, then creates a security certificate and key. It sends these to the target server using Huawei Cloud OpenStack metadata management. Both servers then restart and use the new certificate to set up a secure SSL channel. Services stay up during migration, with just a brief pause needed before starting the target ECS in Continuous Synchronization mode. SMS keeps downtime to a minimum. For details about SMS, see the **Server Migration Service Documentation**.

- **Image import and export**: Server migration can use Huawei Cloud's Image Management Service (IMS) to create private images from existing ECSs or external image files. You can also import existing service cloud images to the cloud platform. This helps with moving services to the cloud and deploying them in batches.

Figure 6-22 Image import and export solution



If SMS cannot move your whole system over the network, use IMS instead. Make private images from the source server's system and data disks, then upload these images to Huawei Cloud OBS. In IMS, create private images from the uploaded files and use them to set up ECSs. After migration, the host OS, system settings, and data files will match those on the source server. For details, see the **Image Management Service Documentation**.

- **Container migration**

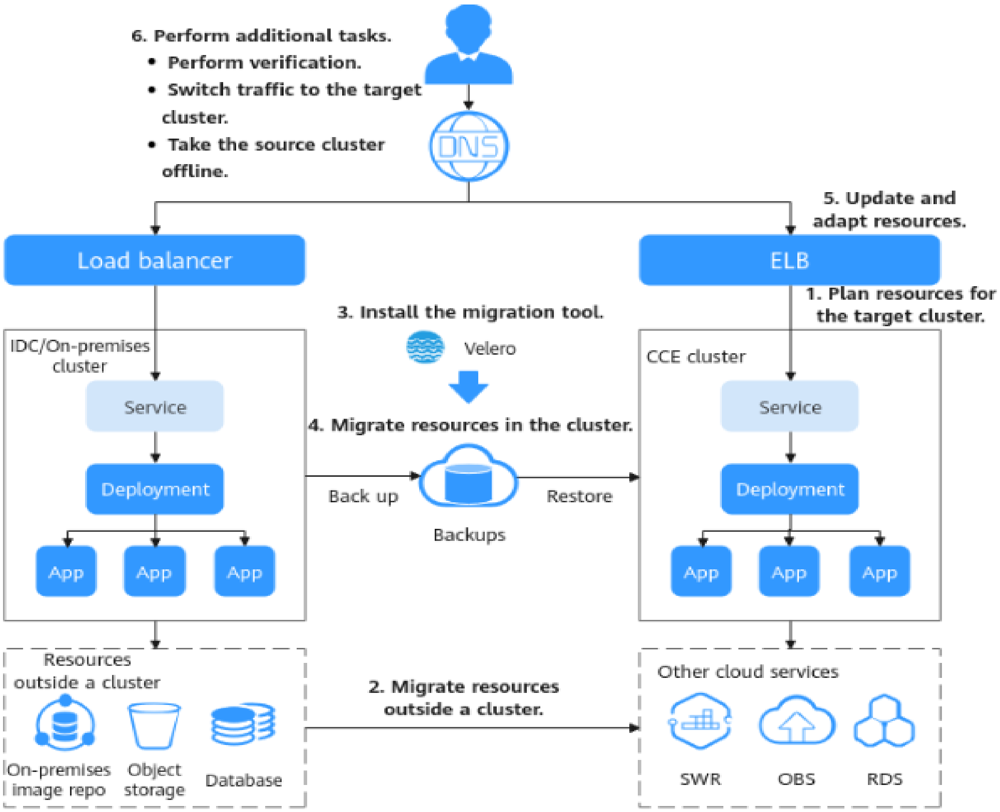
  A container is a lightweight high-performance resource isolation mechanism implemented based on the Linux kernel. It is a built-in capability of the OS kernel. Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications. For developers, Kubernetes is a cluster operating system. Kubernetes provides functions such as service discovery, scaling, load balancing, self-healing, and even leader election, freeing developers from infrastructure-related configurations.

  Cloud Container Engine (CCE) is a highly scalable, high-performance, enterprise-class Kubernetes service to run Docker containers. With CCE, you can easily deploy, manage, and scale containerized applications on Huawei Cloud. CCE is an enterprise-grade container service built on open-source Kubernetes. It helps businesses manage containerized apps with high performance and reliability. CCE supports native Kubernetes apps and tools, making it easy to set up a container runtime in the cloud.

  You can access CCE, a hosted Kubernetes service, using the CCE console, kubectl, or Kubernetes APIs. For details, see the **Cloud Container Engine Documentation**.

  This section uses migrating containerized applications from other clouds to CCE as an example. The following figure shows the details.

Figure 6-23 Container migration


6. Perform additional tasks.
• Perform verification.
• Switch traffic to the target cluster.
• Take the source cluster offline.
DNS
5. Update and adapt resources.
Load balancer
ELB
1. Plan resources for the target cluster.
3. Install the migration tool.
Velero
IDC/On-premises cluster
CCE cluster
Service
Service
4. Migrate resources in the cluster.
Deployment
Deployment
Back up
Restore
Backups
App
App
App
App
App
App
Resources outside a cluster
2. Migrate resources outside a cluster.
Other cloud services
On-premises image repo
Object storage
Database
SWR
OBS
RDS

The container image migration procedure is as follows:

a. Export the container images used by the third-party cluster.

b. Pull the images to the client machine by following the instructions provided in the third-party container image service.

c. Upload the image files to Huawei Cloud SoftWare Repository for Container (SWR).

d. Run the **docker pull** command to upload the image to Huawei Cloud. For details, see the **SoftWare Repository for Container Documentation**.

6.4.6.3 Middleware Layer Migration

- **Redis migration**

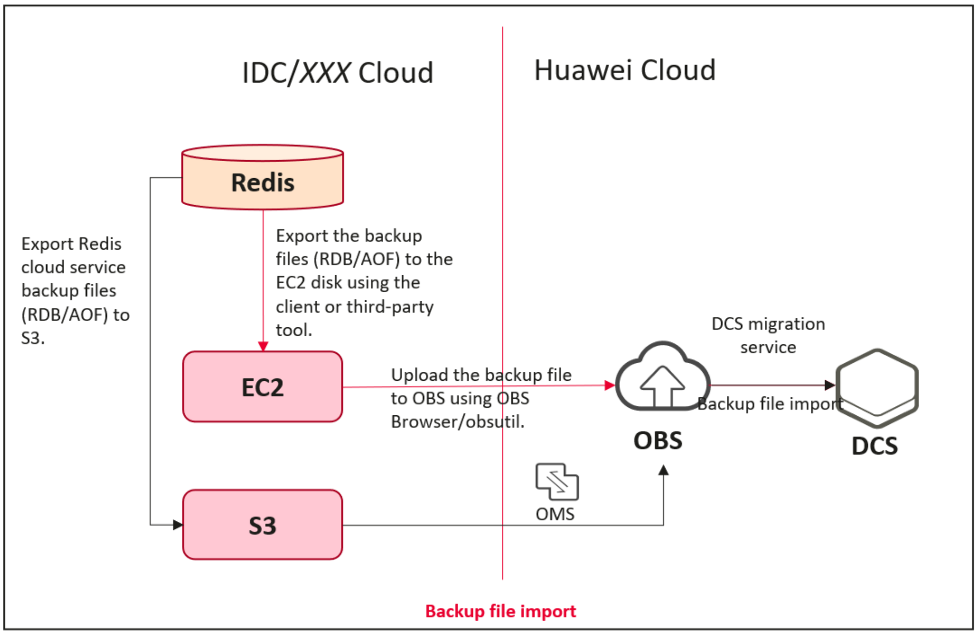
  Use Redis for caching and database storage. You usually do not need to migrate cache data. Most of the time, data is re-cached in Huawei Cloud SQL databases. Migrate data only if Redis serves as your main database. In web scenarios, you do not need to move cached sessions. When a client logs in again, the data is stored in Huawei Cloud DCS.

- **Offline backup import**

  You can import backup files to Redis 3.0, 4.0, and 5.0. RDB backups from self-hosted Redis 5.0 cannot be imported. Export AOF or RDB backup files to use DCS for migrating data from on-premises or third-party cloud Redis services. This is a full migration, so new data created during the process will not be migrated.

  Migration slows down if the source server has more than 10 GB of data. This method is not recommended for servers with over 10 GB of memory usage.

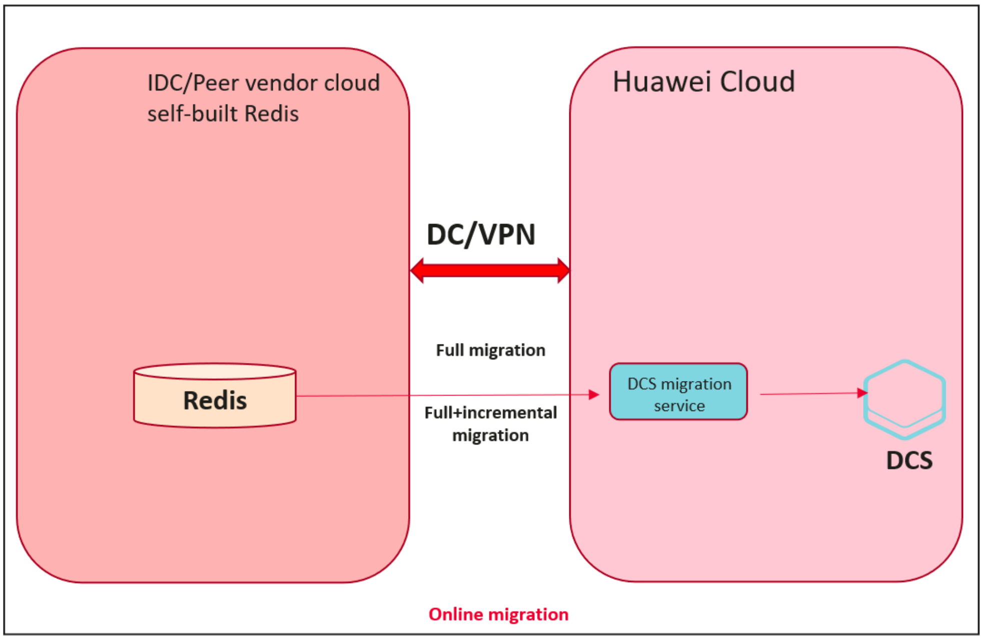
Figure 6-24 Redis migration


IDC/*XXX* Cloud
Huawei Cloud
Redis
Export Redis cloud service backup files (RDB/AOF) to S3.
Export the backup files (RDB/AOF) to the EC2 disk using the client or third-party tool.
EC2
Upload the backup file to OBS using OBS Browser/obsutil.
DCS migration service
Backup file import
OBS
DCS
OMS
S3
Backup file import

- The migration process is as follows:
  a. Back up your source Redis data.
     - In IDC, use a third-party tool or set a policy to save backups to disks. See Redis CLI or Redis port for details.
     - Use the backup feature to export the RDB file to S3.
  b. Upload the backup to OBS.
     - In EC2, use OBS Browser or obsutil to upload the backup files (**aof**/**rdb**) to OBS in the same region as DCS.
     - In S3, create an OMS task to move the RDB backup file to OBS in the same region as DCS.
  c. Set up a DCS migration task.
     - On the DCS console, create a migration task, choose the backup file import option, select the **aof** or **rdb** file from the source OBS, and configure other settings to start the task.
  d. Check and compare data on both sides using DCS.
     - Run the redis query command **info keyspace**.
- **Online migration**

  Online migration moves all data and synchronizes new changes in real time. This needs the source and destination to connect over a private network. Also, the **SYNC** and **PSYNC** commands must be enabled on the source Redis.

Figure 6-25 Online migration



The migration process is as follows:

a. Create a DCS migration task.

On the DCS console, create a migration task. Choose the online migration mode. Select either full migration or full plus incremental migration as needed. Set other parameters to start the task.

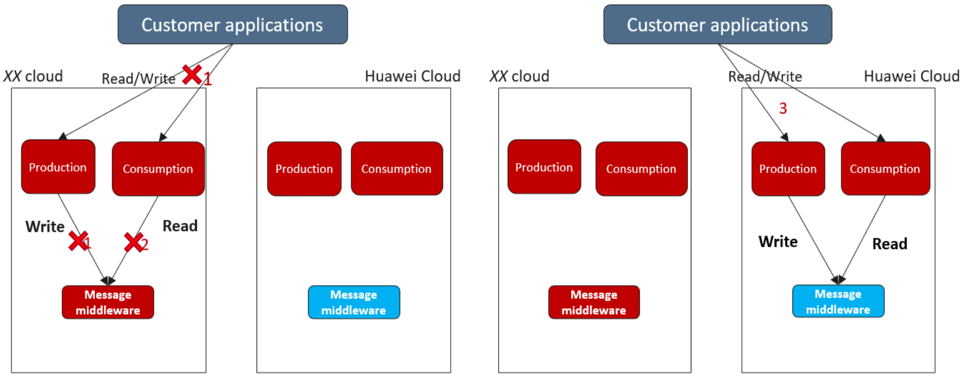
b. Check and compare data on both sides using DCS.

After the migration task is complete, run the Redis query command **info keyspace**.

- **Message middleware migration**

  Message middleware includes Kafka, RabbitMQ, RocketMQ, and IBM MQ. In a migration project, these systems are typically moved using a policy switchover method.

Figure 6-26 Message middleware migration



The migration process is as follows:

a. Stop enterprise applications and message production until all messages are consumed.

b. Stop the consumption service.

c. Start the message production and consumption services on Huawei Cloud, receive customer traffic, and check if services work normally.

6.4.6.4 Data Layer Migration

1. **Object storage migration**

   Object storage is suitable for storing unstructured data, such as documents, texts, images, XML and HTML files, reports, and audio and videos. The following figure shows how to migrate data at different scales.

Figure 6-27 Object storage migration

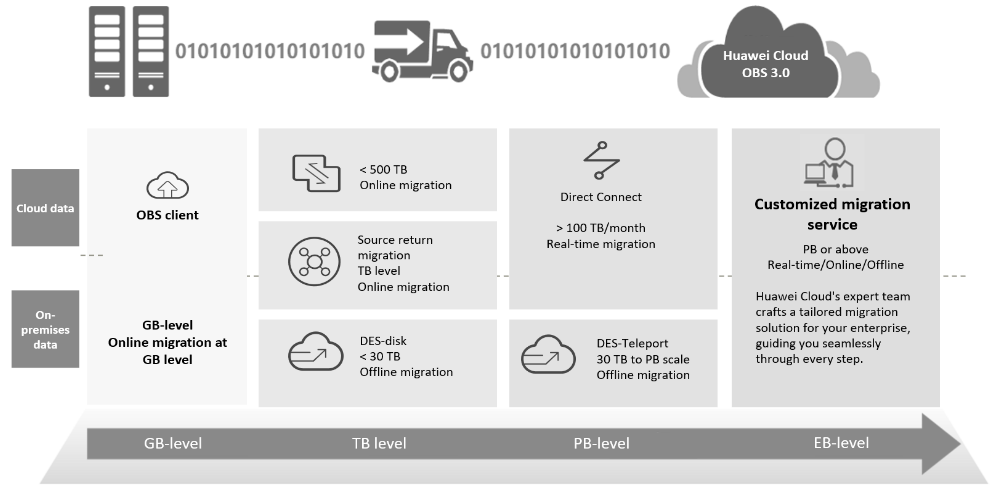

0101010101010101
0101010101010101
Huawei Cloud OBS 3.0
Cloud data
On-premises data
OBS client
GB-level Online migration at GB level
< 500 TB Online migration
Source return migration TB level Online migration
DES-disk < 30 TB Offline migration
Direct Connect
> 100 TB/month Real-time migration
DES-Teleport 30 TB to PB scale Offline migration
Customized migration service
PB or above Real-time/Online/Offline
Huawei Cloud's expert team crafts a tailored migration solution for your enterprise, guiding you seamlessly through every step.
GB-level
TB level
PB-level
EB-level

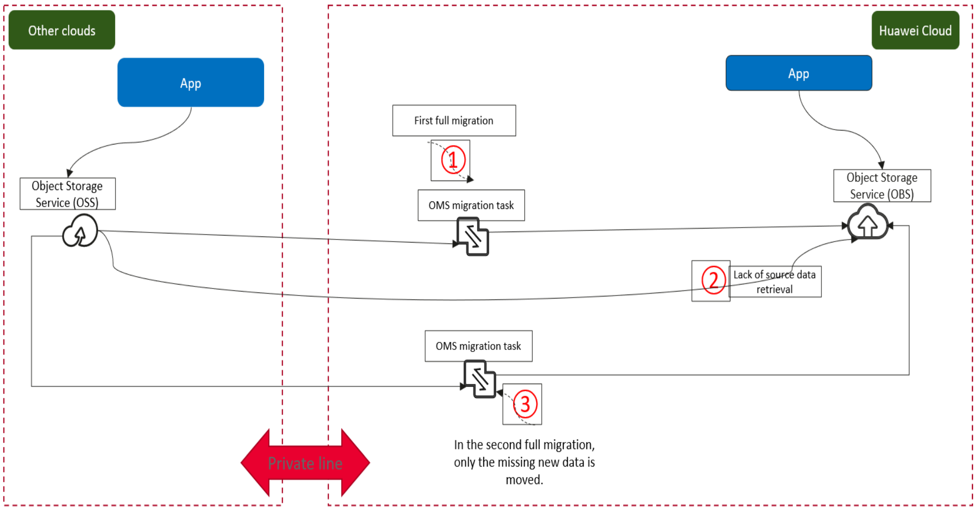
Table 6-23 Object storage migration methods

Item	Method	Scenario	Remarks
Object storage migration	Copying data to OBS	The data volume is small, around a few gigabytes.	-
	Migrating and retrieving data using OMS	The data volume is large, ranging from terabytes to petabytes, with significant increases.	Recommended. Huawei Cloud offers the technical support.
	Copying DES disks offline	The data volume is large, with terabytes of incremental data.	-

- **Migrating and retrieving data using OMS**: OMS is a user-friendly and efficient online data migration service. It uses the SDK of your current object storage to quickly move, encrypt, and copy data to Huawei Cloud OBS. This makes it easy to transfer object storage data from other cloud providers to Huawei Cloud. The first full migration moves all data from your current object storage to Huawei Cloud OBS. When switching services, set up retrieval rules on Huawei Cloud to move new data to Huawei Cloud OBS. After the switch, run another full migration. In this step, already moved objects are skipped, and only the new data is transferred to Huawei Cloud OBS.

Use the OMS to view and manage migration tasks, create task groups, assess bucket data, and check audit logs. See the **Object Storage Migration Service Documentation** for details.

Figure 6-28 OMS


Other clouds
App
Object Storage Service (OSS)
First full migration
①
OMS migration task
Huawei Cloud
App
Object Storage Service (OBS)
②
Lack of source data retrieval
OMS migration task
③
In the second full migration, only the missing new data is moved.
Private line

① Set up an OMS full migration task to move all OSS data to Huawei Cloud OBS.

② Set up Huawei Cloud OBS source return rules in the console to recover data lost during application access to Huawei Cloud.

③ Set up an OMS full migration task to move the missing incremental data to Huawei Cloud OBS.

- **Copying DES disks offline**

  The Data Express Service (DES) helps you move data to Huawei Cloud. It can handle terabytes or petabytes of data. DES uses physical storage devices like USB drives or eSATA disks, or a tool called Teleport, to transfer large amounts of data quickly. This solves issues with high cloud bandwidth costs and long upload times.

  For details, see the **Data Express Service Documentation**.

  Huawei Cloud offers custom migration services for moving storage data. Their expert teams create tailored migration plans and offer complete support. They assist with both full and incremental data moves, ensuring accuracy at each step. This service helps transfer data from local storage to the cloud.

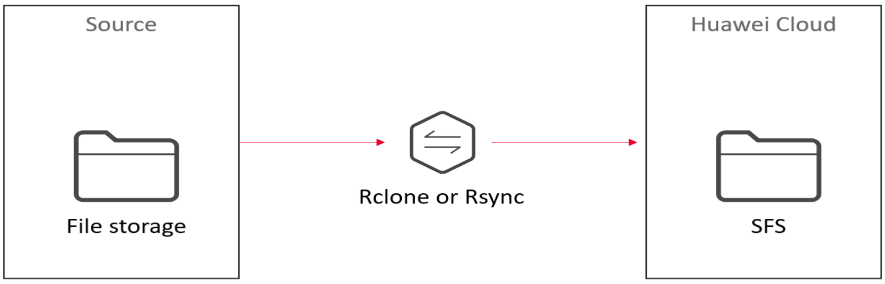
- **File storage migration**

  Network Attached Storage (NAS) is a file system for sharing, scaling, and reliable, fast data access.

  Scalable File Service (SFS) offers scalable, high-performance file storage over a network. It allows shared file access between multiple Elastic Cloud Servers (ECSs), Bare Metal Servers (BMSs), and containers on Cloud Container Engine (CCE) and Cloud Container Instance (CCI).

If the network is connected during NAS migration, you can mount files to multiple platforms. Use Rclone or Rsync to copy files from the source directory to the SFS or SFS Turbo file system on the transit host.

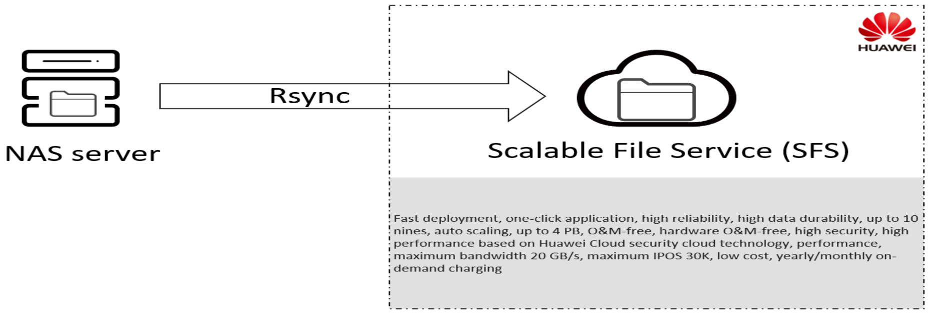
Figure 6-29 File migration



Rclone is a CLI tool for synchronizing, uploading, and downloading data between various cloud storage and web disks. It allows users to set up multiple threads and concurrent tasks, which speeds up migrations and reduces time. Rclone also checks and synchronize data, copying files from one system to another for NAS file migration.

Remote synchronize (Rsync) is a Remote Copy Protocol (RCP)-based remote file synchronization tool that uses its own Rsync algorithm.

Figure 6-30 Rsync migration



Fast deployment, one-click application, high reliability, high data durability, up to 10 nines, auto scaling, up to 4 PB, O&M-free, hardware O&M-free, high security, high performance based on Huawei Cloud security cloud technology, performance, maximum bandwidth 20 GB/s, maximum IPOS 30K, low cost, yearly/monthly on-demand charging

Rsync is enabled on each service platform. The Rsync client script is set up on the file server. A scheduled task, like Crontab on Linux, syncs files from the service server to the disk array incrementally (fully the first time). It has the following advantages:

a. Incremental file synchronization is fast, efficient, and uses fewer resources.

b. It supports network access via SSH and RSH.

c. Stopping the Rsync service does not affect other services.

d. It supports mirrored data copies, ideal for distributing static files online with good scalability.

- **Database migration**

Databases are classified into relational databases and non-relational databases. Relational databases are crucial because they keep data consistent. To move your database to Huawei Cloud, choose one of these methods:

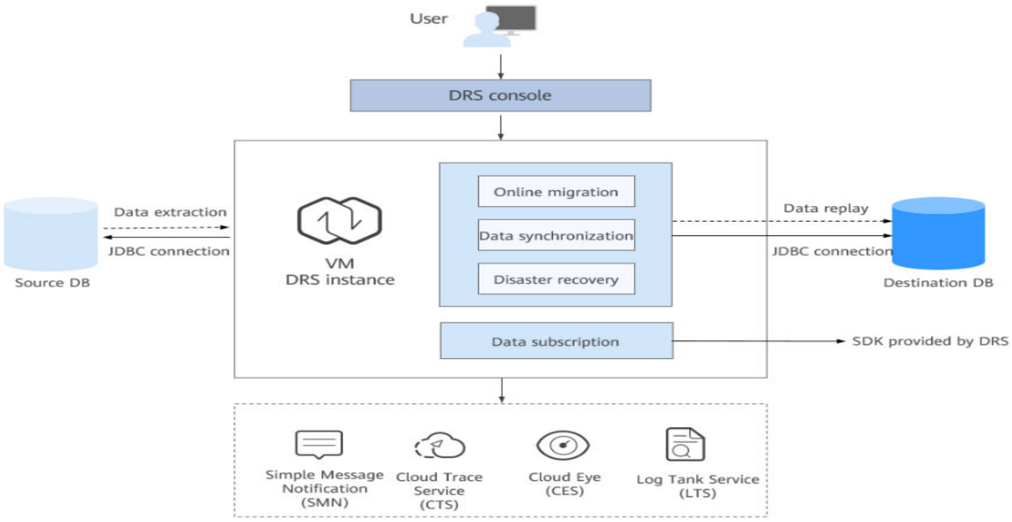
a. Use SMS host migration. If your source is a self-built database on an ECS, migrate the ECS database server to Huawei Cloud's ECS. SMS makes this easy but will pause database services during the process. For minimal downtime, use database backups or synchronization.

b. Use backup and restoration. This method involves saving the source data, stopping the source database, performing a full backup, keeping necessary logs, and restoring the data on the target system.

c. Use Data Replication Service (DRS) for online migration and real-time synchronization. DRS works across cloud platforms, on-premises to cloud, and between regions. It is a reliable, secure, and efficient service that simplifies data flow and reduces transmission costs. It simplifies the data flow between databases, significantly reducing data transmission costs.

