

# Cloud Computing Basics



## Foreword

- Enterprises in the fast-growing ICT industry are posing higher requirements on compute, storage, and networking resources. Within this context, a new architecture - cloud computing - has emerged to meet the need for on-demand resources and accelerate business innovation.

# Objectives

- Upon completion of this course, you will be able to:
  - Understand what cloud computing is, what runs on the cloud, and what you can achieve with the cloud.
  - Understand cutting-edge cloud technologies, future cloud trends, application scenarios, and cases.
  - Understand the benefits and future breakthroughs of cloud computing.

# Contents

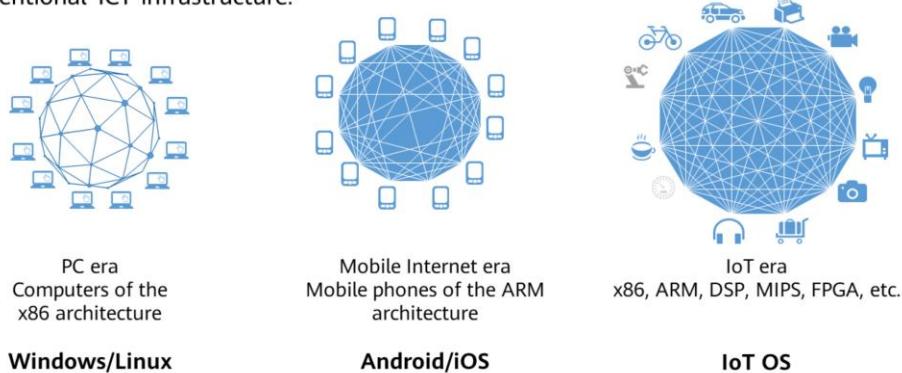
## **1. Cloud Computing Basics**

- Background of Cloud Computing
- Definition of Cloud Computing
- Cloud Computing Is Around Us
- Cloud Computing Models
- Benefits of Cloud Computing

## **2. Cloud Computing Technologies**

# The Information Explosion Is Coming

- With the prevalence of the mobile Internet and fully connected era, more terminal devices are being used and data is exploding every day, posing unprecedented challenges on conventional ICT infrastructure.



5      Huawei Confidential



- The PC era is essentially in which computers are networked, and personal computers are connected through servers. Now, in the mobile era, we can access the Internet through mobile phones. With the advent of 5G, all computers, mobile phones, and intelligent terminals can be connected, and we can enter an era of Internet of Everything (IoE).
- In the IoE era, the entire industry will compete for ecosystem. From the PC era to the mobile era, and to the IoE era, the ecosystem experiences fast changes at the beginning, then tends to relatively stable, and rarely changes when it is stable. In the PC era, a large number of applications run on Windows, Intel chips, and x86 architecture. Then, browsers come with the Internet. In the mobile era, applications run on iOS and Android systems that use the ARM architecture.
- Compared with the previous generation, the number of devices and the market scale of each generation increase greatly, presenting future opportunity. As the Intel and Microsoft in the PC era and the ARM and Google in the mobile era, each Internet generation has its leading enterprises who master the industry chain. In the future, those who have a good command of core chips and operating systems will dominate the industry.

## Challenges Faced by Conventional IT Architecture

- The Internet era has brought a large amount of traffic, users, and data to enterprises, but conventional IT architecture cannot meet the requirements for rapid enterprise development.



- The Internet brings a large amount of traffic, users, and data, so enterprises need to continually purchase traditional IT devices to keep pace with their rapid development. Therefore, the disadvantages of traditional IT devices gradually emerge.
  - Long procurement period causes slow rollout of new business systems.
  - The centralized architecture has poor scalability and can only increase the processing performance of a single node.
  - Traditional hardware devices exist independently, and their reliability depends only on software.
  - Devices and vendors are heterogeneous and hard to manage.
  - The performance of a single device is limited.
  - The utilization of devices is low, while the total cost remains high.