   DRS supports various data sources and offers real-time migration, backup migration, real-time synchronization, data subscription, and DR. For details, see the **Data Replication Service Documentation**.

Figure 6-31 DRS migration



The following are common best practices for database migration. See **DRS Best Practices Summary** for more best practices.

Table 6-24 Best practices for database migration

Source DB	Destination DB	Best Practice
Other cloud MySQL databases	Huawei Cloud RDS for MySQL	**Migrating MySQL Databases from Other Clouds to RDS for MySQL**
	Huawei Cloud GaussDB(for MySQL)	**From Other Cloud MySQL to TaurusDB**

Source DB	Destination DB	Best Practice
Other cloud MongoDB	Huawei Cloud DDS	**From Other Cloud MongoDB to DDS**
On-premises MySQL	Huawei Cloud RDS for MySQL	**From ECS-hosted MySQL to RDS for MySQL**
ECS-hosted MySQL	Huawei Cloud GaussDB(for MySQL)	**From ECS-hosted MySQL to TaurusDB**
ECS-hosted MongoDB	Huawei Cloud DDS	**From ECS-hosted MongoDB to DDS**
On-premises MySQL	Huawei Cloud RDS for MySQL	**From On-Premises MySQL to RDS for MySQL**
On-premises MongoDB	Huawei Cloud DDS	**From On-Premises MongoDB to DDS**
Huawei Cloud RDS for MySQL	Distributed Database Middleware (DDM)	**From RDS for MySQL to DDM**
MySQL schema and logic table	Distributed Database Middleware (DDM)	**From MySQL Schema and Logic Table to DDM**
On-premises Microsoft SQL Server	Huawei Cloud RDS for SQL Server	**Migrating Microsoft SQL Server Backup Data to RDS SQL Server DB Instance**

6.4.6.5 Migration FAQs

Migration FAQs

For migration FAQs, see the following link:

- **SMS FAQs**
- **FAQs in Redis Data Migration**
- **OMS FAQs**
- **FAQs About Database Migration**

6.4.7 Verification

6.4.7.1 Data Verification

- **Data verification criteria**

  After the migration is complete, it is necessary to compare the data consistency between the source and destination ends. The precision of data consistency comparisons varies by scenario. Typically, the database tables for core businesses must be 100% consistent between the source and destination ends. For certain scenarios in big data businesses, such as user profile computing, 90% consistency of raw data is acceptable. Below is reference criteria that can be adjusted based on site requirements.

Table 6-25 Reference for data verification criteria

Category	Data Consistency Requirement	Example Business
Core business	100%	The core member data, transaction data, and payment data of an e-commerce system are the most critical assets for users, involving real financial amounts. Therefore, these core services require 100% data consistency. You are advised to perform data comparison on rows, objects, and sample values.
Non-core business	99.9%	The shopping cart data and customer service communication message data of an e-commerce system are non-core business data. A minor loss will not impact the customer's service usage or experience. If the switchover time is limited, it is suggested to only compare the number of data rows.
Peripheral business	90%	The home page recommendation data, user browsing data, and user profile data of an e-commerce system, if partially lost, will not affect the customer's service usage or experience. It is suggested to perform row-level comparisons and sample value checks.

- **Data verification methods**

  Data is categorized into database data, middleware data, and file data, each with distinct consistency verification methods and tools.

  – The methods for verifying database data consistency are described in the following table.

Table 6-26 Database data consistency comparison methods

Item	Tool	Description
Database- and table-level content comparison	DRS tool	Query and compare each data record in the database table to ensure that every field in each record matches the corresponding fields in the source database table. Content comparison is generally slower than row comparison.
	Python script	Based on the DRS task ID, call an API to batch execute comparison tasks and export the results to an XLSX file. Compared to the DRS tool, Python scripts can be executed in batches, improving execution efficiency.
Database- and table-level object comparison	DRS tool	Compares objects like databases, indexes, tables, views, stored procedures, functions, and table sorting rules.
	Python script	Based on the DRS task ID, call an API to batch execute comparison tasks and export the results to an XLSX file. Compared to the DRS tool, Python scripts can be executed in batches, improving execution efficiency.
Database and table-level row count comparison	DRS tool	Compare the number of rows in tables to check if they are consistent. Only the row count is queried, making the comparison process faster.
	Python script	The batch script creates N concurrent task threads, iterates through all tables individually, and outputs the comparison results to an XLSX file. Compared to the DRS tool, Python scripts can be executed in batches, improving execution efficiency.

– The methods for verifying middleware data consistency are described in the following table.

Table 6-27 Middleware data consistency comparison methods

Item	Tool	Description
Keys quantity comparison	redis-cli	Run the redis-cli command **info keyspace** to view the values of the **keys** and **expires** parameters, subtract **expires** from **keys** in the source Redis, check the difference, and repeat this for the target Redis. If both differences are the same, it indicates that the number of keys is consistent and the migration is successful.

Item	Tool	Description
Key-value content comparison	Open-source Redis-Full-Check	Data verification is conducted by fully comparing the data content in the source and destination Redis. The tool captures data from both ends multiple times for differentiated comparisons, recording inconsistent data for subsequent rounds. Through continuous rounds of comparison, the differences gradually converge. The final discrepancy results are stored in SQLite; if no discrepancies exist, it indicates that the data content is complete and the migration was successful.

– The methods for verifying file data consistency are described in the following table.

Table 6-28 File data consistency comparison methods

Type	Item	Tool	Description
Object storage	Object quantity	OMS	The OMS migration tool verifies file integrity through MD5 and checks if the data volume of objects in both buckets is consistent.
File storage	File quantity	Rclone	Rclone utilizes MD5 hash values to verify file integrity. Post-synchronization, it compares the quantity of files at the source and destination ends.
		Rsync	Rsync utilizes MD5 hash values to validate file integrity; if the checksums do not match, Rsync will retransmit the file to guarantee data consistency. Post-synchronization, it compares the quantity of files at the source and destination ends.
	File size	Python script	After the migration is completed, it determines if there is consistency by comparing the total file sizes at the source and destination ends.
	File content	Python script	After the migration is completed, it compares the hash values of the source and destination files by calculating them to verify if they match.

6.4.7.2 Service Verification

Service verification is crucial for cloud migration. Service verification primarily includes function verification and performance verification. During the cloud migration process, there are two stages where service verification is necessary. Firstly, upon completion of service deployment, function and performance verifications must be conducted prior to switching over. Secondly, during the service transition, once the service traffic has been redirected to the target end, post-switch function and performance verifications should also be carried out.

Function Verification

- **Function testing content**

  Function tests ensure that the application system runs properly prior to launch. Below are the details of function tests:

Table 6-29 Function testing content

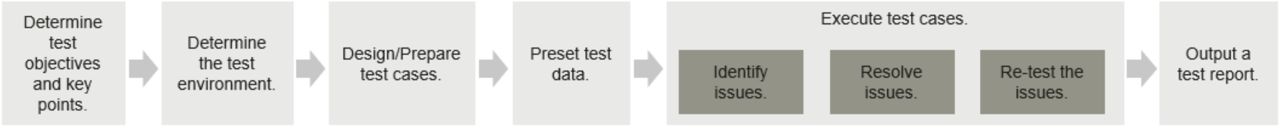
Testing Content	Description
Application functions	The test content heavily relies on the functions of the application system. For example, in an e-commerce system, the core function test cases must encompass online and offline browsing, shopping, placing orders and making payments (via various payment methods and using coupons), printing bills, issuing invoices, promotional activities, inventory synchronization, new member registration, existing member cancellation, order refunds, and returning coupons among other essential features.
Peripheral system integration functions	The test content significantly depends on the integration capabilities of the application system. For example, a major retail e-commerce platform collaborates with a group-buying platform, a food delivery platform, a home-service platform, and a short-video platform. The integration scenarios should minimally verify functions such as placing orders, utilizing coupons, notifying shipments, and leaving reviews across these integrated platforms.

- **Function testing purposes**
  - Verify that the application functions are normal after replacing technical components following the migration of the application system to Huawei Cloud.
  - Confirm that the integration between the application and surrounding systems is functioning properly post-migration to the target environment, ensuring all necessary modifications identified for collaboration with external systems have been implemented appropriately.
- **Function testing methods**

- Smoke testing: Smoke testing is a straightforward form of function testing aimed at verifying the system's availability through the execution of a minimal set of critical test cases. Upon completion of deployment on the target end, conducting a smoke test first ensures the fundamental functions of the system are operational.
- Full-business function testing: This thorough examination verifies the proper functioning of all system features by running test cases tailored to diverse business workflows, thereby ensuring the integrity of every function module.
- Log analysis: Post-deployment of services on the destination end, it is necessary to analyze system logs for any signs of errors. Log analysis serves as a proactive measure to unearth latent issues and vulnerabilities, facilitating prompt rectification and optimization efforts.
- DNS hijacking test: Since services deployed on the cloud are typically configured according to the domain names of the production environment, when testing service functions using a mobile app or browser, it is necessary to employ DNS hijacking methods for testing. This can involve utilizing an internal WiFi along with APISIX modified for operations, coordinating with the DNS resolution on the WiFi to redirect traffic towards the test environment for conducting internal network tests.

- **Function testing process**

Figure 6-32 Function testing workflow



a. **Determine test objectives and key points**: Specify the application functions and scenarios to be tested, as well as the key points and focuses of the test.

   i. System functions: such as promotion activities, coupon payments, returns, and refunds of coupons.

   ii. Batch processing job function: If the migration involves multiple batch processing jobs, focus on their execution during testing, for example, inventory data pushing.

   iii. Third-party service integration function: Verify integrations like those with a group-buying platform, food delivery platform, home service platform, short-video platform, and in-store POS payment functionality.

b. **Determine the test environment**: Identify the environment to use for testing and ensure it does not impact production services. Note: For services calling third-party APIs, isolate external networks to avoid contaminating production data. Set up a dedicated internal WiFi for testers to simulate third-party function tests securely.

Table 6-30 Comparison and analysis of test environments

Scenario	Suggestion on Test Environment Selection	Advantage	Disadvantage
Some applications have been brought online in the Huawei Cloud production environment at the destination end.	Solution 1: Use the destination Huawei Cloud production environment for testing.	1. After testing, the test environment is directly transitioned to production, saving workload. 2. All parameters have been adjusted to the optimal values during the test.	Network isolation is required; there are risks that may affect live-network services.
	Solution 2: Set up a new test environment on the target Huawei Cloud for testing.	No impact on the live network	1. Setting up an environment incurs certain costs. 2. The configuration parameters fine-tuned in the test environment need to be replicated exactly (1:1) into the production environment, involving some amount of work.
The destination Huawei Cloud production environment is a brand-new environment.	Use the destination Huawei Cloud production environment for testing.	1. After testing, the test environment is directly transitioned to production, saving workload. 2. All parameters have been adjusted to the optimal values during the test.	None

c. **Design test cases**: Design and prepare test cases based on testing objectives. The test cases before the switchover should achieve maximum coverage; during the switchover, due to limited testing time, it is advisable to categorize the test cases into three priority levels: P0, P1, and P2.

i. P0 definition: core function test cases. If the cases pass the test, rollback is not required.

ii. P1 definition: critical function test cases. If the cases pass the test, all fundamental functions are available. Once these test cases pass, the switchover is successful, allowing the cancellation of external maintenance notifications.

iii. P2 definition: supplementary test cases. If the switchover time window is sufficient, the cases can be tested on the night of the switchover. If the switchover time window is insufficient, the cases can be tested on the next day.

Table 6-31 Test case execution descriptions

Phase	Test Case	Coverage Rate
Pre-switchover test	All	Test all application functions and third-party integration functions. Special scenarios where tests cannot be performed need to be discussed separately to create a simulation test solution.
Test during the switchover	P0, P1, and P2 levels. At least P0 and P1 test cases must be tested in the switchover time window.	According to the switchover time window, if the time window is sufficient, all test cases are completed. If the time window is insufficient, at least P0 and P1 test cases are completed.

For selecting test cases in the test environment, enterprises need to assess whether testing conditions are met based on the application scenarios. For example, if using a third-party inventory synchronization test case where only the production environments of both parties can connect but not their respective test environments, this test case cannot be executed. It is essential then to identify such unfeasible test cases, evaluate the overall test case coverage, and discuss the simulation test method for test cases that cannot be covered. The table below lists the scenarios.

Scenario	Test Condition	Solution
Order placement in a third-party system	The third-party system cannot connect to the test environment, so testing cannot be conducted in the test environment.	For scenarios that cannot be tested, discuss mitigation solutions such as directly calling the inventory synchronization interface to simulate testing.

Scenario	Test Condition	Solution
Inventory synchronization	Third-party inventory cannot connect to the inventory system in the test environment, so testing cannot be conducted in the test environment.	For scenarios that cannot be tested, discuss mitigation solutions such as directly calling the inventory synchronization interface to simulate testing.
Payment	Online payment can be tested. Offline POS payment cannot be tested as the network cannot be connected to the test environment.	For scenarios that cannot be tested, discuss mitigation solutions such as directly calling interfaces to simulate testing.
And more	And more	And more

d. **Preset test data**: To ensure the authenticity and effectiveness of the tests, it is necessary to preset the test data in advance. You can use data from the source test environment or use masked production data.

e. **Execute test cases**: Some enterprises start test automation late, and a large number of test cases still need to be manually executed. During manual test case execution, the implementer needs to record related information such as the test time, testers, and test case execution results in detail. Some enterprises already have automated testing capabilities. During cloud migration, they only need to add new test cases to the automation platform for automatic execution.

f. **Output a test report**: After all test cases are tested, output a test report. In general, functional testing needs to ensure that the test environment is as consistent as possible with the production environment, with 100% test case coverage, to guarantee the proper functionality of applications after they are migrated to the cloud.

Performance Verification

- **Performance verification**

  After an application system is migrated to the cloud, the underlying technical components are replaced. The default parameters of the cloud technical components may differ from those of the source technical components, or the implementation mechanisms of the source and destination technical components may vary. As a result, performance issues may occur during cloud migration. Therefore, performance testing is required, which includes the following three categories:

Table 6-32 Performance testing content

Testing Content	Description
Cloud service performance	Perform performance tests on cloud services, such as databases, HBase, and storage system IOPS.
API performance	API performance is an aspect of system performance evaluation, with targeted stress tests conducted on specific APIs.
Overall application performance	Perform overall performance tests based on application scenarios, such as during a promotion when thousands of users simultaneously browse and purchase a product.

The following table lists the purposes of the three types of performance tests.

Table 6-33 Performance test purposes

Test Content	Purpose
Cloud service performance	Evaluate whether the specifications of cloud services meet the performance requirements under high concurrency and whether the parameters are set to the optimal values.
API performance	Evaluate the maximum load capability of specific APIs.

Test Content	Purpose
Overall application performance	1. Determine the maximum load capacity of the cloud-based business system: Through high concurrency and high load testing, determine the maximum load that the cloud-based business system can handle, along with its performance and response times at peak load. As the stress incrementally escalates, observe how the cloud-based business system performs under equivalent stress compared to the original setup, comparing gathered metrics to pinpoint potential issues. 2. Verify system stability and reliability: Perform long-term, high-load tests to verify the stability and reliability of the cloud-based business system across diverse scenarios, including system resource management, data transfer, and exception handling. 3. Assess system scalability: As the system load gradually increases, testing the scalability of the cloud-based business system determines whether it can expand to a larger scale and support more users and service requirements. 4. Identify system performance bottlenecks: Conduct stress tests on the cloud-based business system to identify system bottlenecks and determine performance issues arising from environmental changes post-migration, thus enabling targeted optimizations.

The methods for testing the preceding three types of performance are as follows:

1. **Cloud service performance testing (using databases as an example)**

   For most application systems, the bottleneck often lies in the database. While other components like network bandwidth, load balancers, application servers, and middleware can be scaled horizontally relatively easily, databases typically rely on primary-standby configurations due to high data consistency requirements, rather than adopting a distributed architecture.

   Common database-related metrics include:

   – **TPS/QPS**: Transactions per second and queries per second, used to measure database throughput.

   – **Response time**: includes the average response time, minimum response time, maximum response time, and time percentage. The time percentage is of significant reference value, such as the maximum response time of the first 95% of requests.

   – **Concurrency**: the number of query requests being processed simultaneously.

   – **Success rate**: the proportion of requests successfully returned within a specified period.

   Huawei Cloud RDS offers standard performance baselines like TPS and QPS, allowing enterprises to conduct stress tests tailored to their business data. A widely-used database stress testing tool is Sysbench, supporting multi-threading and various databases. The tests include:

- CPU performance
- Disk I/O performance
- Scheduler performance
- Memory allocation and transfer speeds
- POSIX thread performance
- Database performance (OLTP benchmark tests)

2. **API performance stress testing**

   You can use either of following methods to test the performance of APIs: one is using Huawei Cloud's cloud-native performance testing tool, CodeArts PerfTest; the other is using GoReplay. The advantages and disadvantages of both methods are listed in the table below.

Table 6-34 Comparison of API performance stress test methods

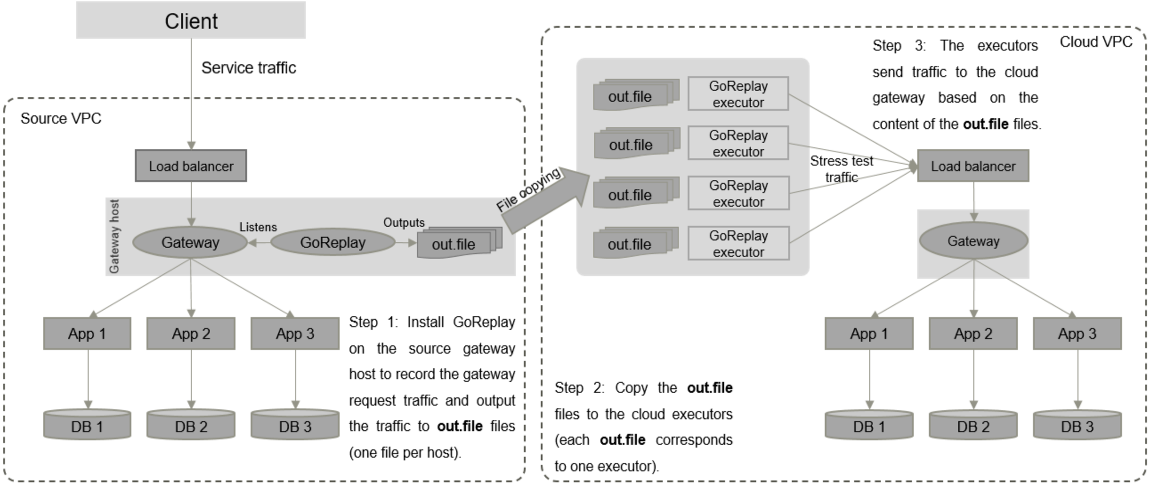
Test Tool	Stress Test Method	Advantage	Disadvantage
CodeArts PerfTest	Complete the API pressure test using the Huawei Cloud performance test tool.	● Supports tests in multi-protocol, high-concurrency, and complex scenarios. ● Provides professional performance test reports, making application performance clear at a glance. ● Does not interact with core services in the production environment, ensuring no impact on the live network. ● API tests have minimal dependency on other services, allowing them to be conducted based on a single service system.	● High execution costs due to extensive initial business analysis and script compilation. ● Requires highly skilled test personnel who are proficient in using test tools and possess relevant test knowledge; otherwise, the test results may be suboptimal.

Test Tool	Stress Test Method	Advantage	Disadvantage
GoReplay	Deploy GoReplay on the service gateway to replicate live network traffic and replay it at the destination end.	● Low cost and high efficiency: No need to sort out the interfaces and service logic of each system; testing can be conducted based on actual traffic. ● On the one hand, a large volume of live real traffic ensures comprehensive coverage. On the other hand, it supports intermediate process verification, such as comparing and verifying all objects in the content of sent messages and intermediate calculations, which is difficult to achieve through traditional manual verification point creation.	● Plugins must be installed on the traffic entry gateway in the production environment, consuming certain CPU and storage resources. ● In batch migration scenarios, since traffic recording is based on all service requests, if some services are not deployed at the destination end, it may lead to 404 errors, requiring manual fault localization, which is labor-intensive. ● Only HTTP traffic can be recorded; HTTPS, TCP, UDP, and other traffic cannot be captured.

For details about how to use CodeArts PerfTest to perform a performance stress test, refer to the **CodeArts PerfTest documentation**. The following describes how to perform performance stress tests through GoReplay traffic replication.

GoReplay is an open-source tool for replicating, replaying, and manipulating HTTP traffic. It captures real-time traffic and sends it to one or more target servers to achieve traffic replication and replay. Using GoReplay, actual HTTP request and response traffic can be replicated to test, development, or production environments for testing, monitoring, and analysis. During the lift and shift migration, we can use GoReplay to replicate request data from the traffic entry point of the live network's service gateway at the source end and replay the service requests on the target cloud executors to test the load on relevant service interfaces on the cloud. The detailed solution is shown in the following figure.

Figure 6-33 GoReplay traffic replication stress test solution



Step 1: Install GoReplay on the source gateway host to record the gateway request traffic and output the traffic to **out.file** files (one file per host).

Step 2: Copy the **out.file** files to the cloud executors (each **out.file** corresponds to one executor).

Step 3: The executors send traffic to the cloud gateway based on the content of the **out.file** files.

When using GoReplay for performance stress testing, pay attention to the following points:

- When GoReplay is used to record request traffic on the source gateway, monitor its impact on host performance and observe relevant host metrics in real time, such as CPU usage and memory usage. Additionally, the **out.file** files generated by GoReplay consume a significant amount of disk space. Monitor disk usage to prevent the gateway application from becoming unavailable due to full disk space. It is advisable to use network storage for storing output files.
- During traffic replay, if the destination service requires access to a third-party interface, it may impact production services. Ensure proper network isolation is in place.

3. **Overall application performance stress testing**

   Overall application performance stress testing involves comprehensive stress testing of all business processes and functions of the service system to assess its stability and performance in real production environments. During the testing, real user behaviors are simulated, and high-load scenarios are generated to evaluate the system's performance and stability under heavy loads, confirming that the service system meets actual user needs.

   Common scenarios for overall application performance stress testing include:

   a. Normal service load: Simulates the system's business load under normal usage conditions, including the quantity, frequency, and types of user requests. Verifies the system's performance under normal load to ensure it meets user requirements.

   b. Peak load: Simulates the highest load the system may face, typically during periods when user requests reach their peak. This scenario determines if the system's scalability can handle peak-hour requests and ensures no performance bottlenecks or crashes occur.

   c. Burst load: Simulates exceptional situations, such as sudden surges in user requests or large-scale data processing. Evaluates the system's stability and fault tolerance under sudden increased stress, ensuring graceful handling of abnormal loads.

   d. Long-term load: Simulates extended system operation, often lasting several hours or more. Tests for issues like memory leaks and resource depletion after prolonged operation to ensure system stability and reliability.

e. Exceptional scenarios: Simulates various system anomalies, such as network failures, server outages, and database disconnections. Tests the system's fault tolerance and recovery capabilities under these conditions to ensure correct exception handling and continued availability.

6.4.8 Cutover

6.4.8.1 Mock Cutover

Once the function and performance tests are completed, and the cloud applications and services run properly, the service cutover can be initiated.

Service cutover is the process of transitioning service operations from the source legacy system to the new cloud platform. It requires meticulous planning and coordination to ensure data integrity and service stability during the cutover.

In most cases, service cutover takes a certain period of time. In this period, requests are redirected to the cloud platform and the legacy service system will be phased out. This process helps achieve a smooth transition between the old system and the new system, and finally migrate services to the cloud. Therefore, it is common to conduct a mock cutover before real cutover.

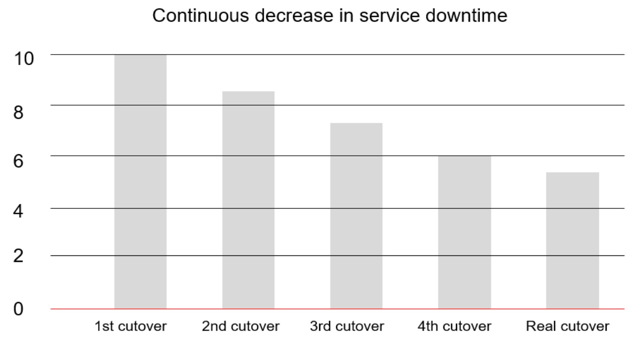
Why Is Mock Cutover Needed?

A mock cutover is critical to cloud migration. It can boost confidence and assurance for the real cutover. To be specific, it mainly has the following benefits:

1. **Identifying issues and risks:** The mock cutover helps uncover, prevent, and resolve problems, such as an incorrect startup/shutdown sequence of applications and batch tasks, network configuration issues, and data inconsistency issues.
2. **Enhancing team cooperation:** The mock cutover enables technical team members to get familiar with the entire cutover process and steps so that they can work together more efficiently.
3. **Optimizing the runbook:** The mock cutover helps find out problems in the cutover workflow, such as the overall serial/parallel sequence errors and the execution timeout for certain steps. Thus, the team can refine the runbook steps and timelines to improve the accuracy and feasibility for the real cutover while boosting efficiency.
4. **Estimating the cutover duration:** The mock cutover allows the team to know the start time, end time, and execution duration of each step. This enables them to precisely forecast the total duration of the real cutover. Then they can better schedule the service downtime announcement and coordination with other supporting teams.
5. **Minimizing downtime:** A large-scale system cutover often involves more than 200 steps, requiring parallel and serial operations and collaboration among multiple roles and personnel. The Mock cutover can improve the team's familiarity with cutover operations, coordination between parties, and problem-solving efficiency. For steps that take a long time, automatic scripts can be used to replace manual operations or further script optimization can

be conducted to reduce the downtime caused by the cutover. For example, a solution where all service systems are cut over at a time was adopted for the cloud migration of a large retail platform. After four mock cutovers, the cutover time was shortened by 40%.

Figure 6-34 Cutover results

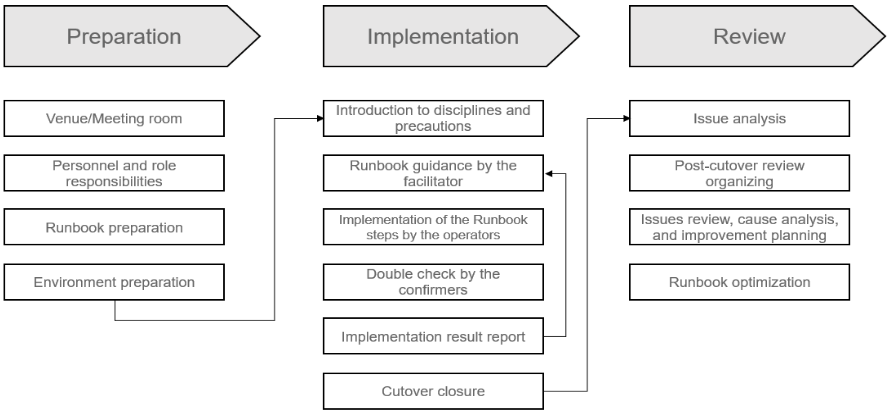


6. **Identifying unforeseen problems:** The mock cutover environment can help reveal hidden problems. For example, during a cutover, despite all applications involved in a system being shut down, some database sessions remain active and the data cannot be static. It is found that a third-party store still performs some operations after system shutdown. Such findings allow enterprises to adjust and optimize the cutover plan to ensure a successful real cutover.

Mock Cutover Process

It is recommended that two or three mock cutovers be performed before the real cutover. The mock cutover process is as follows:

Figure 6-35 Mock cutover process


Preparation
Implementation
Review
Venue/Meeting room
Personnel and role responsibilities
Runbook preparation
Environment preparation
Introduction to disciplines and precautions
Runbook guidance by the facilitator
Implementation of the Runbook steps by the operators
Double check by the confirmers
Implementation result report
Cutover closure
Issue analysis
Post-cutover review organizing
Issues review, cause analysis, and improvement planning
Runbook optimization

1. **Cutover preparations**

   The preparations for a mock cutover include the following aspects:

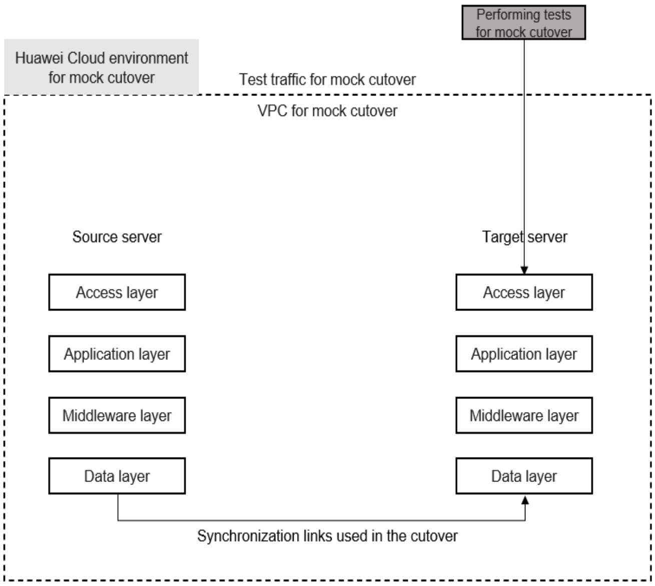
2. **Venue/Meeting room:** Determine the cutover location (floor) and book a meeting room in advance.

3. **Personnel and role responsibilities:** Clearly define the roles and responsibilities of all participants. Ensure that mock cutover participants and the real cutover operators are consistent. Roles can be defined based on the runbook roles in **Designing a Runbook**.
4. **Runbook preparation:** Refine the mock cutover runbook based on the cutover environment. It is best to conduct multiple rounds of review before finalizing the runbook.
5. **Environment preparation:** Prepare the source environment and target environment for the cutover. The source environment should mirror the production environment as closely as possible, including data consistency. It is recommended to preconfigure data and check the environment in advance, along with other preparatory steps.

   There are two solutions for preparing the cutover environment. Enterprises can select one of them based on their actual situations.

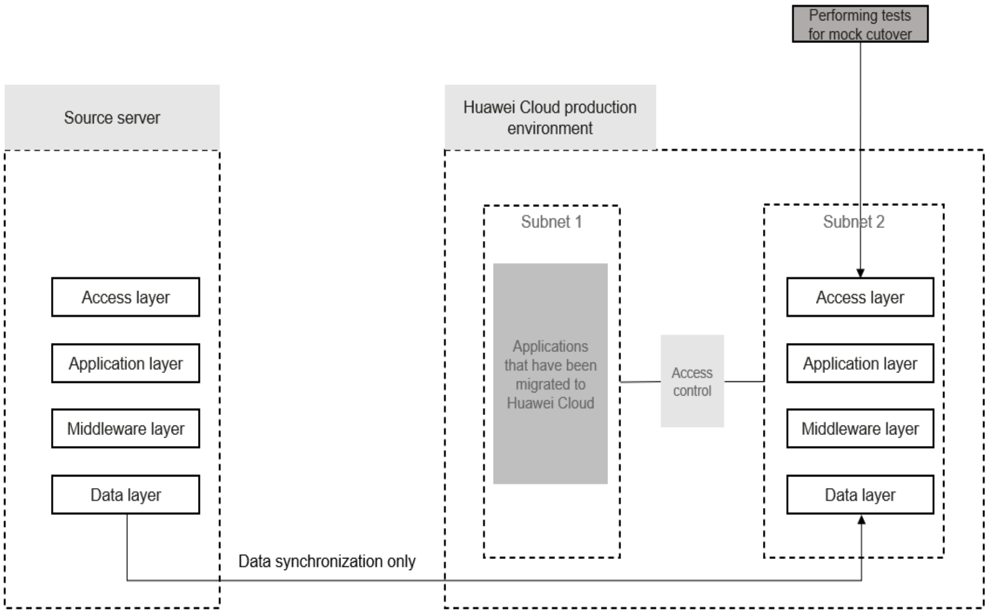
   a. If no test environment that is similar to the production environment is available in enterprises, deploy both the complete source and target environments on Huawei Cloud for the mock cutover.

   Figure 6-36 Cutover environment

   b. If a test or pre-production environment is available for use as the source environment in enterprises, deploy only the target environment on Huawei Cloud. The production environment on Huawei Cloud may be reused as the target environment. In this case, strict isolation measures must be implemented to prevent the mock cutover from affecting the live production system.

Figure 6-37 Cutover environment



6. **Cutover implementation & review**

- **The mock cutover is performed as follows:**
  a. The facilitator announces the cutover disciplines and precautions.
  b. The facilitator introduces the tasks, operators, and confirmers according to the runbook. (Note: Multiple tasks may be executed concurrently.)
  c. The operators perform the assigned tasks according to the runbook.
  d. The confirmers conduct a double check after the operators have completed their tasks.
  e. Once confirmed, the confirmers immediately report the result to the facilitator. (Note: For steps involving multiple confirmers, each confirmer should update their completion status in the online shared document for real-time progress tracking).
  f. Repeat the above steps until all tasks are completed.
  g. Operators and confirmers record the issues encountered during the mock cutover for a review.
- **The cutover review is performed in the following procedure:**
  a. Classify the recorded technical and organizational issues.
  b. Review each issue, analyze the root cause, and discuss improvement measures.
  c. Assign improvement measures to responsible parties and update the measures to the runbook for next mock cutover or real cutover.

6.4.8.2 Real Cutover

The organization, preparations, and roles and responsibilities of the real cutover are largely the same as those of the mock cutover. Nevertheless, during the real cutover, operations must strictly follow the real cutover runbook. The cutover solution and runbook vary from service system to service system. The following cutover steps are for reference only.

Step 1 **Pre-cutover preparations and checks**

Before the cutover, complete all pre-cutover preparations and checks as specified in the runbook checklist. The runbook preparation and check steps vary according to the service system. The following steps are for reference only.

Table 6-35 Preparations and check items before the cutover

Phase	Task	Department	Activity	Involved or Not	Completed or Not
Organizational and assurance preparation	Confirming the cutover window	Enterprise project manager	Determine that the cutover window starts at MM, DD, YYYY HH:MM.	Yes	Yes
	Confirming the images and wording for the downtime notice	Enterprise project manager	Check that the images and wording of the downtime notice have been updated.	Yes	Yes
	Notifying related personnel to release the notice on the official website	Enterprise project manager	Send an email to related personnel for releasing the notice on the official website.	Yes	Yes
	Reserving a meeting room	Enterprise project manager	Reserve a meeting room.	Yes	Yes
	Notifying and checking the personnel involved in the cutover	Enterprise project manager	Check whether the cutover participants can attend the meeting.	Yes	No
		Enterprise project manager	Confirm the third-party personnel involved in the real cutover and their contact information.	Yes	Yes
		Enterprise project manager	Confirm that the personnel in the operations center are on duty during the cutover.	Yes	Yes

Phase	Task	Department	Activity	Involved or Not	Completed or Not
	Establishment of the enterprise's internal communication channel	Enterprise project manager	Establish an internal communication channel for the cutover.	Yes	Yes
	Backend assurance team establishment	Cloud vendor's project manager	Set up a co-managed cutover assurance team with the customer. Establish a communication channel to ensure seamless assurance coordination between the backend O&M and R&D teams.	Yes	Yes
Communication with the third-party/business segments	Items to be communicated with the third-party/business segments	Enterprise project manager	Communicate with the business segments about the impact of downtime and solutions.	Yes	No
		O&M team	In cases where a third party uses fixed IP addresses, check the steps for modifying the configurations.	Yes	No
Environment checklist check	Determining the cut-off date for release pause	R&D team	Determine the cut-off date for release pause.	Yes	No
	Application inventory checking and start/stop scripts refreshing	R&D team	Check the inventory of applications to be migrated to the cloud.	Yes	No
	Job inventory checking and script refreshing	R&D team	Check the latest job inventory.	Yes	No
		R&D team	Check whether the inventory in the script is up-to-date.	Yes	No

Phase	Task	Department	Activity	Involved or Not	Completed or Not
Environment check (including source and target environments, migration tasks, and execution scripts)	Basic cloud service check	O&M team	Check whether the test Wi-Fi provided by the O&M team is available.	Yes	No
		O&M team	Check whether there is an alarm for exceeding the maximum bandwidth of the Huawei Cloud Direct Connect for synchronization.	Yes	No
		O&M team	Perform routine status check.	Yes	No
		O&M team	Check the high availability of cloud services and check whether there is a single-AZ or single-node fault on the target cloud resources.	Yes	No
	Database check	Database-related departments	Check whether the Huawei Cloud database port is the same as that in the production environment.	Yes	No
		Database-related departments	Check whether the NTP clock settings are consistent.	Yes	No
		Database-related departments	Check whether the Redis data migration task is normal and that no error or alarm (including the rollback task) is reported.	Yes	No

Phase	Task	Department	Activity	Involved or Not	Completed or Not
		Database-related departments	Check whether the DRS-MySQL data migration task is in **Incremental migration in progress** state, with no error or alarm (including the rollback task) reported and the dynamic data comparison task configured.	Yes	No
		Database-related departments	Check whether the DRS-MongoDB data migration task is in the **Incremental migration in progress** state, with no error or alarm (including the rollback task) reported.	Yes	No
		Database-related departments	Check whether character sets are consistent between the source and target MySQL databases.	Yes	Yes
		Database-related departments	Check whether users are consistent between the source and target databases.	Yes	Yes
	Supporting system check	Big data-related departments	Change the address of the database for big data extraction to the address of the standby IDC database.	Yes	No
	Script check	O&M team	Store the application service startup script on the executor.	Yes	No
		O&M team	Store the application heartbeat check script on the executor.	Yes	No

Phase	Task	Department	Activity	Involved or Not	Completed or Not
	Log system check	O&M team	Check whether ELK can handle the massive amounts of logs generated when multiple applications are started at the same time.	Yes	No
	Alarm monitoring system check	O&M team	Check whether the monitoring system is normal.	Yes	Yes
	Disk cleanup	O&M team	Check the disk usage in the production environment and execute the script to clean up disk spaces in batches.	Yes	Yes
Preparations of the operation guide, tools, terminals, and the login platform	Updating the runbooks for all participants to the latest version	Project manager	Synchronize the latest production runbook address to all participants (including personnel of the business segments).	Yes	No
	Personnel preparation	Project manager	Organize the participants to review the overall cutover process and provide guidance on their own tasks.	Yes	No
	Personnel operation permissions check	ALL	Log in to the operation environment and check the operation permissions (such as the login system, OS, and operation GUI).	Yes	No
		Test team	Check whether ITSM can be logged in. Check whether issues of cloud migration projects can be recorded.	Yes	No
		ALL	Check whether the operators have the permission to operate the executor on the batch task platform on the day of real cutover.	Yes	No

Phase	Task	Department	Activity	Involved or Not	Completed or Not
	Operation terminal check	ALL	Check the laptops and environments on the evening before the mock or real cutover and ensure that they run properly (DBAs should have a dedicated network cable, and a large switch should be prepared in advance).	Yes	No
	Test client check	Test team	Clear the client and browser cache.	Yes	No

----End

Step 1 **Cutover Runbook**

Once preparations and checks are complete, perform the real cutover according to the steps in the runbook. Each task must strictly follow the instructions. Note that the cutover steps in the runbook may vary depending on the business system. The following steps are for reference only. Steps with the same sequence number are executed concurrently.

If there are many batch processing tasks and the cutover time window is limited, you can start the tasks based on the priority.

Table 6-36 Example cutover procedure

Task	Step	Subtask
Forwarding source service traffic to the maintenance notice page	1.1	Changing the status of a CMDB service to **Maintaining**
	1.2	Forwarding external access traffic to the maintenance announcement page
Stopping the scheduled tasks on the source server	2.1	Stopping the scheduled tasks on the source server
	2.1	Stopping scheduled database tasks on the source server
Stopping the application services and configuration center on the source server	3.1	Stopping *xxx* application services on the source server

	3.1	Stopping the configuration center on the source server
Migrating message queue data	4.1	Migrating MQ data
	4.2	Waiting and confirming that Kafka consumption is complete
Confirming that the data at the data layer on the source server is static	5.1	Confirming that the Redis data on the source server is static
	5.1	Confirming that the MySQL data on the source server is static
	5.1	Confirming that the MongoDB data on the source server is static
Data consistency check	6.1	Checking Redis data consistency and stopping synchronization tasks
	6.1	Checking MongoDB data consistency and stopping synchronization tasks
	6.1	Checking MySQL data consistency and stopping synchronization tasks
Modifying the private DNS resolution at the data layer	7.1	Resolving the private domain name for application-to-application access to the IP address of the application instance on Huawei Cloud
	7.1	Resolving the private domain name of the target Redis to the IP address of the Huawei Cloud instance
	7.1	Resolving the private domain name of the target MySQL to the IP address of the Huawei Cloud instance
	7.1	Resolving the private domain name of the target MongoDB to the IP address of the Huawei Cloud instance
	7.1	Changing the private domain name of the target MQ to the IP address of the Huawei Cloud instance
	7.1	Resolving the private domain name of the target Kafka to the IP address of the Huawei Cloud instance

Starting the configuration center and scheduled task scheduling service, registering jobs, and enabling Kafka consumption	8.1	Starting the configuration center
	8.2	Starting the scheduled task scheduling service
	8.3	Releasing configurations in the configuration center in batches (registering jobs)
	8.3	Releasing configurations in the configuration center in batches (enabling Kafka consumption)
	8.4	Checking whether the configuration center and scheduled task scheduling service are enabled.
Intranet downtime notice revocation, and application startup and checks on the target server	9.1	Starting the target MQ
	9.2	Starting *xxx* application services on the target server
	9.3	Heartbeat check
	9.4	Basic service check
	9.5	Taking down the intranet maintenance announcement page
Starting the scheduled tasks on the target database and the scheduled tasks with the highest priority	10.1	Starting scheduled database tasks
	10.1	Starting the first batch of batch processing tasks on the target server
Main process test (P0 case)	11.1	Main process test (P0 case) ● Verify that the application is running properly in the target cloud environment. ● Verify that core functions and key service processes are the same as those before the migration. ● Monitor logs and metrics to ensure that the system is running properly.
Internet downtime notice revocation	12.1	Taking down the Internet maintenance announcement page
Starting the second batch of batch processing tasks on the target server	13.1	Starting the second batch of batch processing tasks on the target server
	13.2	

P1 service verification (After the job is started, verify the P1 case.)	14.1	Verifying the service functions of the target server
Starting the third batch of batch processing tasks on the target server	15.1	Starting the third batch of batch processing tasks on the target server
Starting the second batch of batch processing tasks on the target server	13.1	Starting the second batch of batch processing tasks on the target server

----End

6.4.9 Assurance

In the assurance phase of cloud migration, you need to perform the following tasks to ensure a smooth transition to the new cloud environment:

- **Cloud platform monitoring:** Establish an effective monitoring system to monitor the performance, availability, and security of the cloud platform. Set an alarm mechanism to detect and resolve potential problems in a timely manner.
- **System monitoring and O&M:** Set system monitoring and alarms to detect and resolve potential problems in a timely manner. Deploy infrastructure monitoring tools to monitor key indicators of servers, storage devices, and networks, and ensure that the log recording and error alarm mechanisms work properly.
- **Security check and bug fixing:** Check for potential vulnerabilities or weaknesses and take appropriate remedial measures to enhance security. Update and patch the system and software to ensure that the components and versions in use are up-to-date and that security patches are applied in a timely manner.
- **Backup and DR policies:** Evaluate and set new backup and DR policies to ensure data security and recoverability. Perform regular backup and DR drills to verify backup availability and recovery procedures.
- **Optimization and adjustment:** Optimize and adjust the system and applications based on the running status. Identify bottlenecks and performance problems by monitoring performance indicators, and adjust and optimize the system accordingly to improve system stability and responsiveness.
- **Training and support:** Provide necessary training and support for the O&M team to ensure that they are familiar with the new cloud environment and tools.
- **Document output:** Record and maintain the documents for future reference and archiving.

6.5 Big Data Migration

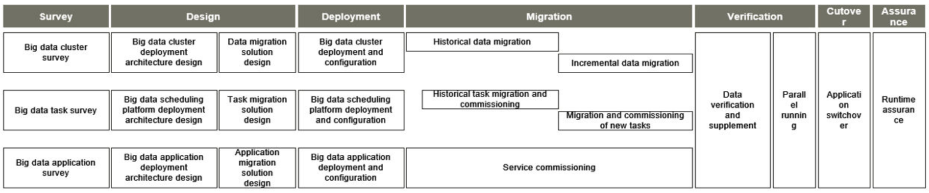
6.5.1 Survey

Big data migration encompasses the process of transferring big data clusters, task scheduling platforms, and applications from one operational environment to another. This process is structured into the following three modules. This section specifically details the migration of big data clusters and task scheduling platforms. For comprehensive information regarding big data application migration, refer to **Application Cloud Migration**. Note that this section primarily highlights the differentiating aspects of each migration type.

- **Big data cluster migration**: This involves the relocation of big data clusters, including their storage, compute, and management components, to a new operating environment. This process necessitates cluster reconfiguration and data migration, taking into account critical factors such as the data migration methodology, network throughput, system compatibility, and data consistency.
- **Big data task scheduling migration**: This includes the migration of existing big data task scheduling systems, workflows, and scheduling policies to a new operating environment. This entails a thorough assessment of task dependencies, task adaptation and reconstruction (if required), performance optimization, deployment procedures, testing protocols, and verification processes.
- **Big data application migration**: This refers to the migration of individual big data applications from one operating environment to another.

  The big data migration process is as follows:

Figure 6-38 Big data migration process


Survey
Design
Deployment
Migration
Verification
Cutove r
Assura nce
Big data cluster survey
Big data cluster deployment architecture design
Data migration solution design
Big data cluster deployment and configuration
Historical data migration
Incremental data migration
Big data task survey
Big data scheduling platform deployment architecture design
Task migration solution design
Big data scheduling platform deployment and configuration
Historical task migration and commissioning
Migration and commissioning of new tasks
Big data application survey
Big data application deployment architecture design
Application migration solution design
Big data application deployment and configuration
Service commissioning
Data verification and supplement
Parall el runnin g
Applicati on switchov er
Runtime assuran ce

For details about how to migrate big data applications, see **Application Cloud Migration**. This section describes only the special precautions for migrating big data applications.

The following outlines each phase of the big data migration process:

1. **Survey**: Conduct a comprehensive assessment of the existing big data platform, detailing its current version, configuration specifications, resource quantities, data types, data volume, and the types and number of associated tasks.
2. **Design**: Design the big data deployment architecture, data migration solution, task migration solution, and data verification solution.
3. **Deployment**: Deploy the big data platform, including cluster deployment and task scheduling platform deployment.
4. **Migration**: Migrate data and tasks.
5. **Verification**: Verify data and tasks.

6. **Switchover**: Switch over the big data application.
7. **Assurance**: Perform real-time monitoring and special O&M assurance for a certain period of time after the service cutover.

Refer to the survey methods in **Big Data Survey** to survey the status of the big data cluster, big data task scheduling platform, and big data applications.

6.5.2 Design

For details about the deployment architecture design for big data platforms on the cloud, see **Designing a Big Data Architecture**. This section focuses on the design of the data migration solution and task migration solution.

Data Migration Solution Design

Big data migration encompasses three data types, as detailed in the table below.

Table 6-37 Three types of data involved in big data migration

Category	Description
Metadata	Hive metadata or external metadata
Inventory data	Historical data that does not change in a short period of time
Incremental data	Data that is updated periodically.