## Discussion:

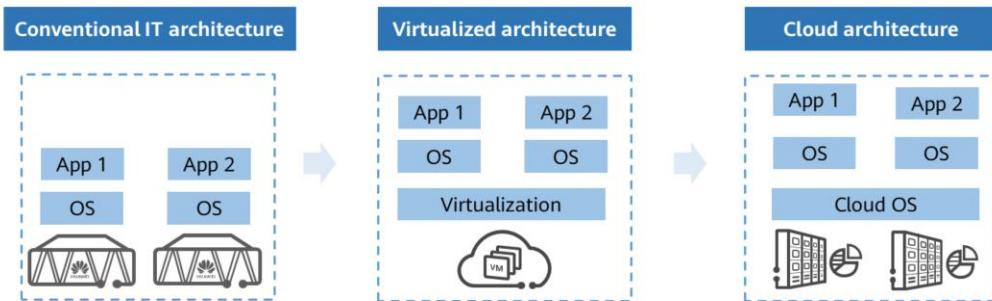
- How can IT enterprises overcome these challenges?

- **IT infrastructure transformation**
- **Resource integration and comprehensive utilization**
- **Business collaboration and continuous optimization**



- Let's discuss
  - How to solve these pain points? Think over advantages of cloud computing that can solve these pain points, so you can have a better understanding of cloud computing.

## Enterprises Are Migrating To the Cloud Architecture



- The traditional IT architecture consists of hardware and software, including infrastructure, data centers, servers, network hardware, desktop computers, and enterprise application software solutions. This architecture requires more power, physical space, and capital, and is usually installed locally for enterprises or private use.
- With the virtualization technology, computer components run on the virtualization environment, not on the physical environment. Virtualization enables maximum utilization of the physical hardware and simplifies software reconfiguration.
- With cloud transformation, enterprise data centers are transformed from resource silos to resource pooling, from centralized architecture to distributed architecture, from dedicated hardware to software-defined storage (SDS) mode, from manual handling to self-service and automatic service, and from distributed statistics to unified metering.

# Contents

## **1. Cloud Computing Basics**

- Background of Cloud Computing
- **Definition of Cloud Computing**
- Cloud Computing Is Around Us
- Cloud Computing Models
- Benefits of Cloud Computing

## **2. Cloud Computing Technologies**

## Cloud Computing Definition

- Cloud computing is a model for enabling ubiquitous, convenient, **on-demand network access** to a **shared pool** of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be **rapidly provisioned** and released with **minimal management** effort or service provider interaction.

--National Institute of Standards and Technology (NIST)

- Simply put, the term "cloud" is a metaphor for networks and the Internet, and refers to an abstraction of the Internet and the infrastructure that underpins it. Computing refers to computing services provided by a powerful computer with a range of functionalities, resources, and storage. As such, cloud computing can be viewed as the delivery of on-demand, measured computing services over the Internet.

- Cloud computing has the following characteristics:
  - Broad network access
  - Fast and elastic scaling
  - On-demand self-service
  - Resource pooling
  - Metered services

# Contents

## **1. Cloud Computing Basics**

- Background of Cloud Computing
- Definition of Cloud Computing
- **Cloud Computing Is Around Us**
- Cloud Computing Models
- Benefits of Cloud Computing

## **2. Cloud Computing Technologies**

## Cloud Services and Applications Around Us (Individuals)



Cloud albums



Cloud music



Cloud videos



Cloud documents

- What are the data sources of cloud computing in daily life?
  - Cloud album, such as Baidu Cloud and iCloud Shared Album
  - Cloud music, such as NetEase Cloud Music, Kugou Music, Kuwo Music, and Xiami Music
  - Cloud video, such as Baidu Cloud and Tencent Cloud Video
  - Cloud documents, such as Youdao Note, and Shimo document
- These apps are based on cloud computing, making our lives more convenient. Cloud computing allows enterprises to provide better products to attract more users.

## Cloud Services and Applications Around Us (Enterprises)

- HUAWEI CLOUD Meeting allows interactive video and voice communications between people in two or more locations.



Videoconference



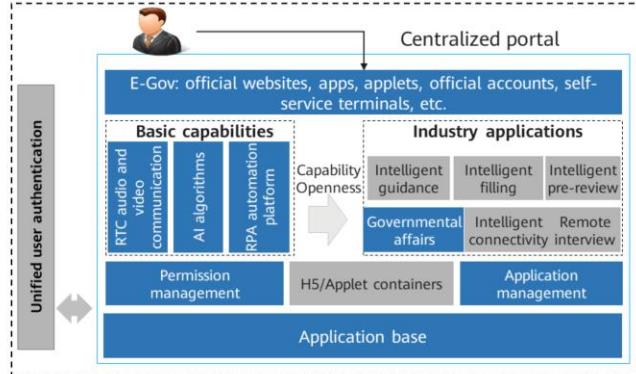
Livestream

- Driven by the requirements of the government, transportation, electric power, medical care, education, finance, and military industries and enterprises, the video conferencing market in China has an average annual growth beyond 20%. Currently, only less than 5% of enterprises in China have video conference rooms, and more and more enterprises are aware of the importance of efficient collaboration. Therefore, the video conferencing system has become indispensable for efficient office work.
- HUAWEI CLOUD Meeting can be used by enterprise office, telemedicine, smart education, and enterprise organization construction.