The migration methods of the three types of data are as follows:

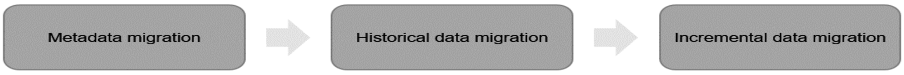
Table 6-38 Migration methods of different types of big data

Data Categorization		Migration Method
Metadata	Hive metadata	Export the Hive metadata from the source end and import it to Huawei Cloud MRS-Hive.
	External metadata MySQL	Use Huawei Cloud DRS to synchronize metadata from MySQL to RDS on the cloud.
Inventory data	Hive historical data is stored in HDFS.	Use Huawei Cloud CDM to migrate all historical data to Huawei Cloud MRS or Huawei Cloud OBS (decoupled storage and compute scenario).
	HBase historical data	1. Use Huawei Cloud CDM to migrate all historical data to Huawei Cloud MRS. 2. Use the HBase snapshot mode to migrate HBase data to Huawei Cloud MRS.

Data Categorization		Migration Method
Incremental data	Hive incremental data	Query the daily changed data based on the source metadata, identify the data directories to be migrated, and use Huawei Cloud CDM to migrate incremental data to the cloud.
	HBase incremental data	Use Huawei Cloud CDM to migrate all incremental data (based on timestamps) to Huawei Cloud MRS.

Enterprises can select the most suitable migration solution based on the specific data types involved. CDM serves as the primary tool during the data migration phase. Big data migration is usually performed in the following sequence:

Figure 6-39 Big data migration sequence



- **Metadata migration**

  Metadata migration is the initial phase. Metadata provides descriptive information about the data, such data structures, data definitions, and data relationships. This process involves exporting the source metadata and subsequently recreating or importing it into the target system. Successful metadata migration is crucial for ensuring the target system can accurately interpret and process the migrated data.

- **Historical data migration**

  Following the completion of metadata migration, the historical data migration phase involves transferring data accumulated over a specific past timeframe. This historical data is migrated to the target system to facilitate subsequent analysis and processing. The process typically includes exporting data from the original storage and loading it into the target system according to predefined rules and formats.

- **Incremental data migration**

  Upon completion of historical data migration, the incremental data migration phase addresses the transfer of new data generated since the historical migration. This incremental data requires timely and accurate migration to the target system, often performed in near real-time or at scheduled intervals. Common techniques for incremental data migration include data synchronization and continuous data transmission, ensuring the target system has access to the latest information.

Data Verification Standard Design

During big data migration, achieving 100% data consistency across all data types may not always be a strict requirement. Instead, data consistency needs should be determined based on specific service requirements and the importance of the data. Consequently, appropriate data migration policies and technical measures must be implemented to guarantee data correctness and integrity according to these defined requirements.

1. **Verification standards based on data type**: For transactional data, such as banking transaction records, stringent data consistency during migration is paramount. Post-migration, the source and target data must exhibit an exact match to prevent operational issues arising from data discrepancies. For data that is not transactional in nature, minor discrepancies after migration may be acceptable.
2. **Verification standards based on data importance**: Data crucial to core business operations demands high data consistency during migration. This data often contains sensitive or critical information, necessitating the utmost accuracy and integrity throughout the migration process. For data that does not directly impact core services, a greater tolerance for minor data differences post-migration may be permissible.

Therefore, prior to initiating data migration, organizations must establish clear and specific verification standards for each data type and based on its importance to the business. The following template can be used as a reference:

Table 6-39 Data types and verification standards

Data Type	Verification Standard	Table Name
Class X data	100% consistency	Table A, Table B, Table C,...
Class Y data	Error < 0.01%	Table D, Table E, Table F,...
...	Customized standard	...

Task Migration Solution Design

Big data tasks are classified into three types: JAR tasks, SQL tasks, and script tasks (Python and Shell). You can select a proper migration solution based on the task type.

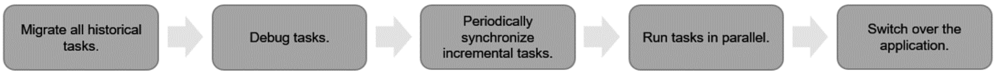
Table 6-40 Task types and migration solutions

Task Type	Migration Solution
JAR tasks	● All JAR tasks must undergo recompilation specifically targeting the cluster version in the cloud environment.

Task Type	Migration Solution
SQL tasks	● Same major version: If the source and target Hive environments share the same major version, SQL tasks can be migrated in parallel. ● Different major versions: In scenarios where the Hive major versions differ, SQL tasks should still be migrated in parallel. However, a subsequent fine-tuning phase is required to adjust SQL statements based on version-specific syntax changes, ensuring compatibility with the cloud environment's syntax.
Script tasks (Python, Shell)	● Same scheduling platform: When the source and target environments utilize the same scheduling platform, script tasks (Python, Shell) can be migrated in parallel. ● Different scheduling platforms: If the scheduling platforms differ between the source and target environments, script tasks can still be migrated in parallel. However, a critical step involves adapting and potentially reconstructing the scripts to align with the functionalities and syntax of the scheduling platform in the cloud environment.

Big data task migration is usually performed in the following sequence:

Figure 6-40 Big data task migration sequence



Migrate all historical tasks.
Debug tasks.
Periodically synchronize incremental tasks.
Run tasks in parallel.
Switch over the application.

1. **Migrate all historical tasks.**

   Initiate the migration of all historical tasks by transferring both the associated data and code to the new big data platform. This involves exporting data from the legacy storage system and subsequently loading it into the new storage system. Furthermore, the original task scripts and their corresponding configuration files must be adapted and migrated to ensure compatibility with the new computing environment.

2. **Debug tasks.**

   Upon completion of the full historical task migration, a commissioning and verification phase is crucial. This includes executing the migrated jobs, meticulously checking the output results against expected outcomes, and rigorously verifying performance and stability during task execution. Any identified issues or exceptions necessitate appropriate adjustments and remediation.

3. **Periodically synchronize incremental tasks.**

   Following the successful migration and commissioning of historical tasks, the migration and synchronization of incremental tasks will commence. Incremental tasks represent new workloads that require periodic execution during the ongoing migration process.

4. **Run tasks in parallel.**

   After incremental jobs have been migrated and synchronized, a parallel execution phase is initiated across both the original and new platforms. This involves simultaneously running jobs from both the legacy and the new big data systems on the new platform. The purpose of this parallel run is to verify the consistency of results between the two systems. This verification process entails comparing job outputs, logs, and key performance indicators to definitively determine the alignment of the new system's results with those of the original system.

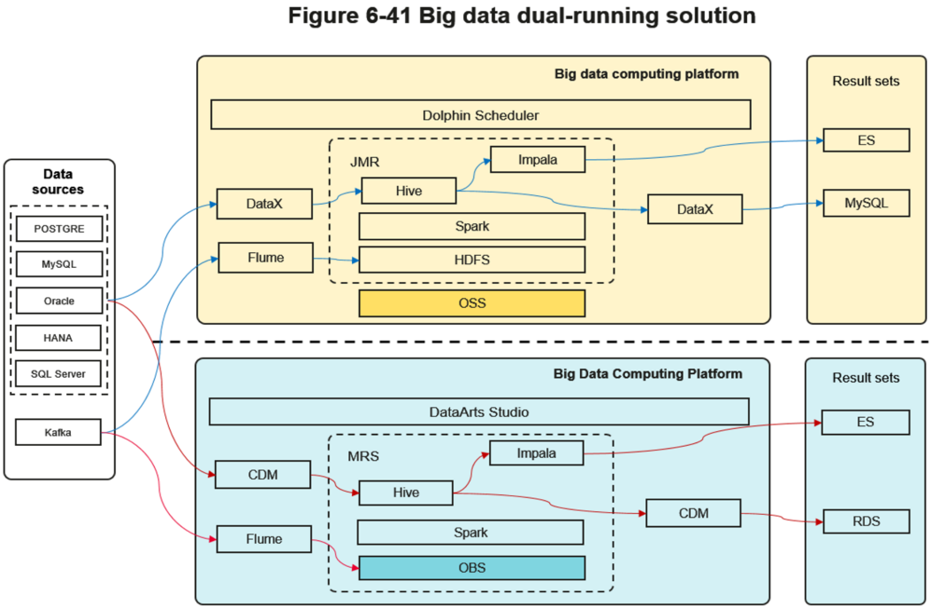
5. **Switchover the applications.**

   Once the parallel execution phase demonstrates stable and consistent performance over a defined period, the migration of big data applications and the complete cutover of all services to the new big data platform can be performed.

6. **Design a big data parallel running solution.**

   Implementing a parallel running solution is a prevalent and effective big data migration strategy. This approach involves operating both the legacy and the new platforms concurrently to facilitate continuous data and task verification. After a period of stable parallel operation, all services are transitioned to the new big data platform, as illustrated in the following figure.

   Figure 6-41 Big data parallel running solution

   The design ideas of the parallel running solution are as follows:

   a. **Data and task migration**

      Prior to establishing data source connectivity, perform a comprehensive migration of data and tasks. This includes migrating historical data from the original platform to the new big data platform, as well as transferring all associated task code, scripts, and configuration files. The selection of appropriate migration tools and methodologies should be based on specific project requirements, potentially leveraging offline data transmission utilities and dedicated big data migration tools such as CDM.

b. **Data source access**

Establish connectivity between the target big data cluster and the same underlying data sources utilized by the original big data cluster. This ensures data source consistency between the two environments. Employ data synchronization tools, ETL (Extract, Transform, Load) tools, or custom-developed scripts to facilitate data source connection and synchronization. For offline computing tasks, data sources can be accessed via data synchronization tools like CDM, ETL tools, or custom scripts. For real-time computing tasks, consider utilizing Kafka MirrorMaker and Nginx traffic mirroring configurations to replicate real-time data streams to both platforms concurrently.

c. **Parallel running**

Initiate the simultaneous operation of both the target and original big data clusters and their respective task scheduling platforms. During this parallel running phase, both platforms will process workloads concurrently, generating independent sets of results.

d. **Running stability verification**

Throughout the parallel running period, continuous monitoring and verification of task execution stability and data consistency on the target big data platform are essential. This includes actively tracking task execution status and meticulously comparing task logs and output results. Any identified issues or anomalies must be addressed and rectified promptly to ensure the reliability of the new platform.

e. **Official service switchover**

Upon thorough confirmation of the target big data cluster and task scheduling platform's stability, coupled with the verified integrity and accuracy of the migrated data and tasks, proceed with the official service switchover. This entails redirecting all service traffic and job execution to the target big data platform and decommissioning the original big data cluster and task scheduling platform.

6.5.3 Deployment

Big Data Platform Deployment

You can deploy the big data platform as follows:

- **Big data cluster deployment**

  Following the established architecture design principles, big data clusters in the cloud environment are typically provisioned utilizing cloud-based services. Huawei Cloud MapReduce Service (MRS) offers a streamlined solution for deploying and managing Hadoop ecosystems on Huawei Cloud. This service enables the one-click deployment of fully controllable, enterprise-grade big data clusters. Within these clusters, users can readily operate key big data components such as Hadoop, Spark, HBase, and Kafka.

  For details, see the **MRS official website**.

- **Big data task scheduling platform deployment**

If the target architecture incorporates Huawei Cloud DataArts Studio for task scheduling, the platform can be deployed and configured according to the guidelines provided on its **official website**.

For target architectures employing a self-built big data task scheduling platform, the scheduling software can be deployed on Huawei Cloud ECSs. Alternatively, Huawei Cloud SMS can be utilized to migrate an existing on-premises scheduling platform to a Huawei Cloud ECS instance.

- **Big data application deployment**

  Big data applications can be deployed on Huawei Cloud ECSs. Alternatively, Huawei Cloud SMS can be employed to migrate existing big data applications to Huawei Cloud ECS instances.

Platform Permission Configuration

- **Platform permission configuration**

  Following the deployment of the target big data platform, ensuring accurate and consistent permission settings is crucial. To achieve this, you can reference the permission configurations of the source platform and implement the following steps:

- **Reviewing source permission settings**

  Conduct a thorough review of the source platform's permission settings. This includes a detailed examination of users, roles, organizational structures, and associated permission levels. The objective is to fully understand the scope of permissions and access rights granted to each user, enabling the accurate replication of these settings on the target platform.

- **Creating users and roles**

  Based on the reviewed source platform permissions, create corresponding users and roles on the target platform. It is imperative to ensure that user identities and role assignments on the target platform precisely mirror those on the source platform.

- **Adjusting permission levels and scopes**

  On the target platform, meticulously adjust permission levels and scopes to align with the configurations identified on the source platform. The goal is to establish identical permission settings on the target platform, guaranteeing that users can only access the resources they are authorized to view and interact with.

- **Assigning and inheriting permissions**

  Implement permission assignment and inheritance mechanisms on the target platform based on the source platform's permission structure. This ensures that users inherit the same permissions and retain their corresponding roles and permission settings in the new environment.

- **Reviewing and adjusting access control**

  Review the access control mechanisms implemented on the target platform and make necessary adjustments to align with the source platform's permission settings. The objective is to ensure that access controls effectively restrict user access to appropriate resources, adhering to the permission rules established on the source platform.

- **Implementing security audit and monitoring**

Establish comprehensive security audit and monitoring mechanisms to ensure the ongoing and effective oversight of permission settings on the target platform. This proactive approach facilitates the detection and prevention of unauthorized access attempts, enabling timely and appropriate incident response.

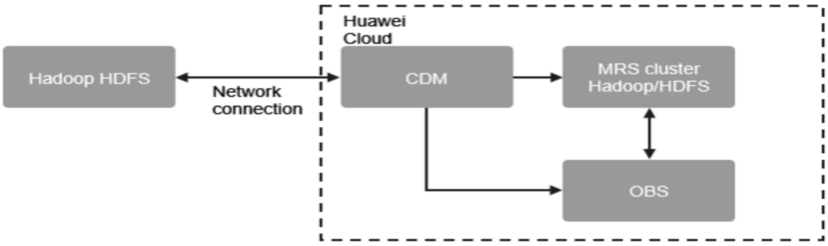
6.5.4 Migration

Data Migration

1. **Migrating Hadoop data to MRS**

   As shown in the following figure, data in the Hadoop cluster in the IDC or other public clouds is migrated to MRS. For details, see **MRS Help Center**.

   Figure 6-42 Migrating Hadoop data

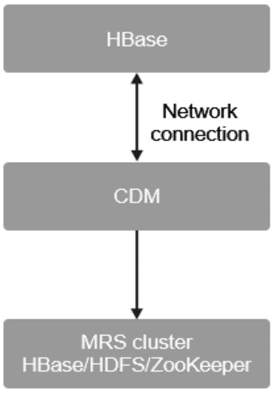


2. **Migrating HBase data to MRS**

   Migrate data in the HBase cluster in the IDC or other public clouds to MRS. HBase stores data in HDFS, including HFile and WAL files. The **hbase.rootdir** configuration item specifies the HDFS path. By default, data is stored in the **/hbase** folder on MRS. Some mechanisms and tool commands of HBase can also be used to migrate data. For example, you can migrate data by exporting snapshots, exporting/importing data, and CopyTable. For details, see the Apache official website.

   You can also use CDM to migrate HBase data. For details, see **MRS Help Center**.

   Figure 6-43 Migrating HBase data

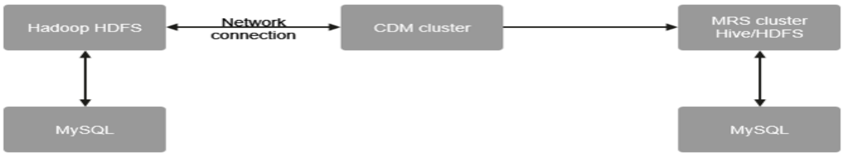


3. **Migrating Hive data to MRS**

   You can use CDM to migrate data in the Hive cluster in the IDC or other public clouds to MRS.

   For details, see **MRS Help Center**.

   Figure 6-44 Migrating Hive metadata

4. **Using BulkLoad to import data to HBase in batches**

   Organizations frequently encounter the need to ingest substantial volumes of data into HBase. While data can be loaded into HBase in batches by invoking the **put** method of HBase APIs or by leveraging MapReduce to load data from HDFS, these approaches can impose significant strain on RegionServers. This is due to the frequent flush, compaction, and split operations triggered by HBase, leading to high consumption of CPU and network resources and consequently, reduced efficiency. When operating within an MRS environment, utilizing the BulkLoad method for batch importing local data into HBase is strongly recommended. During the initial data loading phase, BulkLoad significantly enhances write efficiency and alleviates write pressure on the RegionServer nodes.

   For details, see **MRS Help Center**.

5. **Migrating MySQL data to a Hive partition table in an MRS cluster**

   Hive partitions are implemented by using the HDFS subdirectory function. Each subdirectory contains the column names and values of each partition. If there are multiple partitions, there are many HDFS subdirectories. It is not easy to load external data to each partition of the Hive table without using tools. With CDM, you can easily load data of the external data sources (relational databases, object storage services, and file system services) to Hive partitioned tables.

   For details, see **MRS Help Center**.

6. **Migrating data from MRS HDFS to OBS**

   CDM can migrate MRS HDFS data to OBS. For details, see **MRS Help Center**.

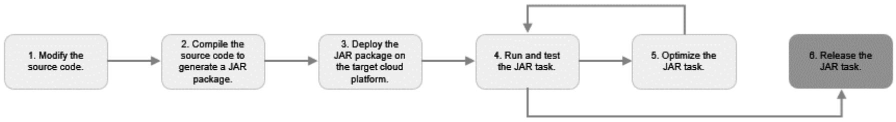
7. **Migrating tasks**

   Big data task migration involves the process of transferring big data workloads from one scheduling platform to another. This primarily encompasses JAR tasks, SQL tasks, and script tasks. The subsequent sections detail the migration procedures for these three task types.

8. **Migrating JAR tasks**

   To migrate JAR tasks effectively, a thorough understanding of the source code and dependency libraries of the original tasks is essential. The migration process necessitates recompiling the code to generate executable JAR files that are compatible with the target cloud environment. Following recompilation, comprehensive verification and optimization of these files are crucial to ensure proper functionality and performance within the new infrastructure. The steps are as follows:

Figure 6-45 JAR task migration process


1. Modify the source code.
2. Compile the source code to generate a JAR package.
3. Deploy the JAR package on the target cloud platform.
4. Run and test the JAR task.
5. Optimize the JAR task.
6. Release the JAR task.

Prerequisite: Ensure that the data dependencies required for JAR task debugging have been successfully migrated to the cloud environment. Refer to the "Data Migration" section for detailed information on the data migration process.

a. Adapt the source code to align with the big data resource configuration in the cloud environment. This includes adjusting parameters such as the version, dependency libraries, database connection strings, and library dependency configurations to match the cloud environment's specifications.

b. Compile the modified source code to generate a JAR package that is executable within the cloud environment.

c. Upload the generated JAR package to the task scheduling platform in the cloud. Subsequently, deploy and configure the JAR package.

d. Execute the scheduled task and monitor its execution status and results by reviewing the generated logs.

e. In the event that the task execution does not meet the expected outcomes (e.g., prolonged execution time), perform thorough root cause analysis. Based on the analysis, implement necessary optimizations and conduct further verification.

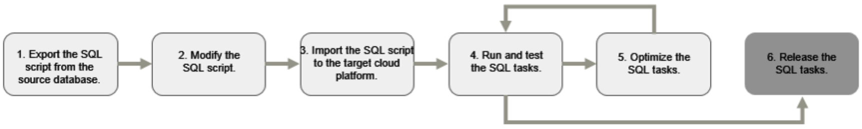
f. Configure the scheduling parameters of the task based on the defined service requirements.

If you use **DataArts Studio** of Huawei Cloud as the big data task scheduling platform, you can configure JAR tasks by referring to **DataArts Studio Help Center**.

9. **Migrating SQL tasks**

To migrate SQL tasks, you need to adapt and reconstruct SQL scripts. The following figure shows the migration process.

Figure 6-46 SQL task migration process


1. Export the SQL script from the source database.
2. Modify the SQL script.
3. Import the SQL script to the target cloud platform.
4. Run and test the SQL tasks.
5. Optimize the SQL tasks.
6. Release the SQL tasks.

Prerequisite: Ensure that the data dependencies required for SQL task debugging have been successfully migrated to the cloud environment. Refer to the "Data Migration" section for detailed information on the data migration process.

a. Export the source SQL script(s): Export the SQL script(s) from the source task scheduling platform.

b. Modify the SQL script(s): Adapt the exported SQL script(s) to align with the specific syntax and resource configuration of the target cloud scheduling platform.

c. Import the SQL script(s) to the target cloud platform: On the target cloud task scheduling platform, configure the necessary SQL tasks and import the modified SQL script(s).

d. Run and test SQL tasks: Execute the configured SQL scheduling tasks and thoroughly monitor their execution status and output by reviewing the generated logs and execution results.

e. Optimize SQL tasks: In the event that the task execution does not meet the expected outcomes (e.g., prolonged execution time), perform thorough root cause analysis. Based on the analysis, implement necessary optimizations and conduct further verification.

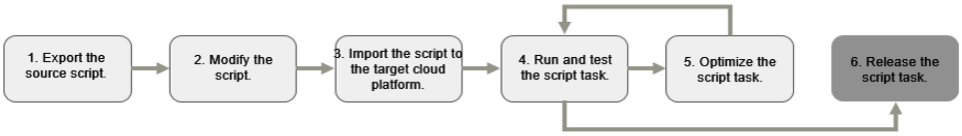
f. Release SQL tasks: Configure scheduling tasks as required by services and configure correct task dependencies.

If you use **DataArts Studio** of Huawei Cloud as the big data task scheduling platform, you can develop and configure SQL jobs by referring to **DataArts Studio Help Center**.

10. **Migrating script tasks (Python and Shell)**

When migrating script tasks, you also need to adapt to the cloud environment. The following figure shows the process.

Figure 6-47 Process of migrating script tasks


1. Export the source script.
2. Modify the script.
3. Import the script to the target cloud platform.
4. Run and test the script task.
5. Optimize the script task.
6. Release the script task.

Prerequisite: Ensure that the data dependencies required for script task debugging have been successfully migrated to the cloud environment. Refer to the "Data Migration" section for detailed information on the data migration process.

a. Export source scripts: Copy the executable scripts associated with the scheduling tasks from the source scheduling platform.

b. Modify scripts: Adapt the copied scripts to align with the cloud environment's configuration. This includes adjusting parameters such as database connection strings, resource allocations, and output directory paths to match the cloud infrastructure.

c. Import scripts to the target cloud platform: Upload the modified scripts to the target cloud scheduling platform and configure the corresponding script scheduling tasks within the platform.

d. Run and test script tasks: Execute the configured scheduling tasks and thoroughly monitor the script's execution status and output by reviewing the generated logs and execution results.

e. Optimize script tasks: In the event that the task execution does not meet the expected outcomes (e.g., prolonged execution time), perform thorough root cause analysis. Based on the analysis, implement necessary optimizations and conduct further verification.

f. Release script tasks: Configure scheduling tasks as required by services and configure correct task dependencies.

If you use **DataArts Studio** of Huawei Cloud as the big data task scheduling platform, you can develop and configure Shell and Python scripts by referring to **DataArts Studio Help Center**.

6.5.5 Verification

- **Verifying data**

  The database comparison methods include database content comparison, object comparison, and row comparison. The file comparison methods include file quantity comparison, size comparison, and content comparison. For details about data comparison methods, see **Data Verification**.

- **Verifying tasks**

  Following the migration of big data tasks, it is imperative to ensure that these workloads execute successfully, produce accurate outcomes, and meet defined performance benchmarks. Task verification generally involves verifying the following key aspects:

- **Verifying the task execution success rate**

  After the task migration is complete, a thorough verification of the migrated big data tasks is required. This includes executing tasks and meticulously assessing their execution success rate. During this verification process, close attention should be paid to task statuses, logs, and any encountered errors or exceptions. Tasks that fail to execute correctly necessitate detailed investigation and debugging to identify and rectify the underlying faults.

- **Verifying the consistency of task execution results**

  Verify the consistency of execution results generated by the big data tasks by comparing the output data of tasks running on both the new and legacy big data platforms. Consistency verification can be performed using specialized comparison tools, custom data verification scripts, or manual inspection. Data inconsistencies found may necessitate a review of data conversion processes, data formats, or data processing logic implemented during the migration, followed by appropriate rectification and adjustments.

- **Verifying the task execution performance**

  Following the migration, evaluate the task execution performance, including metrics such as runtime, resource utilization, and concurrency. By monitoring key execution and performance metrics, you can determine whether the post-migration task performance meets the established expectations. Identified performance anomalies may require adjustments to task configuration parameters, optimization of the task code, or reconsideration of resource allocation strategies.

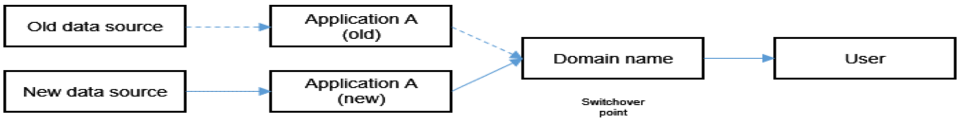
During the job verification process, leveraging monitoring tools, log analysis techniques, and data verification utilities is crucial to ensure the overall reliability and stability of the big data tasks after migration.

6.5.6 Switchover

Big data switchover refers to the transition of big data applications from a source environment to a target environment. Detailed information regarding switchover drills and formal switchover procedures can be found in **Cutover**. This section outlines three key switchover approaches for big data applications to provide comprehensive guidance during the transition process.

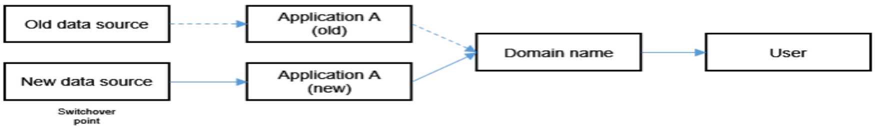
- **Parallel running scenario**: In a parallel running scenario, the big data application is deployed and actively running in both the source and target environments simultaneously. The switchover point in this approach is the domain name. During the service switchover, the transition is executed by simply redirecting the domain name resolution to point to the new application instance in the target environment, thereby shifting service traffic.

  Figure 6-48 Parallel running scenario

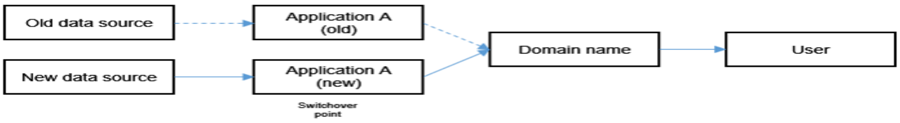
- **Data push scenario**: This scenario is applicable when the data source actively pushes data to the big data application. The switchover point here is the data source itself. The switchover process involves halting the data push from the original data source, configuring and initiating the data push from the new data source, and updating the application's data source configuration to point from the old data source to the new one.

  Figure 6-49 Data push scenario

- **Data extraction scenario**: This scenario applies when the big data application actively extracts data from a designated data source. The switchover point is the application itself. The switchover procedure entails stopping the application's data extraction from the original data source, configuring and starting the application to extract data from the new data source, and reconfiguring the application's data source connection to point from the old data source to the new one.

  Figure 6-50 Data extraction scenario

6.5.7 Assurance

During the assurance phase of big data migration, the following tasks are critical to ensure a seamless and stable transition to the new cloud environment:

- **Monitoring and alerting**: Implement a comprehensive real-time monitoring system to continuously track the operational status of clusters, task scheduling platforms, and applications. Configure robust alerting mechanisms to promptly detect potential issues and enable timely intervention and remediation.
- **Cluster performance optimization**: Conduct thorough evaluation and optimization of big data cluster performance. This includes diligently monitoring resource utilization, fine-tuning configuration parameters, and adjusting cluster size and resource allocation to maximize overall performance.
- **Data security and permission management**: Rigorously review and fortify data access control and permission management frameworks. Ensure that sensitive data is accessible only to authorized personnel and implement appropriate encryption and data masking techniques to safeguard data security.
- **Automatic task scheduling**: Verify the proper operation and scheduling of the big data task scheduling platform. Optimize scheduling policies to guarantee the timely and accurate completion of tasks, and establish robust mechanisms for handling potential faults or exceptions.
- **Exception handling and fault recovery**: Develop a comprehensive fault handling and recovery plan. This plan should include a clear classification of potential issues that may arise within clusters, tasks, and applications, along with well-defined response and recovery procedures for each scenario.
- **Team training and knowledge sharing**: Provide thorough training to team members to facilitate their adaptation to the new cloud environment and technology stack. Establish a proactive knowledge sharing mechanism to foster effective communication and the exchange of expertise within the team.

6.6 Application Modernization

6.6.1 What Is Application Modernization?

Migrating applications and data to the cloud is just the beginning of digital transformation. We need to continuously consolidate and optimize applications and data. We need to modernize applications to meet new IT and service requirements and support cloud-based service development. Migrating to cloud is just a task of porters and architects. However, we need to be an experience officer of the cloud and use new cloud technologies to continuously optimize service experience and support service innovation.

In recent years, major cloud service providers have proposed the vision of application modernization. In the digital era, enterprises can quickly respond to changes and achieve agile innovation, which will become a decisive factor for enterprises to build their sustainable competitiveness. Application modernization has become an inevitable choice for many enterprises to carry out digital transformation. Traditional application should evolve to application modernization. Application modernization should be considered based on application implementation and cloud platform capabilities. The cloud platform supports layered decoupling of application modernization. Applications focus on business logic and build common capabilities such as Design for X and governance on the cloud platform.

Figure 6-51 Development trend of modernized applications

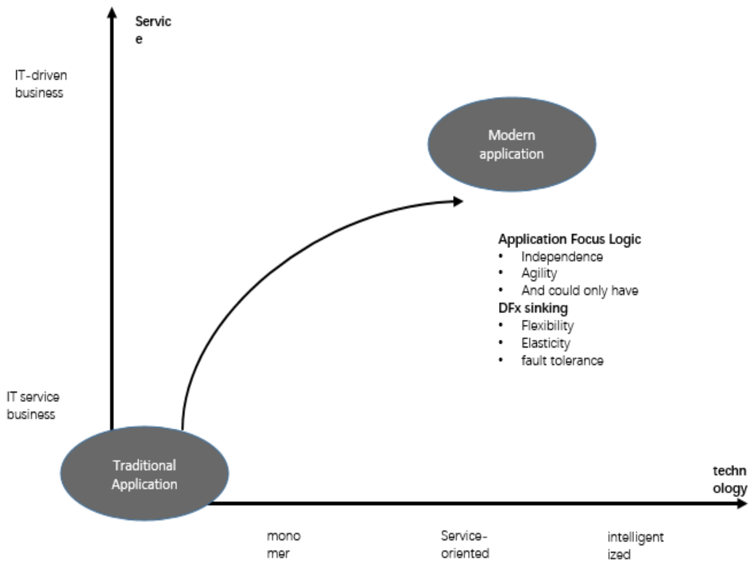


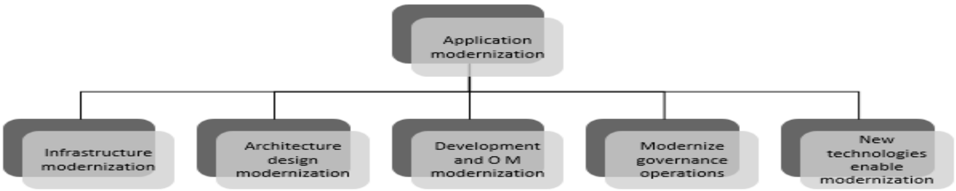
Table 6-41 Comparison between traditional and modern applications

Traditional applications	Modern applications
Monolithic architecture with highly-coupled modules	Microservice-based architecture, fully decoupled applications, and quick combination
Multiple application entries, affecting user experience	User-centric, one-stop personalized experience
Unable to quickly respond to new service changes	Quick combination and on-demand customization for new services
New functions are released with major versions, and the requirement delivery period is long (in years or months).	Fast iteration and rollout, shortening the delivery period (in weeks or days)

Traditional applications	Modern applications
Large team scale and traditional development mode	Split the team into smaller ones and agile operations with DevSecOps
Physical server	Container-based deployment and full cloudification

Application modernization not only uses cloud native technologies (such as containers, microservices, DevOps, and API gateway), but also new technologies (such as AI, virtual human, IoT, and blockchain). These technologies enable businesses to keep up with the trend of the era and improve user experience and innovation capabilities. Application modernization includes the following four aspects:

Figure 6-52 Four aspects of application modernization


Application modernization
Infrastructure modernization
Architecture design modernization
Development and O M modernization
Modernize governance operations
New technologies enable modernization

- **Infrastructure modernization** reduces costs and eases concerns. Cloud native reconstruction of traditional infrastructure delivers high availability and scalability, reduces O&M costs, and frees development and O&M personnel from manual resource allocation.
- **Architecture design modernization** decouples reusable functions from business logic. Microservice and serverless architectures split applications into modules that can be quickly and independently released, so that development and O&M personnel can focus on application innovation.
- **Development and O&M modernization** improves automation and security. Integrated development, O&M, and security capabilities, such as DevSecOps, provide intrinsic security and speed up application release.
- **Governance and operations modernization** promotes architecture evolution by integrating legacy and new assets in all domains. Through all-domain convergence and integration and unified governance and operations, an evolvable application architecture maximizes the value of legacy and new assets.

6.6.2 Infrastructure Modernization

Containerization is a process of migrating traditional applications or services to the containerized environment. The general procedure for containerization is as follows:

- **Evaluation and planning:** First, evaluate the features, dependencies, and architecture of applications or services. Determine which are suitable for containerization and make a reconstruction plan.
- **Container platform**: Select a container platform that meets your requirements. The most common containerized platform is Docker, but there are other options, such as Kubernetes.
- **Containerized application**: An application is split into small modules or microservices, and each module is packaged as an independent container image. Use Dockerfile to define the process of building container images, including dependency installation, configuration, and startup commands.
- **Container orchestration and management**: Container orchestration tools, such as Kubernetes, can be used to manage multiple container instances and implement functions such as automatic scaling and load balancing. You can define the deployment and running modes of containers by compiling configuration files or using command line tools.
- **Network and storage configuration**: Configure the network communication between containers and the mode for accessing external resources. Ensures that a container can interact with other containers, databases, and message queues, and ensures data durability and reliability.
- **Security and monitoring**: Ensure the security of the containerized environment by restricting container permissions, using secure image sources, and scanning vulnerabilities. In addition, a monitoring system is set to monitor the performance and running status of containers in real time.
- **Test and deployment**: After the containerization is complete, perform comprehensive tests: unit test, integration test, and performance test. Ensure that applications run properly in the container environment. Then, use automation tools or scripts to deploy the container to the production environment.
- **Continuous integration and delivery**: Establish a continuous integration (CI) and continuous delivery (CD) process to quickly and reliably build, test, and deploy new versions of containerized applications.

Containerization is a complex process and requires careful planning and evaluation. Before you start, you need to take a deep dive into the containerization technology and the platform you selected, and select the appropriate tools and methods as required.

6.6.3 Architecture Modernization

Microservice-based reconstruction and migration to the cloud

It is a complex process to conduct microservice-based reconstruction on a traditional monolithic application and migrate it to the cloud environment. The following describes the basic steps and considerations for microservice reconstruction and migration to cloud.

1. **Evaluate existing applications and objectives:**

   First, evaluate the traditional monolithic application to understand its architecture, functions, and performance. Then, specify the goals you want to achieve in the cloud environment, such as scalability, high availability, and flexibility. This evaluation phase helps you determine whether the application is suitable for microservice reconstruction and migration to the cloud.

2. **Decompose monolithic applications**

   Before microservice reconstruction, you need to decompose a monolithic application into smaller and independent functional modules. This process is often referred to as decomposing monoliths. By analyzing the service logic and functions of an application, you can identify the modules that can run independently and divide them into different microservices. Each microservice is responsible for specific service functions and should be loosely coupled and independent of each other as much as possible.

   You can use different rules to decompose an application, for example, decomposing based on service domain (domain-driven design) or decomposing based on functional module. Ensure that each microservice has clear responsibilities and defines the interaction mode between microservices through interfaces.

3. **Define service boundaries and interfaces**

   After decomposition, define the boundary and API of each microservice. Determine the external interfaces of each microservice and the communication methods between them (for example, using RESTful APIs or message queues). When defining interfaces, ensure that they are clear, consistent, and easy to use. This facilitates collaboration between teams and supports future expansion and changes.

   In addition, consider using open standards and protocols, such as the OpenAPI specification (formerly known as Swagger) to define interfaces. This simplifies the integration between microservices and facilitates document generation and code generation.

4. **Design and implement service governance design and implementation**

   In the microservice architecture, service governance becomes critical. You need to consider how to discover, register, configure, and monitor your microservices. Select service registration and discovery tools (such as Consul and Eureka) that meet your requirements and ensure that microservices can be effectively managed, monitored, and maintained throughout the service lifecycle. The service registration and discovery tool helps you automate the service registration and discovery process and provides functions such as service health check and load balancing.

   In addition, load balancing, fault recovery, and service security should be considered. The load balancing mechanism is used to balance request distribution to ensure that each microservice can process a proper amount of load. Implement a fault recovery mechanism (such as circuit breaker mode) to handle faults and prevent cascading faults. In addition, proper authorization and authentication mechanisms are used to protect microservice security and restrict access to sensitive data and functions.

5. **Containerization technology introduction**

   The microservice architecture is usually deployed and managed using the containerization technology. The most common method is to use Docker containers. Each microservice is packaged into an independent container for better isolation and deployment. Container orchestration tools (such as Kubernetes) are used to automate container deployment, scaling, and management, improving system scalability and elasticity. Containerization enables more flexible deployment and management of microservices. Containerization also helps to solve the consistency problem between the development environment and the production environment. The development team can use the same container to run microservices locally and ensure that the microservices run properly in the development and test phases. Then, these containerized microservice images are uploaded to the cloud platform for deployment and production.

6. **Data management and persistence**

   When converting monolithic application to a microservice, you need to consider data management and persistence. Each microservice may have its own database or share the same database. Select a database solution that meets your requirements and ensure data consistency and reliability. In the cloud environment, you can use hosted database services, such as Huawei Cloud RDS and GaussDB. In addition, you need to consider how to handle data transactions and data consistency across multiple microservices. A common method is to use a distributed transaction manager (such as the Saga mode) to ensure that data operations between microservices are consistent and atomic.

7. **Monitoring and logging**

   For the microservice architecture, it is important to implement comprehensive monitoring and logging. Use appropriate monitoring tools and log systems to collect and analyze metrics and logs of each microservice, as well as performance and fault information of the entire system. This helps you quickly detect and resolve potential problems and ensures system availability and stability. You can use the monitoring and log services provided by cloud providers, such as Huawei Cloud Eye and LTS, to centrally manage and analyze monitoring data and logs. In addition, the visualization and alarm mechanism enables the team to monitor the system running status in real time and take measures in a timely manner when an exception occurs.

8. **Automatic deployment and CI/CD**

   The microservice architecture usually needs to be frequently deployed and updated. To simplify and accelerate the deployment process, automatic deployment and continuous integration/continuous delivery (CI/CD) processes can be introduced. Use appropriate tools and technologies, such as Jenkins and GitLab CI/CD, to automate the build, test, and deployment processes. The automatic deployment and CI/CD process may include a series of steps such as code compilation, unit test and integration test running, container image building and pushing, and deployment to a cloud environment. This can speed up delivery, reduce human errors, and provide reliable deployment pipelines.

9. **Security and permissions management**

   In the microservice architecture, security is an important consideration. Ensure that each microservice has an appropriate access control and permission management mechanism to prevent unauthorized access and data leakage. Identity authentication and authorization technologies (such as OAuth and JWT) can be used to verify the validity of requests and transfer identities between microservices. In addition, proper network security measures, such as firewalls and SSL/TLS encryption, are taken to protect communication between microservices. In addition, security review and vulnerability scanning are performed periodically to ensure system security and reliability.

10. **Progressive migration**

    It is a complex process to reconstruct traditional monolithic application microservices and migrate them to the cloud, which may take some time and resources. To reduce risks and interruptions, you can use the progressive migration method.

    First, select a small and relatively independent module for microservice reconstruction and cloud migration. Through this lab project, you can verify the feasibility of architecture design, technology selection, and processes, and learn valuable experience and lessons. After the first module is successfully migrated, other modules are gradually reconstructed and migrated. Progressive migration also helps you gradually develop your team's capabilities and familiarize yourself with new architectures and tools. In addition, you can collect feedback and continuously adjust and optimize the system to ensure smooth reconstruction.

To sum up, microservice reconstruction and cloud migration are complex and critical processes. Multiple aspects need to be considered, such as architecture design, splitting, API definition, service governance, containerization, data management, monitoring and logging, automatic deployment, and security. By evaluating existing applications and objectives, split a monolithic application into independent microservices, introduce appropriate technologies and tools, and use the progressive migration method. Then, you can successfully transform the traditional monolithic application into a highly scalable, elastic, and reliable microservice architecture, and migrate it to the cloud.

Microservice architecture optimization

We often see that the services of some Internet enterprises develop rapidly. Software engineers of different service units continuously add new microservices or repeatedly develop microservices that implement the same service functions. As a result, the microservice architecture is disordered, which severely affects the TTM and makes fault locating time-consuming, in the face of the disordered microservice architecture, the following optimization policies can be used to shorten the time to market (TTM) and improve the fault locating efficiency:

1. **Evaluate the existing architecture.** First, comprehensively evaluate the current microservice architecture. Understand the overall architecture, dependencies between services, communication protocols, and data flows. This will help you sort out the complexity of the architecture and identify the key areas for improvement.
2. **Reconstruction and splitting:** Based on the evaluation result, reconstruct and split the existing microservices. Identify services that are too large, have unclear responsibilities, or are highly coupled, and split them into smaller, more focused units. This simplifies the system structure and improves maintainability.
3. **Service governance**: Use a proper service governance mechanism to manage the microservice architecture. Technologies such as service registration and discovery, load balancing, and fuses are used to enhance service visibility, elasticity, and stability. This helps reduce faults and delays and improve fault locating efficiency.

4. **Implement automated testing:** Establish a comprehensive automated testing strategy and tool chain. Automated testing at all levels, such as unit testing, integration testing, and end-to-end testing, can quickly capture and solve problems, ensuring that modifying a service does not affect other services.
5. **Emphasis on documentation and standards:** Establish clear documentation and standards, including architecture design specifications, interface specifications, and development specifications. This helps team members understand the overall architecture and follow consistent practices during development. Documentation and standards can also help new team members adapt and contribute more quickly.
6. **Real-time monitoring and logging**: The real-time monitoring and logging system is introduced to collect and analyze the running status and performance metrics of microservices. In this way, potential problems or exceptions can be detected in a timely manner, and the problems or exceptions can be quickly located and resolved. In addition, a proper alarm mechanism can help you quickly respond to faults and exceptions.
7. **Continuous delivery and deployment**: Use continuous integration and continuous delivery (CI/CD) tools and processes to automate the build, test, and deployment of microservices. This will shorten the release cycle, reduce release risks, and accelerate the rollout of new features and fixes, thereby improving TTM.
8. **Establish cross-team collaboration:** Encourage cooperation and communication between different teams, especially in the microservice architecture. Facilitate knowledge sharing, problem collaboration, and experience exchange, accelerate problem locating and resolution, and avoid repeated work.

These optimization policies can gradually improve the disordered microservice architecture, shorten the TTM, and improve the fault locating efficiency. The optimization of the microservice architecture is a continuous process, which requires continuous evaluation, adjustment, and improvement.

6.6.4 Modern development and O&M

Development and O&M modernization can be implemented through DevOps practices. The following are some steps and implementation suggestions:

- **Culture change:** First, to modernize development and O&M, we need to build a culture that emphasizes cooperation and sharing in the organization. The development team and O&M team should trust and cooperate with each other to pursue system stability and continuous delivery.
- **Automation**: Automation is one of the core principles of DevOps. Automated tools and processes can reduce manual operations, decrease error risks, and improve efficiency. For example, use continuous integration and continuous delivery (CI/CD) tools to automatically build, test, and deploy applications.
- **Infrastructure as code (IaC):** Infrastructure configuration and management can be incorporated into the code base using an IaC approach. This ensures repeatability, versioning, and automated deployment of the infrastructure, improving the stability and reliability of the entire environment.
- **Centralized log and monitoring**: By centrally managing logs and monitoring data, you can learn about the system running status in real time and detect

and rectify faults in a timely manner. Select appropriate log management and monitoring tools, and define key performance indicators (KPIs) and alarm rules for system availability and performance.

- **Container-based and microservice-based architecture:** Container-based technologies (such as Docker) and microservice-based architecture can be used to decouple and expand applications. This allows the development team to deploy, update, and maintain applications more flexibly, while improving scalability and resilience.
- **Continuous learning and improvement:** DevOps is a continuous evolution process. The team should continuously learn and improve the work process. Through continuous feedback, iteration, and improvement, the collaboration between development and O&M can be gradually optimized to improve the delivery speed and quality.

These are some key practices for modernizing development and operations. Note, however, that specific implementations may vary depending on the needs and status of the organization. It is recommended that the evaluation be performed based on the actual situation, and corresponding practices be gradually introduced and adjusted.

6.6.5 Modern governance and operations

Enterprise cloud migration does not mean modernization of all applications. Instead, old and new applications will coexist for a period of time. Huawei ROMA Connect helps enterprises integrate old and new applications so that they can coexist on the cloud without damaging the existing application environment. The following are some steps and suggestions for reference:

- **Understanding old and new apps**: First, you need to have a comprehensive understanding of the existing apps and the new apps to be integrated. This includes understanding their functions, data structures, interfaces, and communication modes. This will help determine the integration strategy and technology selection.
- **Select an appropriate integration method**: Select an appropriate integration method based on your requirements and app characteristics. ROMA Connect provides multiple integration methods, such as API integration, message queue, and event trigger. Select the most appropriate integration mode based on the dependency and communication mode between applications.
- **Integration solution design**: Design a detailed integration solution based on the application analysis and selected integration mode. This includes defining interface specifications, data mapping, message passing mechanisms, and so on. Ensure that the design solution is compatible with the interaction between the old and new applications and does not interrupt the existing service processes.
- **Integration implementation**: Implement the integration based on the integration solution. Use the tools and platforms provided by ROMA Connect to configure and set necessary integration components and connectors. Ensure that key parameters such as data mapping, message route, and security authentication are correctly configured.
- **Perform comprehensive testing and verification** before the integration application is put into production. Ensure that the data transfer and function invoking between the old and new applications are normal and no damage or conflict occurs.

- **Perform monitoring and maintenance.** Once the integration application goes live, establish a monitoring mechanism to track the running status of the integration environment. The application performance, API availability, and data consistency are monitored. Handle exceptions in time and perform maintenance and optimization periodically.

6.7 Cloud-based Innovation

6.7.1 Overview

Cloud platforms are driving new waves of product and service innovation.

- AI is used with foundation models to make products smarter and offer more precise personal services, like AI customer service and intelligent recommendation systems.
- Blockchain technology improves the security and trust in products and services. It can be used for supply chain management and digital identity verification to create a transparent and trackable system.
- Virtual human technology makes virtual avatars for virtual anchors and online classes, offering an immersive experience.
- Big data analytics helps companies understand user needs, improve products and services, and target marketing precisely.
- IoT technologies connect devices to the cloud for real-time data collection and remote control, enabling innovations like smart homes and cities.

Huawei Cloud provides new technologies to businesses at all times. Businesses can use these technologies to innovate quickly and experiment, speeding up innovation. This helps them create new products and services, improve service processes, boost decision-making, enhance user experiences, and open up new business models and markets.

6.7.2 AI

AI mimics human intelligence and is vital in many areas. AI drives service innovation, enhances services, and supports modernization in these ways:

- **Automation and intelligent decision-making:** AI technologies boost service efficiency and accuracy with automation and intelligent decisions. For example, enterprises can use machine learning to process big data, identify patterns, and predict trends. This speeds up service tasks, cuts down on staff work, and improves decision-making.
- **Personalization and customer experience:** AI improves customer experience with personalized recommendations, intelligent customer service, and virtual assistants. AI analyzes user behavior and preferences to offer personalized product recommendations and services. AI can improve customer service with natural language processing and sentiment analysis, boosting customer satisfaction.
- **Intelligent production and supply chain management:** AI can boost production efficiency and enhance supply chain visibility and planning. For example, machine learning and IoT can improve production line efficiency and device reliability through smart manufacturing and predictive maintenance. AI

can also improve inventory management, transportation planning, and delivery routes in the supply chain. This reduces costs and enhances response times.

- **Innovative business models:** AI technologies open up new business models and market opportunities for companies. For example, cloud computing and AI together enable flexible computing and on-demand services, boosting the growth of the Software as a Service (SaaS) model. AI and IoT together can support new business models in areas like smart homes, smart cities, and smart healthcare.