## E-Gov Cloud - Online Services

- The e-Government cloud enables access to comprehensive public services through the Internet and serves as an extensive service platform with software, application, and information resources provisioned for governmental bodies. It utilizes the compute, storage, network, security, and application resources in existing equipment rooms and leverages cloud computing to offer high reliability, availability, and elasticity.

- 1 **24/7 e-Gov services:**  
Public services are available to citizens and enterprises online.
- 2 **One-stop shop for all services**  
The e-Gov cloud allows information sharing and makes collaborative approval possible.
- 3 **Virtual lobby managers**  
Virtual lobby managers are always ready to provide assistance.
- 4 **AI & RPA robots:**  
AI robots assist with intelligent pre-review and RPA robots assist with system data synchronization, reducing the pressure on staff and improving service efficiency.



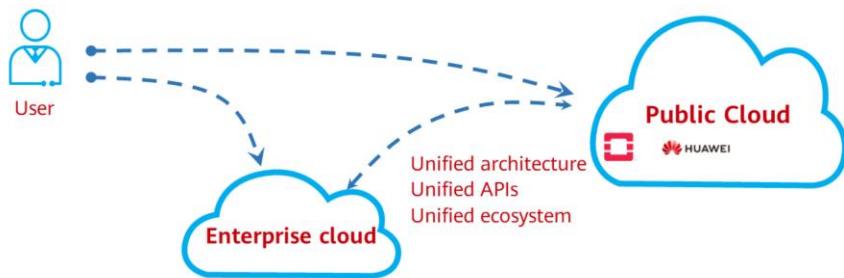
14 Huawei Confidential



- Online services are the most typical application of e-Gov and are used in multiple cities. With online services, applicants fill in the application information and submit supporting documents. The administrative approval center approves applications by streamlining cross-department data. By moving services to the cloud, governments can greatly reduce expenditure, cloud service providers can gain new development benefits, and citizens can acquire services more and more conveniently.
  - Guidance: All policies, bulletins, and processes are released through information guidance, making service handling processes clear for citizens and enterprises. Intelligent Process Automation (IPA) robots are provided to guide users.
  - Handling: Based on big data and AI technologies, fields in documents can be auto-populated.
  - Review: AI technologies pre-review documents, improve review efficiency and quality, and reduce pressure on staff. Real-Time Communication (RTC) audio and video technologies can implement contactless online pre-review.
  - Collaboration: RPA technology handles all work items through the workbench, effectively collaborating with functional agencies. All application and service entries can be managed in a unified manner.

## Public Cloud

- Simply put, the public cloud enables IT resources to be as accessible as electricity and water through the Internet.



- Public cloud is the main form of cloud computing, which is developing considerably in China. Public cloud vendors can be classified as follows:
  - Traditional telecom infrastructure carriers, including China Mobile, China Unicom, and China Telecom
  - Local government cloud computing platforms
  - Public cloud platforms built by Internet giants, such as Alibaba Cloud and Tencent Cloud
  - Some IDC carriers, such as 21Vianet Group
  - Enterprises with foreign technical background or introducing foreign cloud computing technologies, such as Fengqi.Asiatic Cloud

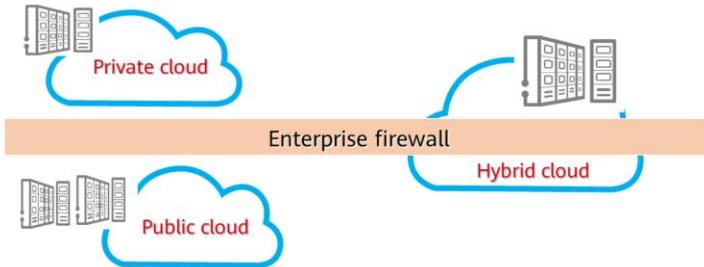
# Contents

## **1. Cloud Computing Basics**

- Background of Cloud Computing
- Definition of Cloud Computing
- Cloud Computing Is Around Us
- **Cloud Computing Models**
- Benefits of Cloud Computing