6.7.3 Big Data

Big data involves large, complex datasets. It is crucial for businesses to gather, store, and analyze this data. Big data drives service innovation, enhances services, and supports modernization in these ways:

- **Data-driven decision-making:** Big data analytics helps companies get useful insights from large amounts of data to support decisions. Enterprises can find market trends, demand changes, and potential risks by analyzing historical and real-time data. This helps them make accurate decisions and boost their competitiveness.
- **Personalized marketing and customer relationship management:** Big data helps companies understand customers better and use personalized marketing and customer care. Enterprises can boost sales and customer satisfaction by studying customer behavior, interests, and preferences to tailor product recommendations and marketing efforts.
- **Predictive analytics and supply chain optimization:** Big data analytics helps companies predict future needs, improving production and supply chain planning. Enterprises can analyze past sales, market trends, and supply chain data to forecast demand, optimize inventory, plan deliveries, cut inventory costs, boost operations, and improve supply chain response.
- **Innovative products and services:** Big data strongly supports product and service innovation for companies. Enterprises can find market gaps and opportunities by analyzing big data. They can also understand user needs and create more competitive, innovative products and services. For example, some companies use big data to analyze medical records and genomic data for personalized medical solutions.

6.7.4 Blockchain

Blockchain is a decentralized and distributed ledger technology that ensures data security and reliability. Blockchain drives service innovation, enhances services, and supports modernization in these ways:

- **Transparency and reliability:** Blockchain technology stores all transactions and data records transparently and securely. This improves data reliability and transparency for businesses, removes the need for traditional intermediaries, and lowers operational risks.
- **Smart contracts and automated execution:** Smart contracts are automatic agreements for the blockchain that execute based on set conditions and rules. They are widely used in supply chain management and financial services. Smart contracts boost transaction efficiency, cut manual steps, lower costs, and prevent fraud and disputes.

- **Disintermediation and friction reduction:** Blockchain removes many intermediaries, making transactions faster, cheaper, and smoother. For example, enterprises can use blockchain for quick cross-border payments and fund clearing, cutting out intermediary banks or payment firms.
- **Decentralized applications and community economy:** Blockchain technology supports decentralized applications. Enterprises can use blockchain to create platforms for decentralized applications that allow direct user transactions and value transfers. This model can boost user involvement, share value, and foster innovation and collaboration.

6.7.5 Metaverse

The metaverse is a virtual digital world that connects with the real world using technologies like augmented reality (AR) and virtual reality (VR). Metaverse drives service innovation, enhances services, and supports modernization in these ways:

- **Interaction and collaboration:** Metaverse technologies offer more immersive and interactive experiences, letting users work together in virtual settings. Enterprises can use the metaverse to create virtual meetings, training sessions, and team collaborations for remote work. This will improve workflows and create global collaboration chances.
- **Virtual stores and digital assets:** The metaverse allows businesses to create virtual stores and sell digital assets. Enterprises can use the metaverse to show and sell virtual products, digital art, and virtual real estate. This model can open new revenue streams and reach a global audience.
- **Virtual tourism and entertainment:** The metaverse can revolutionize tourism and entertainment. With virtual reality, users can visit famous places, attend concerts, or watch sports events virtually. This will expand markets and introduce new business models to the tourism and entertainment sectors.
- **Data collection and personalized experience:** Metaverse technologies gather user behavior data in virtual settings, giving companies deeper insights and enabling personalized experiences. Enterprises can improve user satisfaction and loyalty by studying behavior, interests, and preferences in virtual spaces and customizing products and services accordingly.

6.7.6 Internet of Things

The Internet of Things (IoT) connects devices and sensors to the internet for communication and data sharing. IoT drives service innovation, enhances services, and supports modernization in these ways:

- **Smart home and smart city:** IoT technology connects home devices, urban infrastructure, and public services to manage them intelligently and optimize resource use. IoT allows people to control home devices with phones or other devices, making their home smart. IoT can also improve smart cities by optimizing urban transportation, energy management, and public security.
- **Industrial automation and smart manufacturing:** IoT technology helps with industrial automation and smart manufacturing. Connecting devices to the IoT lets businesses do collaborative work, monitor remotely, and predict maintenance needs. This boosts production efficiency, cuts the failure rate, and streamlines supply chain management.
- **Data collection and analytics:** IoT devices collect a lot of sensor data, like temperature, humidity, and pressure. Enterprises can gain valuable insights by

analyzing data to improve product quality, optimize operations, and predict demand changes.

- **Customer experience and value-added services:** IoT devices connect to customers' mobile phones or other devices to offer personalized services and enhance their experience. For example, smart home devices can automatically change the room temperature, lighting, and security settings based on user habits. This creates a more comfortable, convenient, and secure home. IoT also lets businesses offer custom products and services to meet users' specific needs.
- **Asset tracking and supply chain management:** IoT technology helps track and manage assets and items in supply chains. Enterprises can track item locations, statuses, and transport in real time by using sensors and tags. This improves logistics and supply chain efficiency and reduces the risk of loss or damage.
- **Health monitoring and healthcare:** IoT technology has many uses in healthcare. Sensors in medical devices, wearables, and health monitors track patient health data in real time and performs remote monitoring and diagnosis. This improves healthcare efficiency, reduces waste of medical resources, and enhances patients' quality of life.

6.8 Anti-patterns in Cloud Adoption

During cloud adoption and implementation, you may encounter some anti-patterns. If not identified and avoided, these patterns may lead to low migration efficiency, service interruptions, unnecessary costs, and increased maintenance difficulties. The following are some common anti-patterns in the cloud adoption and implementation phase:

- **Non-automated deployment**

  Enterprises rely on manual processes for code and cloud resource configuration and deployment. This results in low efficiency and high error rates.

  Optimization suggestion: Use automated configuration and deployment tools, such as Terraform and CI/CD, to improve the efficiency and accuracy of cloud resource deployment.

- **No mock cutover**

  Insufficient mock cutovers are performed, leading to unexpected issues during the real cutover.

  Optimization suggestion: Perform comprehensive mock cutovers in various near-real environments before the real cutover. This helps detect and resolve problems in a timely manner, and ensure that the system can run properly after the cutover.

- **Insufficient testing**

  Insufficient tests are performed before service system cutover. Potential functional, performance, and security problems are not detected and resolved in a timely manner, causing poor user experience after the service system goes live.

  Optimization suggestion: Perform thorough function, performance, availability, and security tests before the cutover to ensure that each function module can run properly in the cloud environment.

- **Untagged cloud resources**

  Cloud resources are not correctly tagged, so they are difficult to query, monitor, and manage.

  Optimization suggestion: Tag all created cloud resources to facilitate subsequent O&M management and cost reduction.

It is crucial to identify and avoid these anti-patterns. Meanwhile, more importance should be attached to the best practices and success stories in the industry. In this way, the cloud migration solution can be implemented more scientifically, the efficiency of cloud migration and utilization can be improved, the advantages of the cloud platform can be better leveraged, and the value of cloud technologies can be brought into full play.

7 O&M Governance

7.1 Overview

After applications are migrated to or deployed on the cloud, cloud transformation officially enters the O&M governance phase. This phase is critical because it determines the performance, reliability, security, and cost-effectiveness of IT infrastructure and service systems on the cloud. Through continuous and effective O&M governance, enterprises can efficiently use cloud resources, maintain service continuity and stability, gain full control over the cloud environment, and maximize the benefits of cloud transformation.

The O&M governance phase involves lean governance, deterministic O&M, all-round security operations, and refined FinOps for cloud IT infrastructure, applications, and big data platforms. Continuous optimization is performed based on the WAF framework in this phase. First, lean governance runs through the entire process. It provides a set of standard governance frameworks and best practices to efficiently use cloud resources, meet security requirements, and minimize risks of cloud transformation. Second, deterministic O&M aims to build a preventable, controllable, and governable O&M management system, turning the "uncertainty" brought by digital transformation and rapid business development into "certainty". It ensures that applications can run stably for a long time and reduces faults and downtime.

All-round security operations cover data security, network security, and access control. The comprehensive security system protects enterprises' core data and applications. Finally, refined FinOps implements cost analysis, optimization, and budget management. It helps enterprises control cloud expenditures and maximize the return on investment (ROI) of cloud transformation. All these activities should be continuously optimized based on WAF. In this way, the cloud environment always remains in the best state and can be flexibly adjusted to fit business requirements.

Setting Up O&M Governance Teams

The main tasks in the O&M governance phase include lean governance, deterministic O&M, security operations, and FinOps. All these tasks must be efficiently completed by personnel with the required skills. Set up O&M

governance teams, including a cloud governance team, cloud O&M team, cloud security team, and FinOps team, for your enterprise. For details about the roles, responsibilities, and skill requirements of these teams, see **CCoE**.

7.2 Lean Governance

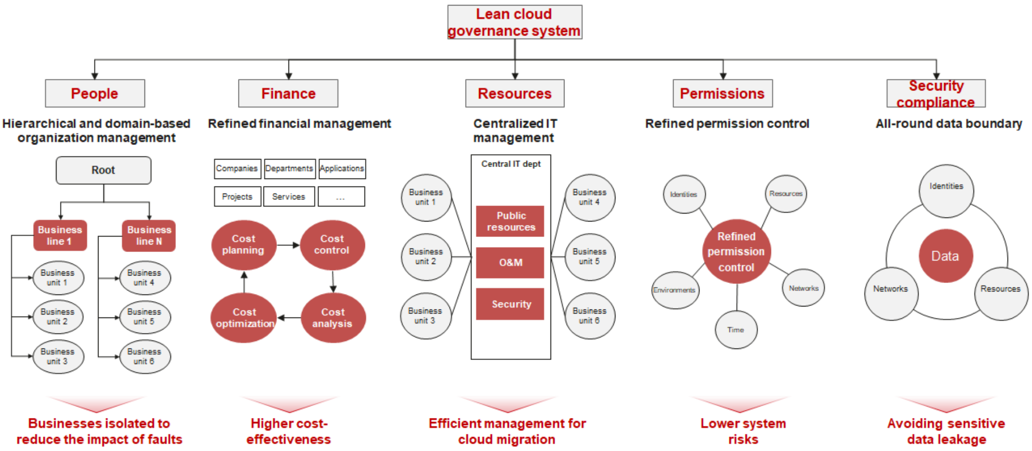
7.2.1 Overview

As enterprises move more and more applications to the cloud, they run dozens or even hundreds of applications with mass cloud resources. A large number of users, including enterprise employees, outsourced employees, and partners, need to access these cloud resources. Risks such as resource idleness, misoperations, malicious operations, data leakage, and permission misconfiguration increase exponentially as the scale of cloud use grows.

You must build a lean, centralized, and structured IT governance system to effectively control these risks, maximize your business benefits, and ensure continuous growth.

Huawei Cloud's IT governance practices and customer experience have led us to create a cloud-based lean governance system, as shown in the following figure. This system centrally manages people, finance, resources, permissions, and security compliance. In this figure, hierarchical and domain-based organization management, centralized IT management, all-round data boundary, and refined permission control are part of Huawei Cloud's Landing Zone solution. Refined cost operations belong to Huawei Cloud's FinOps solution.

Figure 7-1 Cloud-based lean governance system



7.2.2 Hierarchical and Domain-based Organization Management

Hierarchical and domain-based organization management is key to modern enterprise governance, especially in cloud computing. As businesses grow and diversify, this approach becomes even more crucial.

Huawei Cloud's **Organizations** service helps you create a structured, manageable cloud setup that aligns with your business needs. You can assign separate cloud accounts for each unit—like subsidiaries, business systems, product lines, departments, and projects—to keep them organized and independent.

This approach aligns with your current governance structure while effectively isolating faults and security issues. It confines faults and risks within individual business units, minimizing their impact and reducing potential damage.

First, hierarchical and domain-based management aligns effectively with your enterprise's governance architecture. An enterprise usually consists of multiple business units, each with specific goals, teams, and operation modes. Enterprises can mirror their organizational structure by allocating each business unit to separate cloud accounts. This clarifies each business unit's responsibilities and permissions while simplifying resource allocation and management. For example, a company has multiple subsidiaries, each with their own product lines and project teams. Each subsidiary uses its own cloud account for resource deployment. This allows independent management while ensuring consistent monitoring under the company's overall strategy.

Second, hierarchical and domain-based management has significant advantages in fault and security isolation. Cloud environments offer convenient and efficient shared resources, but they can also raise potential risks. Running all business units under one cloud account increases the risk that a fault or security issue in one business unit could impact others or the whole system. Using separate cloud accounts for each business unit helps enterprises contain faults and risks and stop them from spreading. For example, if a business system faces a security attack, the attacker can only access resources within that specific cloud account, and cannot affect other business units. This isolation mechanism significantly enhances both security and stability across the enterprise.

In addition, each business unit can view and manage only its own cloud account's resources, data, and applications. This approach protects business data by maintaining their security and privacy. Isolating data across business units prevents data leakage and tampering caused by human errors or malicious behavior. This also allows each team to concentrate on their growth while minimizing unnecessary interference and conflicts.

The IT department manages all business units with a mix of centralized and decentralized control. In this mode, each business unit operates separately within its own cloud account while the IT department uses consistent policies and tools to monitor, manage, and optimize cloud resources company-wide. For example, the IT department can create consistent security policies for all cloud accounts to meet company security requirements. They can also automate monitoring of resources in each cloud account to quickly spot and fix problems. This management approach balances business flexibility with centralized control. Business units can independently adjust resources and policies as needed to quickly respond to market changes. The IT department ensures the company's overall security, compliance, and resource utilization efficiency globally.

Finally, hierarchical and domain-based management also helps with refined cost control. Enterprises can clearly see the resource usage and costs of each business unit by billing each cloud account separately. This strongly supports cost optimization and budget management. Business units plan and adjust resources based on their budgets. The company sets and adjusts budgets and cost strategies for each business unit based on overall financial goals.

7.2.3 Refined Permission Control

As security rules get stricter, enterprises give users only the permissions they need for their jobs. Refined permission control allows enterprises to accurately set five access control elements: Who, What, How, Where, and When. "Who" shows who can access cloud resources. "What" lists the resources that can be accessed. "How" explains the actions users can perform on these resources. "Where" specifies the locations from which users can access the resources. "When" sets the time period when users can access the resources. Refined permission control is implemented in the following aspects:

- **Refined resources**: Authorize users to access only a specific resource, resources with a specific tag, or the resources under an enterprise project. To allow user access to multiple resources, put them under the same enterprise project or add the same tag to them. Then, configure permissions for the project or tag.
- **Refined operations**: Configure permissions for read, write, and list operations performed on specific objects. For example, the read permissions for cloud server specifications, tags, server details, mounted disks, and NICs are separately configured. In this way, you can grant users the minimum permissions they need.
- **Refined attributes**: Attribute-based access control (ABAC) is more flexible and refined than role-based access control (RBAC). You can add attribute-based conditions to permission settings to allow only access requests that meet the conditions. Attributes include: identity details like username, MFA status, and root user status; network details like source IP address and VPC ID; resource details like tags and names; time details like access and token issuance time; and environment details like requesting and target accounts.

  For details about the global and service-specific condition keys supported by Huawei Cloud, see **SCP Syntax**. ABAC allows for more granular access control permissions. For example, O&M engineer Shane can only shut down and restart a specified ECS after MFA authentication is enabled and between midnight and 4:00 AM.

7.2.4 Centralized IT Management

In centralized IT management, a single IT department oversees all IT resources, services, and functions. This includes network management, O&M, security, compliance audits, identity and permission control, and public resource management for different business units. This model boosts IT management efficiency and consistency while cutting operation costs. Business units do not need to deploy or maintain infrastructure, speeding up service cloudification.

Centralized IT management lets a central IT department manage and coordinate IT resources, services, and functions across business units. This model is crucial in today's fast-paced IT development. Centralizing IT functions helps enterprises improve management efficiency and consistency and cut operation costs. In centralized IT management, the central IT department (or CCoE) can manage multiple business units in these areas:

- **Centralized network management**: The central IT department plans, deploys, and maintains the enterprise's cloud network infrastructure, including Direct Connect, Enterprise Router, VPN, Cloud Connect, NAT Gateway, and

VPC. This unifies, stabilizes, and secures the network across the enterprise, preventing inconsistency and potential security risks that arise from network management by various business units. In addition, unified network management improves data transmission and ensures smooth information flow between departments.

- **Centralized O&M management**: Services like AOM and COC provide unified monitoring and O&M management across accounts. This allows the central IT department to handle the O&M of cloud resources for all business units. We set up standardized O&M processes and specifications to centrally handle performance monitoring, troubleshooting, and upgrades for cloud resources. This approach sets unified O&M standards, boosts efficiency, and cuts costs.
- **Centralized security management**: Network security is the top priority in enterprise operations. SecMaster allows the central IT department to manage security for all business units. It handles cloud assets, security situations, information and events, orchestration, and automated responses. Centralized security management keeps security policies consistent, prevents network attacks and data leaks, and safeguards core assets.
- **Centralized compliance audit**: Compliance is now a major concern for enterprises due to new information security rules and industry standards. The central IT department manages and audits each business unit's cloud resources to meet national, industry, and enterprise standards using the multi-account audit function of services like CTS and Config. This lowers compliance risks and avoids legal issues and reputational damage from non-compliance.
- **Centralized identity and permission management**: The central IT department manages Huawei Cloud users, sets up single sign-on (SSO) with the enterprise's identity system, and controls user access to accounts by using the multi-account identity management and access control function of IAM Identity Center. The administrator creates users, sets passwords, and manages users by group. Centralized permission management strengthens user permission control to block unauthorized access and secure the system.
- **Public resource management**: The central IT department deploys and manages public IT resources like DNS servers, container image repositories, CA authorities, and cloud disks. Centralized management prevents redundant building and wasted resources, boosts resource utilization, and reduces procurement and maintenance costs.

Centralized management in all these aspects improves IT efficiency and consistency for enterprises. First, it standardizes processes and specifications, making IT work more organized and transparent. This reduces errors and security issues that arise from loose management. Second, the central IT team has strong technical skills and extensive experience, and can bring in advanced technologies and best practices to offer top-notch IT support for enterprises.

Centralized IT management lowers operation costs effectively. Enterprises can get better prices and cut extra costs through unified resource planning and bulk purchasing. Centralized O&M and management improve human resource allocation and prevent waste from separate business unit management. In this way, enterprises can get more efficient and reliable IT services without spending more money.

Centralized IT management can speed up IT project delivery for each business unit. The central IT department has set up comprehensive infrastructure and service frameworks. This allows business units to quickly integrate and deploy new systems or applications as needed, avoiding duplicate setup and debugging. Business units can put more efforts and resources to their core business growth. This division of labor lets business units react faster to market needs, speed up product and service innovation, and boost market competitiveness.

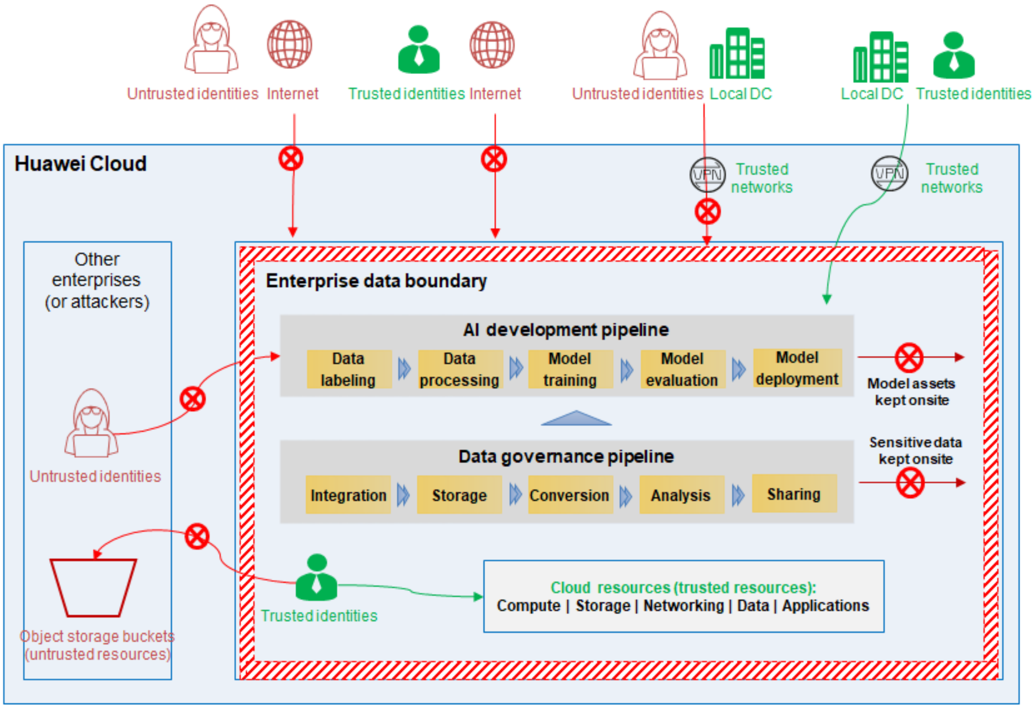
However, implementing centralized IT management requires enterprises to adjust their organizational structure, management model, and culture. First, enterprises should set up clear management systems and processes. They must define roles and responsibilities between the central IT department and business units. Smooth communication between them is essential. Second, the central IT department should have a service-focused mindset and be flexible to offer tailored support and solutions for each business unit's needs. Finally, executives must pay close attention and provide strong support to centralized IT management to remove implementation hurdles.

In summary, centralized IT management meets the needs of modern enterprise growth. Centralizing IT resources, services, and functions in the IT department improves efficiency and consistency, reduces operation costs, and boosts competitiveness. Each business unit can focus on its core business without handling infrastructure deployment or O&M. This speeds up cloud adoption and boosts business development and innovation. Enterprises should plan strategically when setting up centralized IT management. They need to balance every party's needs and use centralized management to build a strong foundation for long-term growth.

7.2.5 All-Round Data Boundary

The all-around data boundary uses identity, network, and resource control policies to create a strong security shield. It ensures that only trusted, verified identities can access specific resources in a trusted, secure network environment, thereby maintaining data security. See the following figure. Requests from trusted identities to access cloud resources via the internet (untrusted network) are denied. Requests from untrusted identities to access cloud resources via the local data center network (trusted network) are also denied. Requests from trusted identities to access object storage buckets of other enterprises (untrusted resources) are still denied. Only requests from trusted identities to access cloud resources of their own enterprises via the local data center network (trusted network) are allowed.

Figure 7-2 All-round data boundary



Identity Control Policies

Identity control policies are the first line of defense for data boundaries. They ensure that only trusted, verified identities can access enterprise cloud resources. Identity control policies are implemented using service control policies (SCPs) and IAM policies.

SCPs control access based on organizational structure. The organization management account can use SCPs to limit permissions for member accounts to ensure that they stay within your organization's access control guidelines. SCPs can be attached to an organization, organizational units (OUs), and member accounts. Any SCP attached to an organization or OU applies to all the accounts within the organization or under the OU.

For details about SCPs, see **SCP Introduction**.

SCPs can affect many areas. Test them thoroughly before applying them in your production environment to prevent issues with your cloud resources.

IAM policies control access to Huawei Cloud resources for users, user groups, and agencies. Enterprises can use IAM policies to give users only the permissions they need for operations on specific resources, following the principle of least privilege. This improves security and reduces permission abuse risks.

For details about IAM policies, see **Basic Concepts**.

On Huawei Cloud, identity control policies are implemented using SCPs and IAM policies. When both SCPs and IAM policies are set, user permissions are the overlap between the two.

Network Control Policies

Network control policies are the second line of defense for data boundaries. They are implemented using Virtual Private Cloud Endpoint (VPCEP) policies.

VPCEP is a network service that sets up private connections in a VPC. It lets you securely connect your VPC to Huawei Cloud services and resources without using public IP addresses. Data is not transmitted over the internet, reducing the risk of interception and attacks. VPCEP policies control which principals (or identities) can use VPC endpoints to access cloud resources. By setting these policies, enterprises can create a closed, secure network. This keeps data safe inside the network and allows only trusted identities to access cloud resources and sensitive data. For example, enterprises can deploy key service systems in VPCs and connect them to RDS DB instances on Huawei Cloud using VPCEP, keeping them off the public network. Only database administrator Shane can access these RDS DB instances via VPCEP, adding another layer of security for sensitive data.

For details about how to manage VPCEP policies, see **Managing the Policy of a VPC Endpoint**.

Resource Control Policies

Resource control policies are the third line of defense for data boundaries. They are implemented using cloud services' access control policies, such as OBS's bucket policies.