## **2. Cloud Computing Technologies**

## Deployment Models for Cloud Computing



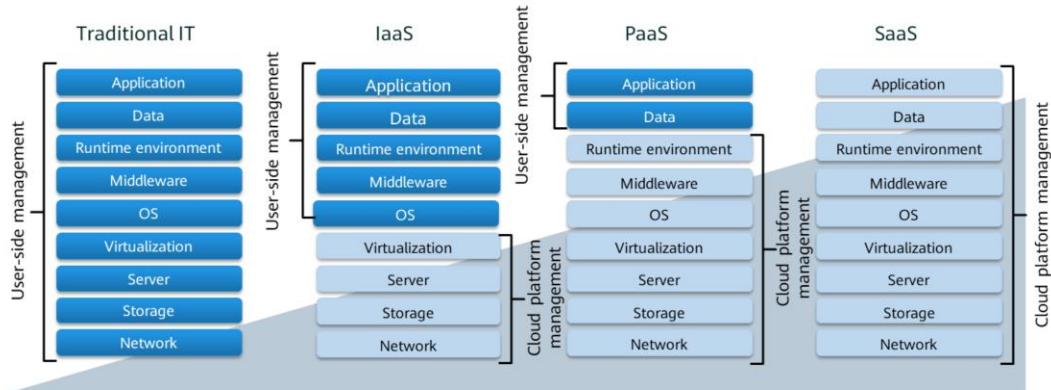
**Private cloud:** The cloud infrastructure is provisioned for exclusive use by a single organization.

**Public cloud:** The cloud infrastructure is owned and managed by a third-party provider and shared with multiple organizations using the public Internet.

**Hybrid cloud:** This is a combination of public and private clouds, viewed as a single cloud externally.

- Private cloud is a cloud infrastructure operated solely for a single organization. All data of the private cloud is kept within the organization's data center. Attempts to access such data will be controlled by ingress firewalls deployed for the data center, offering maximum data protection.
- Public cloud service provider owns and operates the cloud infrastructure and provides cloud services open to the public or enterprise customers. This model gives users access to convenient, on-demand IT services, comparable to how they would access utilities like water and electricity.
- A hybrid cloud is a combination of a public cloud and a private cloud or on-premises resources, that remain distinct entities but are bound together, offering the benefits of multiple deployment models. Users can migrate workloads across these cloud environments as needed.

# Service Models for Cloud Computing



- Infrastructure as a Service (IaaS): The cloud platform provides infrastructure (such as servers, storage devices, networks, and virtual resources) and maintains related resources. Users only need to pay attention to the system and application layers.
- Platform as a Service (PaaS): The cloud platform provides infrastructure (such as servers, storage devices, networks, and virtual resources) and application deployment environment (such as the operating system, middleware, and software running environment) and maintains related resources. Users only need to focus on applications and data.
- Software as a Service (SaaS): The cloud platform provides all resources, services, and maintenance. Users only need to use applications.
- Compared with the conventional IT entire-process and all-device procurement mode, the cloud service-oriented mode provides IT devices as services that allow customers to select on demand, which has more advantages in flexibility, and low cost.

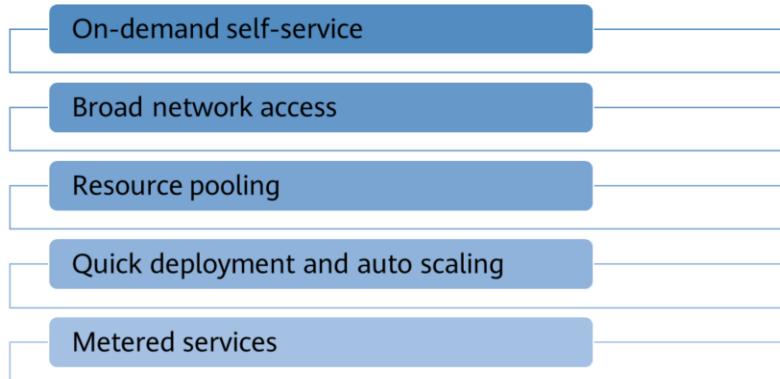
# Contents

## **1. Cloud Computing Basics**

- Background of Cloud Computing
- Definition of Cloud Computing
- Cloud Computing Is Around Us
- Cloud Computing Models
- **Benefits of Cloud Computing**

## **2. Cloud Computing Technologies**

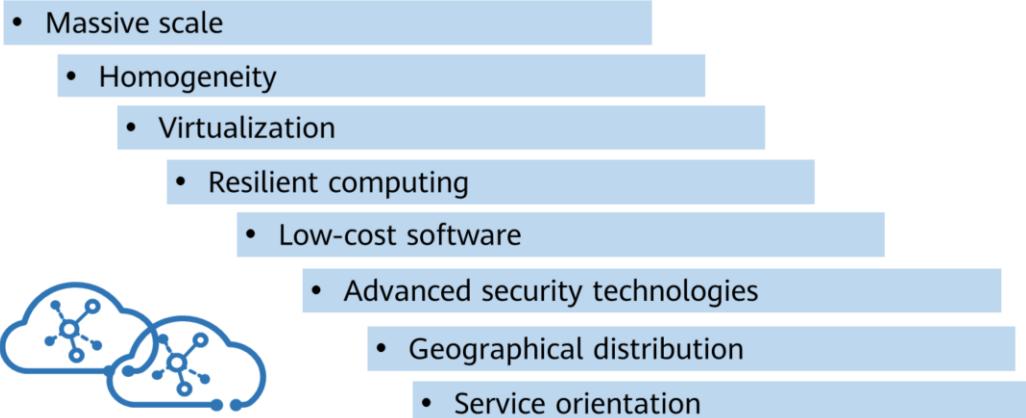
## Benefits of Cloud Computing



- Cloud computing integrates hardware resources into one in software, and dynamically allocates the resources to applications in software, which greatly improves the resource usage. Cloud computing also supports auto scaling, and optimizes the work efficiency. By building high-specification cloud data centers and introducing the automatic scheduling technology, data storage is more centralized, and data assets are more effectively used, achieving energy saving, emission reduction, and easier maintenance. In this way, lower costs and higher efficiency are achieved in each dimension.
- Cloud computing has five main benefits.
  - On-demand self-service: Customers can deploy processing services based on actual requirements on the server running time, network, and storage, and do not need to communicate with each service provider.
  - Broad network access: Various capabilities can be obtained over the Internet, and the Internet can be accessed in standard mode from various clients, such as mobile phones, laptops, and PADs.
  - Resource pooling: Computing resources of the service provider are centralized so that customers can rent services. In addition, different physical and virtual resources can be dynamically allocated and reallocated based on the customer requirements. Customers generally cannot control or know the exact location of the resources. The resources include the storage devices, processors, memory, network bandwidth, and virtual machines.
  - Quick deployment and auto scaling: Cloud computing can rapidly and elastically provide computing capabilities. A customer can rent unlimited resources and purchase required resources at any time.
  - Metered services: Cloud services are billed based on the actual resource usage, such as the CPU, memory, storage capacity, and the bandwidth

consumption of cloud servers. Cloud services provide two billing modes: pay-per-use and yearly/monthly.

## Eight Common Characteristics of Cloud Computing



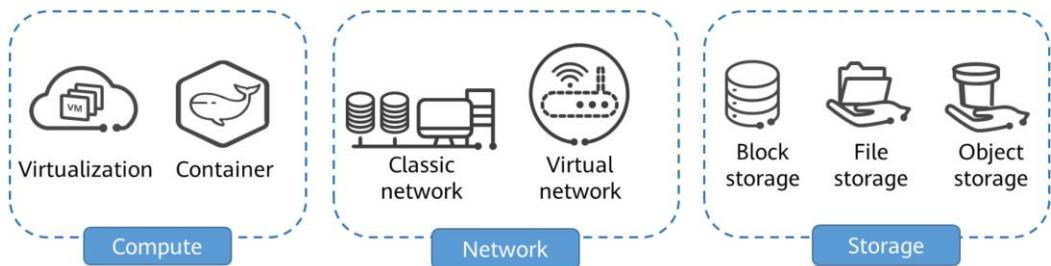
- Massive scale: Cloud computing service is in large scale as it centralizes IT resource supply. This makes cloud computing different from conventional IT.
- Homogeneity: Homogeneity can also be understood as standardization, which is similar to power utilization. Voltage and socket interface should be the same for various electrical appliances and devices.
- Virtualization: Virtualization has two meanings. One is accurate computing units. If a cake is too large for one person, it is better to divide it into small pieces to share. That is, with smaller computing units, IT resources can be fully used. The other meaning is the separation of software and hardware. Before virtualization, software and specified hardware are bound together, and after virtualization, software can be freely migrated on all hardware, which is like renting a house instead of buying one.
- Elastic computing: Elastic computing means that IT resources can be elastically provided.
- Low-cost software: Low-cost software is provided to meet the market competition and requirements. Cloud computing, with low individual technical skill and financial requirements, makes IT easy to use. Small and micro startups are always willing to enjoy the more IT services at the lowest cost. Based on this situation, low-cost software is required to earn money at small profits but quick turnover.
- Geographic distribution: As the broad access mentioned above, IT services can be provided anytime and anywhere. From the perspective of users, cloud computing data centers, are geographically distributed and the performance of network bandwidth varies by regions. Large public cloud service providers have dozens or even hundreds of data centers or service nodes to provide cloud computing