Bucket policies let owners set permissions for their buckets and objects. Enterprises can use bucket policies to specify which users can or cannot perform operations on buckets. For example, you can allow a user to read a specific bucket but not write to or delete it. Bucket policies allow flexible access control based on conditions like requester IP address and request time. For example, an enterprise can set a bucket policy to allow only internal network requests or limit access to a specified period.

For details about how to create a bucket policy, see **Creating a Bucket Policy with a Template**.

Together, bucket policies and VPCEP policies enhance security. You can set VPCEP policies to let servers (ECS, CCE, and BMS) in a VPC access only certain OBS resources. You can also set bucket policies to allow only servers in a specific VPC to access OBS buckets. This maintains security by controlling both request sources and accessed resources.

Building an all-round data boundary is a systematic and comprehensive project. Enterprises must apply strict control policies for identities, networks, and resources. Coordinating all these three aspects creates a strong data security shield.

7.2.6 Refined Cost Operations

Refined cost operations are based on the FinOps concept, combining financial management with cloud resource operations. It helps enterprises optimize cloud resource utilization and manage costs.

Enterprises can use the FinOps-based cost management system for refined cost control and resource allocation in the cloud. This system helps enterprises analyze and manage costs across various levels, such as subsidiaries, business units, product lines, departments, projects, and even microservices. They can accurately identify each business unit's costs and make better decisions. Using FinOps enables enterprises to improve resource utilization, cut waste, and lower costs while maintaining strong service performance. For details about FinOps practices, see the following sections.

In short, enterprises can use lean governance to protect data boundaries, manage complex organizations, control user permissions, and optimize resource costs. This helps mitigate cloud-related risks and maximize business benefits.

7.3 Deterministic Operations

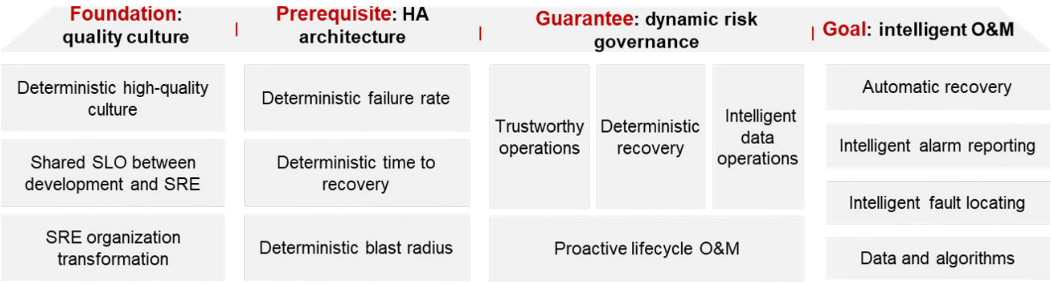
Deterministic operations is a set of O&M concepts, methodologies, and best practices developed by Huawei Cloud based on years of cloud service O&M experience. It helps enterprises efficiently operate and maintain self-built and purchased service systems on the cloud, ensuring that these service systems can run continuously, efficiently, and stably on the cloud.

Deterministic operations is intended to create an O&M management system that makes risks avoidable, controllable, and manageable. It aims to minimize fault probability and strives for zero faults through high-quality product development and rigorous O&M processes and regulations. This also involves technical means to manage possible faults, ensuring that the occurrence interval, impact scope, and recovery time are avoidable, controllable, and manageable. In a word, the "uncertainty" brought by digital transformation and rapid service development can be managed through O&M.

Deterministic operations can help enterprises improve resource utilization through proper resource planning, allocation, and scheduling. In addition, deterministic operations can use automated, intelligent methods to help enterprises improve O&M efficiency, reduce O&M costs, and save a large number of resources.

Deterministic operations is a comprehensive system that involves quality culture, high-availability (HA) architecture, dynamic risk governance, and intelligent O&M tools, as shown in the following figure.

Figure 7-3 Deterministic operations framework


Foundation: quality culture
Prerequisite: HA architecture
Guarantee: dynamic risk governance
Goal: intelligent O&M
Deterministic high-quality culture
Shared SLO between development and SRE
SRE organization transformation
Deterministic failure rate
Deterministic time to recovery
Deterministic blast radius
Trustworthy operations
Deterministic recovery
Intelligent data operations
Proactive lifecycle O&M
Automatic recovery
Intelligent alarm reporting
Intelligent fault locating
Data and algorithms

- **Quality culture: the foundation**

  A quality culture is the cornerstone of deterministic operations. It can be a powerful tool to motivate team members to take responsibility for providing standardized and refined O&M. These are some best practices for building a high-quality culture:

  - Emphasize quality from the top down and make it a core value.
  - Establish shared quality goals and methods for R&D and O&M teams.
  - Transform the O&M team and continuously improve their capabilities. Use software engineering methods to solve problems and move from reactive to proactive approaches.

- **HA architecture: the prerequisite**

  An HA architecture is the prerequisite for deterministic operations. By designing and deploying a thoughtful architecture, you can minimize system failures, recover faster, and mitigate their impacts. To achieve this, you need to:

  - Target at SLOs, design the architecture using scientific methods, and manage the selection and implementation time.
  - Assign O&M teams rights and responsibilities during product planning, design, and launch, and set restrictions on the development and commercial-use plans to ensure the implementation of HA requirements.
  - During O&M, verify the HA design as scheduled to ensure that the system meets the HA requirements.

- **Dynamic risk control: the guarantee**

  Dynamic risk governance is crucial to responding to uncertainty and sudden events. It is essentially a process of identifying changes, fault modes, and service data to support proactive O&M throughout the lifecycle.

  - For managing the risks involved in change jobs, you need comprehensive abilities, including creating version release architectures, managing account permissions, and automating changes.
  - For managing known and unknown faults, you need to use scientific methods to create a fault mode library and develop rapid recovery capabilities. This involves contingency plans to quickly respond to sudden events and regular drills and reviews to verify the architecture's availability and the team's emergency response capabilities.
  - Intelligent operations of service runtime data is essential for continuous improvement. A real-time data collection and operations system is required to support decision-making.

- **Intelligent O&M: the goal**

  Intelligent O&M tools can help improve O&M efficiency and quality and reduce labor costs. Especially in the AI era, you can manage and maintain systems more efficiently by using automation and intelligence technologies.

  - Select appropriate tools and technologies to ensure that they match service requirements and technology stacks, such as automatic deployment, fault prediction, and intelligent demarcation and locating.
  - Integrate the tools into existing systems, and customize and optimize the tools to meet specific O&M requirements.
  - Leverage new technologies to continuously update and upgrade intelligent O&M tools.

7.4 Secure Operations

7.4.1 Overview

Security relies more on operations (70%) than on technology (30%). Security operations protect cloud resources, data, and applications through continuous monitoring, detection, response, and improvement. This highlights that security is an ongoing process, not a one-time task. Effective security operations are essential to coordinate multiple defense layers and ensure the secure, stable operations of service systems and key data. However, security operations face many challenges.

- **Complex security systems**

  As digital transformation progresses, the ICT environment of enterprises becomes more complex. Cloud computing, network channels, devices, edge computing, operating systems, databases, and applications interweave to form a large, complex ecosystem. Each component may involve potential security vulnerabilities, increasing the difficulty of overall security management. Additionally, the fragmented security industry exacerbates this complexity. The market is saturated with numerous security vendors, each offering different products and solutions. This results in a lot of logs and data in various formats, with no unified standards. Consequently, integrating and analyzing security information is challenging, and forming a global security situation awareness is nearly impossible.

  Moreover, increasing compliance requirements pose new challenges for enterprises. Laws and regulations, such as China's Cybersecurity Law, Data Security Law, and Personal Information Protection Law, the EU's GDPR, the financial industry's PCI-DSS, and the healthcare industry's HIPPA, impose strict requirements on data privacy and cybersecurity. Enterprises must invest significant resources to meet compliance standards across different regions and industries, adding to their management burden.

  In addition, the attack methods are becoming increasingly complex. Attackers use AI and machine learning technologies to accelerate the iteration of attack tools and methods. For example, an advanced persistent threat (APT) is a covert and persistent network attack. Attackers, usually well-resourced organizations or criminal groups, have clear objectives. They lurk for extended periods, using various advanced technologies to steal sensitive data or damage target systems. APT attacks are difficult to detect and defend against and are extremely harmful.

  In summary, the complexity of the security system arises from the diversified technical environments, fragmented security industry, stringent compliance requirements, and complex attack methods. To address current security challenges, enterprises need to establish a unified security management platform to integrate various security information and enhance overall protection capabilities.

- **Lack of security experts**

  The shortage of security experts is a significant bottleneck for enterprises' secure operations. First, due to limited investment, many enterprises cannot build large security teams and thus lack professional security talent. The security field is highly specialized, and it takes considerable time to cultivate a qualified security expert with extensive experience and skills. Moreover, there is no effective mechanism for systematically accumulating and transferring the experience and knowledge of security experts. When experts leave, valuable experience is lost, causing significant harm to enterprises.

Frequent security incidents also overburden experts, whose energy is often consumed by routine operations such as handling numerous security alarms, analyzing logs, and performing routine security checks. These tasks, though important, are repetitive, time-consuming, and labor-intensive, preventing experts from focusing on more valuable work like security strategy planning, complex threat analysis, and security system optimization.

Additionally, as attack technologies evolve, security experts must continuously learn and update their knowledge to maintain their professional level, which further increases their pressure and burden. In the competitive talent market, retaining security experts is also a major challenge.

To address the shortage of security experts, enterprises need to increase investment in security talent cultivation and establish robust training and promotion mechanisms. They can also leverage automation and intelligent tools to reduce repetitive tasks for experts, allowing them to focus on core security affairs. Establishing a knowledge management system to accumulate and share expert experience can mitigate risks caused by talent loss.

- **Inefficient security operations**

  Inefficient security operations are a common issue for enterprises. First, the sheer volume of risk alarms is overwhelming. Security devices generate numerous alarms daily, many of which are false positives or redundant. Security personnel struggle to filter and address all alarms promptly, leading to potential oversight of real threats amidst the noise.

  Second, threat identification is slow. Complex security events require extensive manual analysis due to the lack of intelligent tools, delaying the determination of threat nature and severity. This reactive approach can miss critical response windows, allowing security incidents to escalate.

  Additionally, event response and handling are slow. The process from detection to action involves multiple departments and personnel, making coordination complex. Manual operations are prone to omissions and errors, affecting the response efficiency.

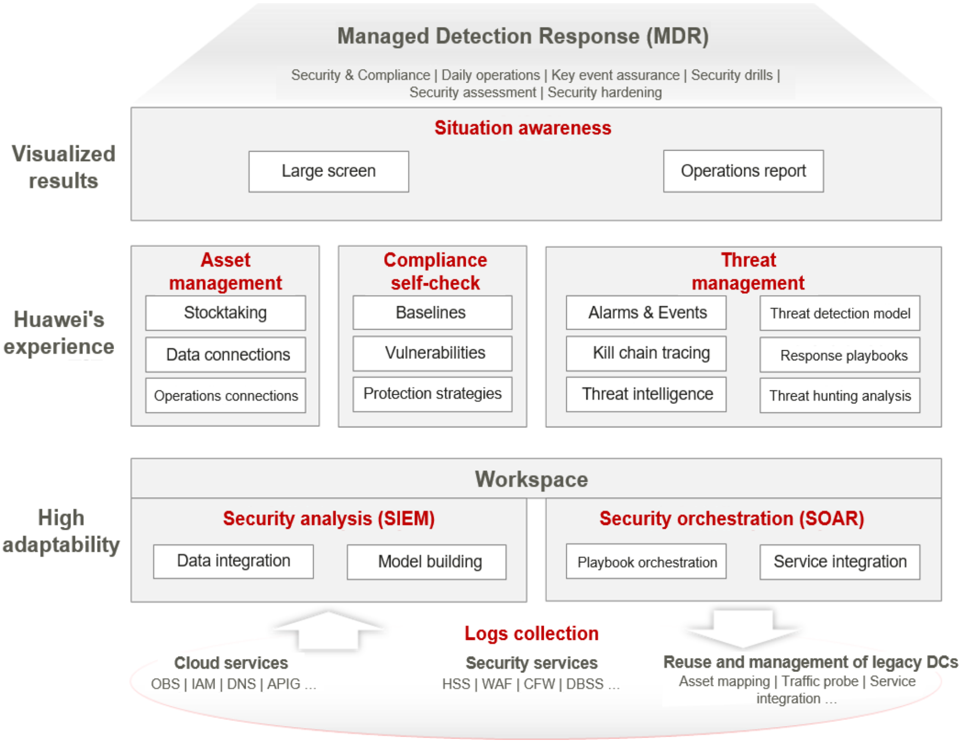
  The root cause is the absence of efficient security operations mechanisms and tools. Traditional methods cannot keep pace with the rapidly evolving security landscape. To enhance security operations efficiency, enterprises must adopt advanced Security Operations Centers (SOCs) and leverage big data analysis and machine learning to automate alarm correlation and prioritization. They should also implement automated response tools to expedite event handling and establish standardized processes and collaboration frameworks to boost cross-departmental efficiency. Furthermore, it is crucial to train security personnel to enhance their analytical and decision-making skills.

  In summary, improving security operations efficiency requires advancements in both technology and management. Only by building an efficient and agile security operations system can enterprises respond to threats promptly and safeguard their core service systems and data.

7.4.2 Secure Operations Framework

With years of experience in security operations and continuous support for customers, Huawei Cloud **SecMaster** offers the following security operation frameworks and processes to help you get started.

Figure 7-4 Secure operations framework



- **Define security operation responsibilities**

  First, you need to clarify the responsibilities of application teams and the Cloud Center of Excellence (CCoE) team. This is subject to each enterprise's cloud operations mode. For example, in the enablement and collaborative operation mode, the CCoE team is responsible for security protection and centralized security operations at the platform layer, and application teams are responsible for security operations of application systems and required cloud resources. Then, SecMaster allocates workspaces for the CCoE team and each application team, ensuring clear division of responsibilities and providing a unified security operations management view for the CCoE team. Finally, you need to connect the security data and assets of the service system running environment to the unified security operations platform. SecMaster can manage the running environments of Huawei Cloud and IDCs, as well as environments across regions and accounts.

- **Identify and clear asset risks**

  SecMaster automatically inventories assets on Huawei Cloud, including hosts, IP addresses, websites, and databases, and connects on-premises asset information via third-party software.

  You can enable HSS to scan for vulnerabilities in Linux, Windows, and web applications, integrating results to SecMaster for lifecycle management. OS vulnerabilities can be patched with a few clicks.

  Additionally, SecMaster identifies compliance risks, aligning with over 10 standards like DJCP, PCI-DSS, and ISO27001. It also provides best practices, hardening suggestions, and reports. You can use SecMaster to comprehensively check the protection coverage of assets and centrally manage security policies at each defense line.

- **Perform data integration and security analysis**

  First, integrate security data and cloud service logs to SecMaster. Build a unified on- and off-cloud data integration solution to access logs of third-party ecosystem software.

  Then, build a security threat detection model based on security data. This model identifies kill chains, analyzes behaviors of each line of defense, and implements in-depth monitoring. It uses big data technologies and intelligent analysis algorithms to accurately detect abnormal behaviors and show them in pop-up messages. The model uses User and Entity Behavior Analytics (UEBA) to restore kill chains and automatically update user profiles, asset fingerprints, and intelligence profiles.

  Then, after you connect your platform with threat intelligence, the data, logs, and alarms in daily operations are displayed. The information on the Huawei Cloud platform is shared with you to help you perform security operations.

- **Handle security incidents**

  With SecMaster, you can respond to, investigate, and handle alarms and events. It provides entities, assets, intelligence, and historical information to help you check and trace kill chains, along with handling suggestions. Playbooks and processes can be orchestrated to quickly address and manage security threats.

These approaches should be carried out in a long term to continuously discover and handle new security threats, thereby enhancing system security and improving the capabilities of the security operations team.

7.4.3 Security Configuration Baseline

Security configuration is essential for protecting information systems. For cloud environments, proper security configuration is fundamental to safeguarding cloud services and assets.

Failure to meet security configuration baseline requirements can lead to significant risks. To enhance cloud security, Huawei Cloud offers the *Huawei Cloud Security Configuration Guide*. This guide covers key areas such as identity and access management, logging and monitoring, VMs and containers, networks, storage, databases, and enterprise intelligence. However, it does not cover all possible configurations. It is advised to use this guide as a starting point and customize it based on your specific needs.

You can manually verify if your cloud service's security configuration meets the baseline requirements using the *Huawei Cloud Security Configuration Guide*. If discrepancies are found, follow the guide's remediation steps to address them. Alternatively, you can use SecMaster's automated **baseline check** feature for this purpose.

7.4.4 Software Engineering Security

Software engineering security involves applying security principles, practices, and technologies throughout the software development lifecycle to reduce vulnerabilities and ensure software confidentiality, integrity, and availability. It covers all phases from requirement analysis, design, coding, testing to deployment and maintenance.

- **Security design**

  Enterprises must adhere to security and privacy design principles, specifications, and legal requirements. During security requirement analysis and design, they analyze threats using service scenarios, data flow diagrams, and network models. Threat, mitigation, and security design solution libraries, derived from the enterprise's security engineering experience and industry best practices, are used for this analysis. Identified threats lead to the development of mitigation measures and corresponding security solutions by application architects. These measures are then translated into security requirements and functions. Security test cases are created based on the company's test case library to ensure system security.

- **Secure coding and testing**

  Enterprises must develop secure coding guidelines and ensure that development and test personnel pass relevant training and exams before their onboarding. Additionally, enterprises should implement static code scanning tools for regular checks. The results are integrated into the CI/CD pipeline and assessed against quality thresholds to evaluate application security. All application systems must clear static code scanning alerts before release to ensure no coding-related security issues.

  To ensure application security, all cloud services undergo multiple rounds of testing by application test personnel before release. This includes testing for authentication, authorization, API security, and database security. Test cases cover both security requirements from the design phase and penetration test cases from an attacker's perspective. Systems that fail security tests are not permitted to go online.

- **Third-party software security management**

  Enterprises must establish clear security requirements and comprehensive control processes for open-source and third-party software. This includes strict controls during selection, security testing, code security, risk scanning, legal review, application, and exit phases. For instance, during selection, open-source software must meet cybersecurity assessment requirements. When in use, third-party software should be integrated into the application system and evaluated in combination with self-developed software to ensure no new security issues arise.

  Additionally, when vulnerabilities in open-source or third-party software are disclosed, they must be promptly detected and fixed. These software components must be tested as part of the application system to verify that known vulnerabilities are addressed. The list of fixed vulnerabilities should be included in the application system's release notes.

- **Configuration and change management**

  Configuration and change management are crucial for maintaining application system security. Enterprises must manage configurations of all application systems, including extracting configuration models (item types, attributes, and relationships) and recording configuration details. Professional CMDB tools are used to manage these configuration items and their relationships.

  Changes to application systems, such as operating system, database, middleware, and application updates, can impact system security and stability. These changes must be managed through structured processes. After change requests are generated, the change manager assesses the change level and submits the requests to the change committee for approval, and then the changes can be implemented as planned. Before implementation, changes must be fully tested in staging environments, using techniques like gray release and blue-green deployment, to ensure the committee understands the actions, duration, rollback procedures, and potential impacts.

- **Security approval for rollout**

  To ensure that application systems comply with laws, regulations, and enterprise security specifications, and to minimize cybersecurity and privacy compliance risks, cloud security experts from the CCoE team must participate in application rollout activities. They work with application teams to analyze and verify if the related versions or services meet regional security and privacy compliance requirements.

  To facilitate quick rollouts of application systems with low and medium security and compliance risks, cloud security experts provide a security and privacy compliance self-check list. This list outlines the compliance requirements that enterprises must meet. Application teams use this list for self-checks during development, deployment, and rollout. Systems with medium and low risks can be rolled out after passing the self-check, with results submitted to cloud security experts for audit. High-risk systems require more resources for stricter detection and approval, ensuring both security and timely rollout.

7.4.5 Personnel Security Management

Enterprises need to manage the security of IT personnel who may access sensitive data. This includes security awareness education, capability training, key position management, and accountability for security violations.

- Security awareness education

  To enhance the information security awareness of all employees, avoid information security violations, and ensure normal business operations, enterprises can conduct security awareness education in three areas: general education, publicity activities, and commitment letter signing.

  - General education: Regularly organize cybersecurity awareness training, requiring employees to continuously learn about cybersecurity, understand relevant policies and regulations, and be aware of acceptable behaviors.
  - Publicity activities: Conduct various information security publicity activities for all employees, such as information security community operations, case studies, information security activity weeks, and animated promotions.
  - Commitment letter signing: Integrate information security into the employee conduct guidelines. Convey the company's information security requirements through annual routine learning, exams, and signing activities to improve employees' information security awareness. Employees sign the information security commitment letter, promising to comply with the company's information security policies and regulations.

- Security capability training

  Establish a comprehensive information security training system by referencing to industry best practices. Integrate various forms of security skills training into employee onboarding, on-the-job training, and promotion processes to enhance employees' security skills.

  - Basic cybersecurity training: Enterprises need to develop role-based training plans for foundational security skills. New hires must complete cybersecurity and privacy onboarding training and exams within their probation period. Current employees should select and complete relevant courses and exams based on their roles. Managers are required to participate in cybersecurity training and workshops.
  - Precise training: Identify typical security issues and responsible parties during product R&D through big data analysis, and provide targeted security training plans (including case studies, courses, and exercises) to continuously improve security quality.
  - Practical drills: Adopt industry best practices to develop an information security practical drill platform. Conduct red-blue team exercises and offer scenario-based drill environments for employees to practice and communicate, thereby enhancing their security skills and response capabilities.
  - C&Q guidance for security capabilities: To facilitate more conscious and effective cybersecurity learning, enterprises should integrate cybersecurity requirements into Competency and Qualification (C&Q) criteria. Employees must attend cybersecurity courses and pass exams before promotion to boost their cybersecurity capabilities.

- Key position management

  To ensure orderly internal management and reduce the impact of personnel risks on business continuity and security, enterprises should implement special management for key positions, such as O&M engineers. The details are as follows.

  - Onboarding security review: Conduct background and qualification checks on new hires to ensure they meet the company's information security requirements.
  - On-the-job security training: Provide cybersecurity training and exams based on awareness, service specifications, user data, and privacy protection. Update the training and exam content regularly to reflect service changes.
  - Onboarding qualification management: Key position employees must pass a cybersecurity exam and obtain a certificate. Issue a two-year valid e-Cert to those who pass, and remind them to retake the exam before the certificate expires.
  - Off-job security review: Conduct security reviews for transferring or departing employees, including account clearance or modification, according to the transfer and resignation checklists.

- Accountability for security violations

  Enterprises must establish a strict security responsibility system and implement an accountability mechanism for violations. Each employee is responsible for their actions and outcomes at work, including legal responsibilities. Security issues can significantly impact the enterprise, so accountability is based on behavior and results, regardless of intent. Accountability levels are determined by the nature of the violation and its consequences. If a legal violation occurs, the employee will be handed over to the authorities. Supervisors, both direct and indirect, are also accountable for inadequate management. The violator's attitude and cooperation during investigations will influence the severity of the punishment.

7.4.6 Cloud Native Security Service

Huawei Cloud offers a comprehensive suite of cloud native security services. These services are deeply integrated with Huawei's cloud platform, providing superior performance, elasticity, and ease of use. Additionally, Huawei's experience in security operations as a cloud service provider continually enhances the capabilities of these services. Enterprises are recommended to prioritize cloud native security services.

- Data Encryption Workshop (DEW)

  DEW is a cloud data encryption service. It provides Dedicated Hardware Security Module (Dedicated HSM), Key Management Service (KMS), Cloud Secret Management Service (CSMS), and Key Pair Service (KPS). DEW uses HSMs to protect your keys, and can be integrated with other Huawei Cloud services. Additionally, DEW enables customers to develop customized encryption applications.

  For details about DEW, see the **DEW Documentation**.

- Host Security Service (HSS)

  Host Security Service (HSS) is designed to protect server workloads. It protects your system integrity, enhances application security, monitors user operations, and detects intrusions. HSS can provide unified visualization and control capabilities for hosts and containers, no matter where they are located.

  For details about HSS, see the **HSS Documentation**.