services to global customers.

# Contents

1. Cloud Computing Basics
2. **Cloud Computing Technologies**
  - Compute
  - Network
  - Storage

## Overview



- Compute services provide computing power required for running services such as websites, office software, and data analysis. Currently, typical compute cloud services are VMs and containers.
- Network services provide resource connectivity and isolation, such as data center networks and campus networks. On the cloud, VMs use virtual networks (for example, VPC) that have the logical topology similar to that of traditional networks.
- Storage services include:
  - Block storage: features high performance and low latency, meeting different high I/O service requirements.
  - File storage: allows file sharing among multiple servers or enterprise departments.
  - Object storage: features a flat, easy scale-out architecture, which is suitable for cloud storage. It is mainly used for massive data storage, cold data backup, and software repository.

## What Is Virtualization?

- The virtualization technology refers to the process of creating multiple VMs that share the hardware resources of a physical server.
  - A VM consists of **disk files** and **description files**, which are encapsulated in the same folder.
  - Multiple VMs running on the server are separately encapsulated in multiple folders and mutually isolated.
    - These folders can be stored in the **file system** provided by the underlying storage. Therefore, multiple VMs can be **stored or run on a shared medium**.

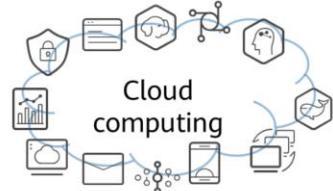


- The key concept behind virtualization involves separating software from hardware by converting "physical" devices into "logical" folders or files.

- In computer technologies, virtualization is a resource management technology. It abstracts various physical resources of a computer, such as CPU, memory, disk space, and network adapters, converts the resources, and presents the resources for segmentation and combination into one or more computer configuration environments. In this way, the uncut barriers between physical structures are broken, allowing users to use computer hardware resources in a better way than the original configuration.
- As shown in the figure, a physical server is divided into multiple files through virtualization, and each file represents a VM.

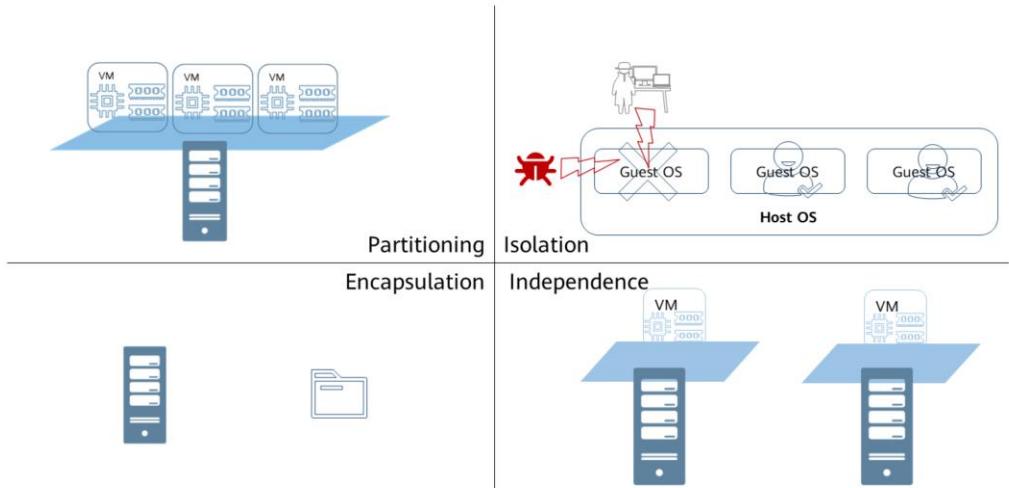
## Virtualization vs. Cloud Computing

- Virtualization is the fundamental technology that powers cloud computing. It transforms physical hardware into virtual resources. On the other hand, the cloud is an environment that delivers virtualized resources on-demand through the internet.