- Web Application Firewall (WAF)

  WAF keeps web services stable and secure. It examines all HTTP and HTTPS requests to detect and block the following attacks: Structured Query Language (SQL) injection, cross-site scripting (XSS), web shells, command and code injections, file inclusion, sensitive file access, third-party vulnerability exploits, Challenge Collapsar (CC) attacks, malicious crawlers, and cross-site request forgery (CSRF). You can start to use it by adding a website on the WAF console. If WAF is enabled, all public traffic to the website will first go through WAF. Malicious traffic will be detected and filtered by the WAF, while normal traffic will be returned to the source IP, ensuring the safety, stability, and availability of the source IP.

  For details about WAF, see the **WAF Documentation**.

- Database Security Service (DBSS)

  DBSS is an intelligent database security service. Based on the machine learning mechanism and big data analytics technologies, it can audit your databases, detect SQL injection attacks, and identify high-risk operations. It provides functions such as user behavior detection and audit, multi-dimensional analysis, real-time alarms, refined reports, sensitive data protection, and audit log backup. Database audit provides you with the database audit function in out-of-path pattern, enabling the system to audit risky behaviors in real time and generate alarms. In addition, database audit generates compliance reports that meet data security standards. These measures help you find the person accountable for internal violations and improper operations.

For details about DBSS, see the **DBSS Documentation**.

- Cloud Firewall (CFW)

  CFW is a next-generation cloud native firewall that provides protection for Internet and VPC borders on the cloud. It supports on-demand elastic capacity expansion and provides basic network security protection for services migrated to the cloud. It provides border protection between VPCs, access control policies, intrusion prevention policies, antivirus, traffic analysis, and system management.

  For details about CFW, see the **CFW Documentation**.

- Data Security Center (DSC)

  DSC is a next-gen cloud-based data security platform that offers basic data security capabilities such as data classification, risk identification, watermark tracing, and static data desensitization. It integrates the various stages of the data security lifecycle and presents an overall view of the cloud-based data security situation.

  For details about DSC, see the **DSC Documentation**.

- SecMaster

  SecMaster is a next-generation cloud native platform that enables integrated and automatic security operations. You can manage cloud assets, security posture, security information, and incidents in one place and enjoy intelligent threat detection, easy security orchestration, and automatic response.

  For details about SecMaster, see the **SecMaster Documentation**.

- Anti-DDoS Service (AAD)

  AAD provides powerful protection for the continuity of important enterprise services. It can protect your servers against large volumetric DDoS attacks so your services can be reliable and stable. AAD protects your mission-critical workloads from DDoS attacks by routing all traffic destined for origin servers to AAD IP addresses and scrubbing malicious attacks. This service can be deployed on hosts used on Huawei Cloud, other clouds, and on-premises data centers.

  For details about AAD, see the **AAD Documentation**.

- Cloud Certificate Manager (CCM)

  CCM is a cloud service that provides one-stop lifecycle management of digital certificates. It includes the SSL Certificate Manager (SCM) and Private Certificate Authority (PCA) services.

  For details about CCM, see the **CCM Documentation**.

- **CodeArts Inspector**

  CodeArts Inspector scans vulnerabilities on your websites, hosts, mobile applications, software packages, and firmware. It provides vulnerability assessments, customized scanning, and vulnerability lifecycle management. After a scan is complete, a scan report is generated for you to check vulnerability details and solutions.

- Cloud Bastion Host (CBH)

  CBH is a unified security management and control platform. It provides account, authorization, authentication, and audit management services that enable you to centrally manage cloud computing resources. CBH provides various functional modules, such as department, user, resource, policy, operation, and audit modules. It integrates functions such as single sign-on (SSO), unified asset management, multi-terminal access protocols, file transfer, and session collaboration. With the unified O&M login portal, protocol-based forward proxy, and remote access isolation technologies, PBH enables centralized, simplified, secure management and maintenance auditing for cloud resources such as servers, cloud hosts, databases, and application systems.

For details about CBH, see the **CBH Documentation**.

7.5 FinOps

7.5.1 Overview

According to the *State of the Cloud Report* released by Flexera in 2024, managing cloud costs is the top challenge for enterprises. The average cloud cost of enterprises exceeds the budget by 15%, and 27% of public cloud costs are wasted. 51% of enterprises have established dedicated FinOps teams, and 20% plan to establish FinOps teams in the next year.

As enterprises increasingly leverage the agility, efficiency, innovation, and elastic scalability of the cloud, they face the following four difficulties in cloud cost management:

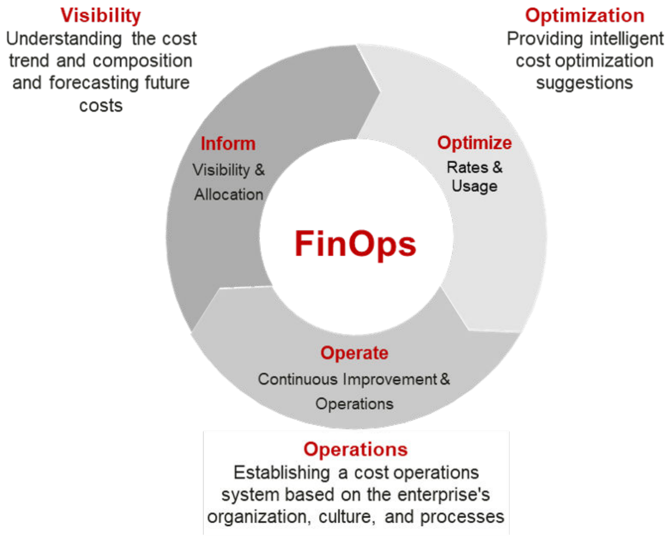
- **Difficult cost planning:** Traditionally, IT costs are fixed after procurement. However, cloud resources are used on demand and dynamically, so cloud costs change with services. For example, cloud resource usage increases during peak hours, and new resources are dynamically provisioned during upgrade and capacity expansion. The dynamic cloud costs cause a large deviation between the budget and the actual spending.
- **Difficult cost control:** Traditionally, IT procurement is centrally managed by the procurement department, which is easy to control. However, cloud resource consumption runs through the entire cloud usage process, and the procurement becomes decentralized. Resources are purchased directly by engineers instead of by procurement personnel. Engineers' weak cost awareness and the large number of people purchasing cloud resources make it difficult to control cloud costs.
- **Difficult cost optimization:** Cloud service providers usually provide hundreds of cloud services and diversified billing units, and there is no unified optimization solution across services. In addition, new services, instance types, and discounts continuously emerge. This makes it difficult for enterprises to optimize costs.
- **Difficult refined management:** While flexible expansion and little expenditure limitation of the cloud facilitate innovation, they also cause resource waste. For example, the service team may configure more resources than required for running the workload to achieve high performance and quality, or forget to dispose of resources added for specific projects, resulting in idle resources.

  When enterprises struggle with finding a cost optimization approach or sustaining optimization effects, FinOps comes into play.

  FinOps is the combination of "Finance" and "DevOps". It encourages communication and collaboration between business teams and engineering teams (IT teams) to solve enterprises' cloud cost management problems. According to **FinOps Foundation**, "FinOps is an operational framework and cultural practice which maximizes the business value of cloud and technology, enables timely data-driven decision making, and creates financial accountability through collaboration between engineering, finance, and business teams."

Cloud cost management needs continuous optimization as enterprise cloud resource consumption runs through the entire cloud transformation process. The **FinOps framework** consists of three phrases: cost visibility, cost optimization, and continuous cost operations, as shown in the following figure. Note that you need to balance cost, quality, and efficiency during cost optimization to prevent extremely low costs from affecting business efficiency and stability.

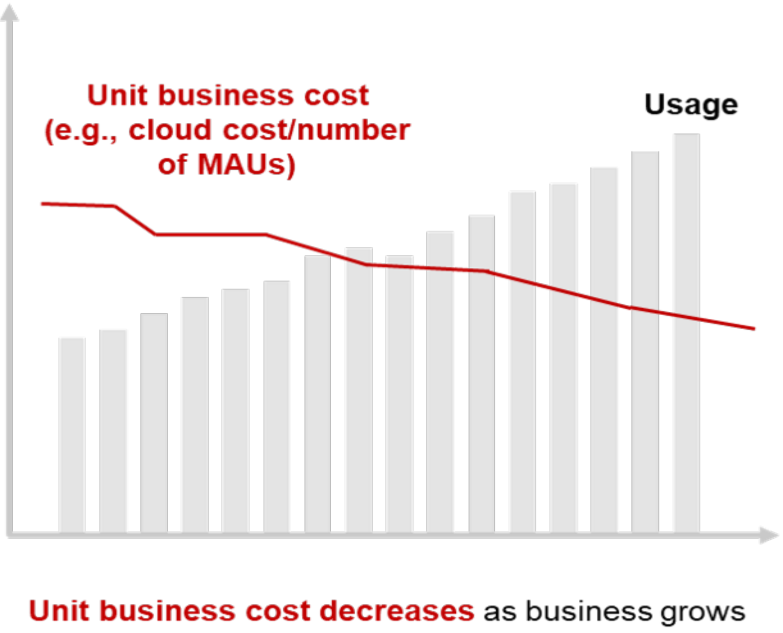
Figure 7-5 Three phases of FinOps


Visibility
Understanding the cost trend and composition and forecasting future costs
Optimization
Providing intelligent cost optimization suggestions
Inform
Visibility & Allocation
Optimize
Rates & Usage
FinOps
Operate
Continuous Improvement & Operations
Operations
Establishing a cost operations system based on the enterprise's organization, culture, and processes

The FinOps framework guides enterprises to build a cost operations system from considering the organization, culture, and processes. Multi-team collaboration and data-based decision-making are used to manage cloud costs in a refined manner. Costs of business service team are visible, and overspending and waste are proactively controlled. Enterprises make data-based decisions on cloud investment to ensure the expenditure of core and strategic businesses. With FinOps, enterprises can continuously reduce the **unit business cost**.

For more information about FinOps, visit the **FinOps Foundation website**.

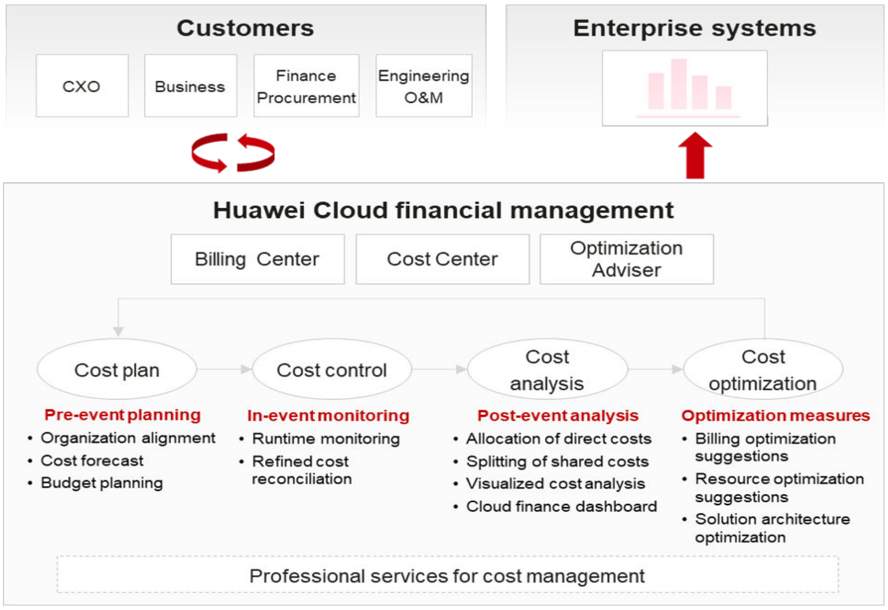
Figure 7-6 Decreasing unit business cost



7.5.2 FinOps Reference Architecture

Huawei Cloud has developed a set of FinOps references based on the FinOps Foundation's framework and Huawei's own practices. The references mainly cover four phases: cost planning, control, analysis, and optimization, as shown in the following figure.

Figure 7-7 Huawei Cloud FinOps reference architecture



Huawei Cloud also provides various cost management tools to help enterprises practice FinOps on Huawei Cloud and continuously improve cloud cost-effectiveness.

Table 7-1 Huawei Cloud cloud cost management tools

Category	Tool	Scenario
Cost planning	**Price calculator**	Estimating the cost of migrating any combination of new services to the cloud
	Cost and usage forecasting	Forecasting future expenditures of installed base services
	Enterprise organization	Isolating resources and costs for accounts created by large enterprises and group companies
	Enterprise project	Grouping costs based on the organizational structure of small- or medium-sized enterprises which may use individual accounts
	Cost tag	Grouping costs in a fine-grained, more flexible manner, with multiple dimensions supported
	Budget management (including budget reports)	Budgeting or tracking cloud expenditures at different granularities, receiving budget overrun notifications in a timely manner, and regularly learning your latest expenditures
Bill management and cost control	Fund management	Monitoring available amount and tracking unusual expenditures
	Resource package management	Checking whether a resource package is about to be used up
	Bill management	Learning the actual expenditures for each billing cycle and performing multi-dimensional reconciliation
	Cost monitoring	Identifying unexpected cost spikes and reducing cost anomalies
Cost allocation and visualization	Cost analysis (including analysis reports)	Learning cost trends and distributions and gaining insight into the drivers of cost changes
	Cost category	Allocating costs to meaningful categories and splitting shared costs based on your business needs

Category	Tool	Scenario
	Cost details	Obtaining cost allocation details (download and OBS subscription)
	Bill storage	Storing bill details in an OBS bucket
	Cloud intelligence dashboard (CXO dashboard)	Dashboard collection provided for CXOs, executives, and organization managers
Cost savings and optimization	Optimization option of changing pay-per-use to yearly/monthly	Finding cost-savings opportunities for pay-per-use resources
	Resource package usage/coverage analysis	Learning the usage of purchased resource packages
	Resource package recommendations	Getting recommendations for purchasing resource packages
	Resource optimization	Identifying idle resources
	Optimization Advisor	Identifying underutilized or idle cloud resources

7.5.3 Cost Planning

- **Aligning with the enterprise management model for future cost traceability**

  A good cost accountability system is the basis of cloud financial management. It ensures that departments, business teams, and owners are accountable for their respective cloud costs.

  Huawei Cloud provides you with a wide range of tools to help you more easily plan and track your organization's costs.

If you have a large-sized enterprise or corporate group, you can create a hierarchical organization with multiple accounts for unified accounting management. For details, see **Enterprise Management and Organizations**. With Enterprise Center enabled, you can isolate resources and costs on an account-by-account basis, helping you quickly expand business.

If you have a small- or medium-sized enterprise or you are using a standalone account, you can enable **Enterprise Project Management Service (EPS)** to manage personnel, expenditures, resources, and permissions in the cloud based on the organizational structure and service management model of your enterprise. If you have other multi-dimensional and fine-grained planning requirements, you can use **costs tags** as a supplement to organization planning. For example, you can use a specific tag to identify the team and owner of a specific resource.

You can use any of these tools to complete organization planning as early as possible and keep your formulated plan consistently executed throughout the operations of your enterprise.

- **Estimating future costs through intelligent forecast**

  As enterprises move more workloads to the cloud, their cloud expenditures continue to grow rapidly. Estimating future cloud costs becomes critical to planning. There are two types of future cloud costs: costs continuously generated by services running on the cloud and costs generated by new services, such as in new cloud migration or regional expansion projects. Cloud expenditures are variable. There is no one-size-fits-all prediction method, but you can predict costs based on historical expenditures and service drivers, such as cloud migration of service rollout or regional expansion. This will help you get a relatively accurate result.

  You can access Cost Center and enable cost analysis to predict future costs based on your historical expenditures. **Cost analysis** of Cost Center considers different billing model characteristics and uses machine learning and rule-based models to predict costs and usage of all consumption models.

  You can also use Huawei Cloud **Price Calculator** to estimate costs for your desired services or any combination of services based on their usage required for service rollout or regional expansion.

- **Tracking future resource usage and expenditure through budget management**

  Budget overruns are one of the crucial challenges facing cloud cost management. After you make organizational and cost plans, you can make a budget for each business team and track their budget execution. By tracking budgets, each team's budget execution progress becomes visible. The difference between actual costs and budgets is controlled in a timely manner to avoid budget overrun.

  Huawei Cloud provides **budget management** for you to track your costs by product, team, or project in a refined manner.

  In addition to viewing the budget progress in Cost Center, you can also configure budget alerts. If the actual or predicted usage reaches the specified threshold, you will receive an alert via SMS or email so that you can take measures in a timely manner.

  You can create a daily, weekly, or monthly **budget report** to inform relevant personnel so that the costs are always visible to the relevant business departments, finance departments, and CTO of your enterprise.

7.5.4 Cost Control

- **Budget management**

  Cloud costs change during the use of the cloud because cloud resources can scale flexibly on demand. To avoid unexpected bills and high costs, you need to control cloud resources precisely during cloud use, establish risk monitoring and warning mechanisms to promptly respond to risks. When there is an unusual spending, it is important to analyze the root cause, such as business development, resource over-allocation, or resource idleness. Increase the budget or optimize resource utilization based on the root cause.

  You can use **cost anomaly detection** provided by Cost Center to identify any unexpected cost spikes in a timely manner.

  Cost Anomaly Detection uses machine learning to analyze your historical expenditures, establish a baseline, and identify root causes for cost surprises.

  You can configure alerts to keep you up to date on any cost anomalies, so you can quickly take action based on identified potential causes, service conditions, and cost analyses.

- **Tailored reconciliation**

  Huawei Cloud offers a diverse set of charts to help you quickly understand bills and make multi-dimensional reconciliation.

  Huawei Cloud offers **bills** across various dimensions. This helps you track your overall expenditure trend and identify the most costly service types, enterprise projects, regions, and billing modes. With this information, you can quickly evaluate whether your monthly expenditures align with your budget. You can also check transactions and detailed bills, detailed fund transactions and resource deductions of each month, total expenditure, payment details, and bill details.

  You can export your Huawei Cloud bills for reconciliation with the bills generated by your company to check for resource purchase and usage.

- **Funds monitoring and resource package alerting**

  To keep costs down, Huawei Cloud monitors the account balance and resource packages usage to help you evaluate whether there are any expenditure spikes.

  You can set balance alerts to receive timely warnings about insufficient balance and monitor expenditure amounts.

  If you have purchased a **resource package**, you can learn about the remaining usage of the resource package in a timely manner. You can enable usage alerts so that you can be notified before the resource package is used up. This way, you can avoid unexpected expenditures by limiting pay-per-use usage or by purchasing new resource packages in a timely manner.

7.5.5 Cost Analysis

Cost Allocation Facilitates Financial Accountability

Cost allocation enables enterprises to allocate costs to business teams, making costs of each business team clearly visible. Based on clear costs, business departments can provide cost-effective solutions with properly priced products and a balance among costs, stability, and performance. Enterprise managers can make informed decisions on cloud expenditures of each business, ensuring that expenditures of core and strategic businesses are within the budget and no waste occurs. Typical cost allocation scenarios include:

- Project-based cost allocation, such as to innovation projects and expansion projects
- Department-based cost allocation, such as to R&D departments and test departments
- Brand-based cost allocation, such as to the R&D, manufacturing, and store sales of a specific brand
- System-based cost allocation, such as to the IT systems shared by multiple businesses

Cost Allocation

Huawei Cloud provides multiple cost allocation capabilities.

- **Direct cost allocation:** Enterprises can allocate costs based on the account, enterprise project, or cost tag associated with the resource when the costs are generated.
- **Cost allocation for special products:** The costs of special products such as yearly/monthly subscriptions and resource packages are **amortized** to enterprise projects, tags, and resources based on actual usage.
- **Usage-based shared cost splitting for traffic resources:** Cost Center allows you to split the costs of shared cloud resources such as CDN and Live to domain names or IP addresses based on the actual usage.
- **Splitting container cluster costs:** Huawei Cloud provides **CCE cost insights**. After this function is enabled, the CCE cluster management fee and the fees of ECS and EVS resources associated with the CCE cluster can be split to clusters, namespaces, and workloads.
- **Multi-dimensional categorization and summary:** You can also use **cost categories** to categorize and summarize costs with multiple conditions (product type, bill type, linked account, enterprise project, cost tag, brand, and subsystem)
- **Shared cost allocation:** You can define **cost categories** to allocate shared costs proportionally within your organization, for example, the costs of shared resources and platform services, and untagged costs. This makes it easier to fairly allocate costs to different teams or business departments.

Making Cost Analysis from Various Dimensions to Explore Costs and Usage

Enterprises perform refined analysis on cloud costs to understand the cost structure and trend to examine:

- Do costs increase with the business growth? Is the increase trend stable?
- Which cloud services have the highest expenditure? Which cloud services have the fastest expenditure growth?
- Is the pricing mode reasonable? Are new, cost-effective resources used?
- Is the resource deployment reasonable? Are low-cost region resources used?
- What caused cost fluctuation?

Huawei Cloud **Cost Center** provides you with multi-dimensional summary and filtering mechanisms to help you analyze costs and usage trends, and their drivers. Considering common cost data filtering conditions, Cost Center has preconfigured the following reports for quick analysis.

Table 7-2 Preconfigured reports in Cost Center

Report Name	Description
Multi-Dimensional MTD Costs	Shows your MTD original costs grouped in various different ways, helping you learn about your cost breakdowns and flows.
Monthly Costs by Service Type	Shows the monthly costs by service type. You can learn which types of services have had the highest original costs over the last six months.
Monthly Amortized Costs	Shows the monthly costs amortized over the last six months.
Daily Costs	Shows the daily original costs over the last three months and in the following one month.
Monthly Costs by Linked Account	Shows the monthly costs by linked account. You can learn the linked accounts with highest original costs over the last six months.
Monthly Costs by Enterprise Project	Shows the monthly original costs for each enterprise project over the last six months.
Monthly Costs by Region	Shows the monthly original costs for each region over the last six months.
Pay-Per-Use ECS Monthly Costs and Usage	Shows the monthly original costs and usage of pay-per-use ECSs over the last six months.
Container Cost Insights	Shows the cost breakdowns and trends of CCE clusters by namespace or workload.

7.5.6 Cost Optimization

- **Selecting appropriate billing mode**

  Huawei Cloud provides you with flexible billing modes, such as pay-per-use, yearly/monthly, resource package, and spot pricing. You can choose whichever is best suited to your enterprise needs. The following describes the application scenario for each billing mode, so you can more easily select the one most likely to help you effectively reduce your expenditures:

  Pay-per-use: You may only need to use cloud resources temporarily or for unplanned, urgent purposes.

  Yearly/monthly: You want to cut down costs by choosing a long-term subscription, and your demand is stable during the subscription term.

Resource package: Similar to a yearly/monthly subscription, you can buy a resource package in advance to cover your usage of certain types of resources that you want to use in a long term.

Spot pricing: You do not have high requirements on service stability, and you can tolerate service interruption. This billing mode is currently available only for ECSs.

- **Optimizing your billing modes**

  Huawei Cloud provides optimization recommendations on billing modes to help you reduce costs by changing to a more appropriate billing mode while maintaining resource performance.

  Changing pay-per-use to yearly/monthly: Cost Center helps you analyze the usage of your pay-per-use resources (such as ECS, EVS, RDS) and identify places where you can save money by changing the billing mode from pay-per-use to yearly/monthly. You are advised to focus on optimization recommendations with high savings (large value for **Estimated Monthly Savings**) and low risks (small value for **Break-Even Time**).

- **Identifying idle and underutilized resources**

  Huawei Cloud monitors your historical expenditures and resource usage to identify idle resources, for example, idle cloud servers, and then gives you **resource optimization recommendations**. You can decide whether to adopt the recommendations based on the resource utilization, estimated monthly savings, and opinions of service teams.

  Huawei Cloud Optimization Advisor (OA) provides cost inspection to help you quickly and accurately identify existing risks based on the provided optimization suggestions.

- **Architecture optimization and continuous operations**

  FinOps professional services can optimize the architecture from the service layout, resource planning, and data storage dimensions based on enterprise scenarios. For example, hybrid deployment of online and offline services improves resource utilization; decoupling of storage and compute enables on-demand use of compute and storage resources to avoid waste; and cold and hot data separation reduces cold data storage costs.

7.6 Continuous Optimization

O&M governance is a continuous improvement cycle. You need to periodically review and evaluate the cloud environment based on the Well-Architected Framework (WAF) and make adjustments and optimizations based on service requirements and WAF best practices.

You also need to continuously learn and apply new Huawei Cloud services and functions to continuously improve the maturity of the cloud environment.

By combining the five pillars of WAF with lean governance, deterministic operations, comprehensive security operations, and refined FinOps, you can build an evolving, secure, reliable, high-performance, and cost-effective cloud environment to better support business development.