- Virtualization is a key technology of cloud computing. It aims to abstract physical resources into logical resources for flexible allocation. Virtualization offers scalable, distributed, and HA resources for cloud computing.
- Cloud computing allows users to use cloud resources on demand, relying on the virtualization technology.

## Main Features of Virtualization

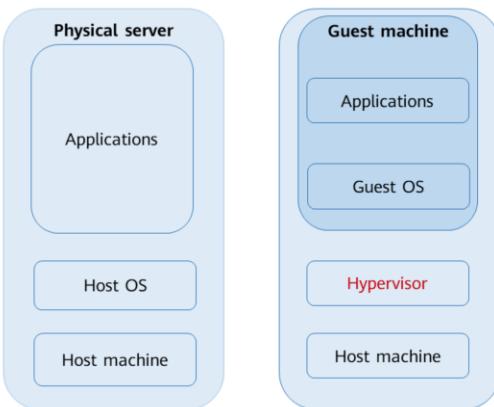


27      Huawei Confidential



- Features of VMs:
  - Partitioning: Multiple VMs can run on one physical server, which means that the virtualization layer can allocate resources of a physical server to multiple VMs. This is called partitioning.
  - Isolation: If one VM on a server is faulty or infected with viruses, the other VMs can still run properly.
  - Encapsulation: VMs exist in the virtualization system as files. You can migrate VMs by cutting/copying and pasting files.
  - Independence: After being migrated to another physical server, a VM can properly run without any modification on the server because VM OSs are decoupled from physical hardware.

# Important Virtualization Concepts

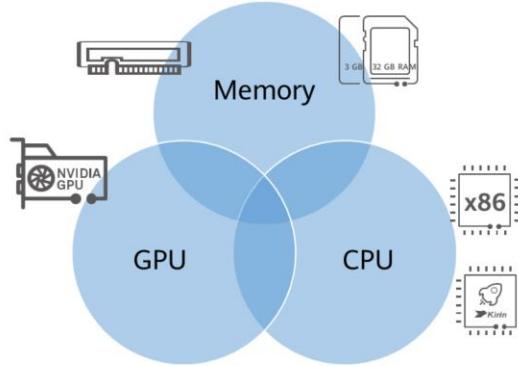


- **Guest OS:** Virtual machine (VM) OS
- **Guest Machine:** VM
- **Hypervisor:** Virtualization software layer/Virtual machine monitor (VMM)
- **Host OS:** OS running on a physical machine
- **Host machine:** physical machine

- Hypervisor: It is also called virtualization software or VM monitor. Hypervisor is used to create and run VMs on physical servers. The mainstream open-source virtualization technologies are Xen and KVM.

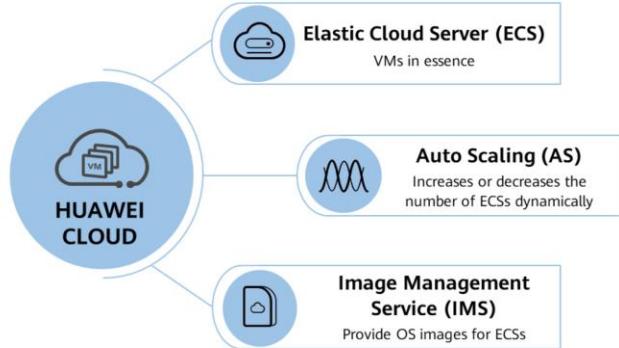
## Computing Resources Around Us

- Computing essentially refers to the process of obtaining information. In the ICT industry, several resources are needed to calculate data and obtain information.



- A computer system consists of CPU, memory, disk, and network resources. Compute resources include CPU, GPU, and memory.
- Central Processing Unit (CPU) is the computing and control core of a computer system, which processes information and executes programs.
- Memory is an important component of a computer system. It is used to store CPU computing data and exchange data between memory and external storage (such as hard disks).
- Graphics Processing Unit (GPU) is a microprocessor that performs image computation on PCs, workstations, game consoles, and mobile terminal devices such as tablet and smartphones.

# HUAWEI CLOUD Compute Services



- An Elastic Cloud Server (ECS) is a VM on the cloud, consisting of vCPUs, memory, OS, and EVS disks. After buying an ECS, you can use it on the cloud just like you would use your local PC or physical server.
- Auto Scaling (AS) automatically scales compute resources based on your demands and the AS policies you have configured, properly adjusting the number of ECSs as the service load changes over time.
- An image is a template used to create servers or disks. Image Management Service (IMS) provides image lifecycle management. With the IMS, you can create a system or data disk image from a server or an external image file, or create a full-ECS image from an ECS or a backup of an ECS.

## What Is a Container?

- A container is a lightweight, portable technology for application packaging. It is a standard unit that packages an application's code and all its dependencies, enabling the application to run across different computing environments. Simply put, containers are like standardized boxes that can hold different types of things and be put into different cabinets.



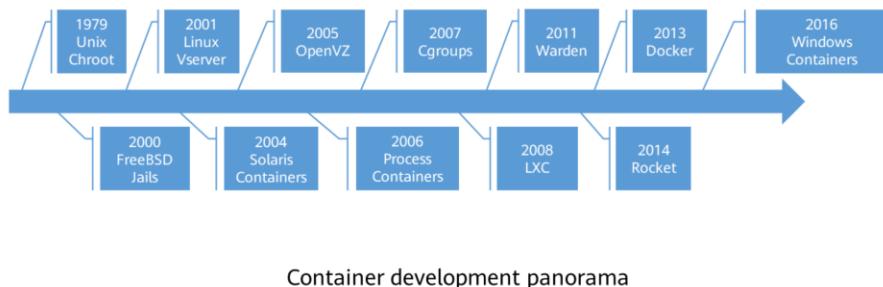
31      Huawei Confidential



- Containers can:
  - Package software into standardized units for development, migration, and deployment.
  - Isolate compute, storage, network, and other resources.
  - Start, stop, deploy, and migrate applications agilely and instantly.
  - Allow developers to focus on R&D and O&M engineers to focus on system maintenance.

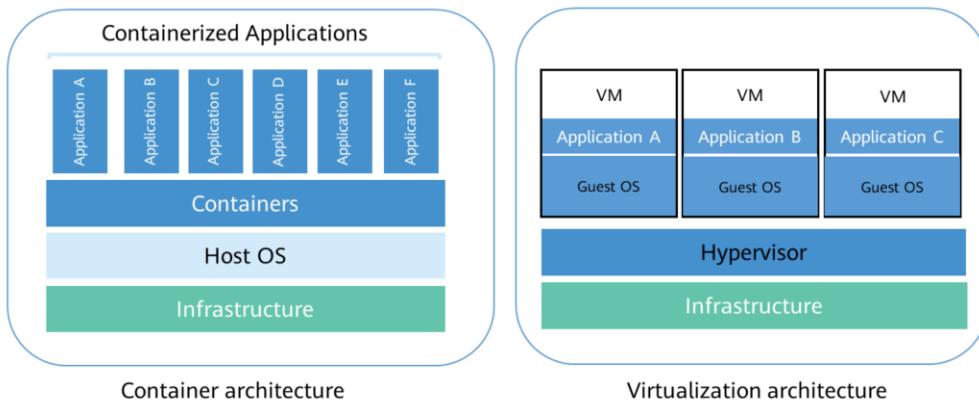
# Container Technology Development

- Two challenges in the development of container technology:
  - Unified platform
  - Usability



- Container technology was born in 1979, introduced as the chroot operation in UNIX. Chroot provided an isolated file system for each process so their root directories can be easily changed. This is the origin of OS virtualization.
- In 2000, BSD released FreeBSD Jails based on chroot. In addition to file system isolation, FreeBSD Jails isolate users, networks, and other resources. An IP address was assigned to each jail, which is an independent, smaller computer system, for independent software installation and configuration.
- In 2005, SWsoft released OpenVZ, which was similar to Solaris Containers. OpenVZ uses a modified Linux kernel to provide virtualization, isolation, resource management, and checkpoints. Since then, kernel virtualization has become a mainstream solution.
- In 2006, Google launched Process Containers. Process Containers, renamed as control groups (cgroups) later, were designed for limiting, accounting, and isolating resource usage (CPU, memory, disk I/O, network) of a collection of processes. In 2007, cgroups were merged into Linux kernel 2.6.24.
- In 2008, LXC (the first, most complete implementation of Linux container manager) was implemented using cgroups and Linux namespaces. LXC can work on a single vanilla Linux kernel without requiring any patches.

## Difference Between Containers and VMs (1)



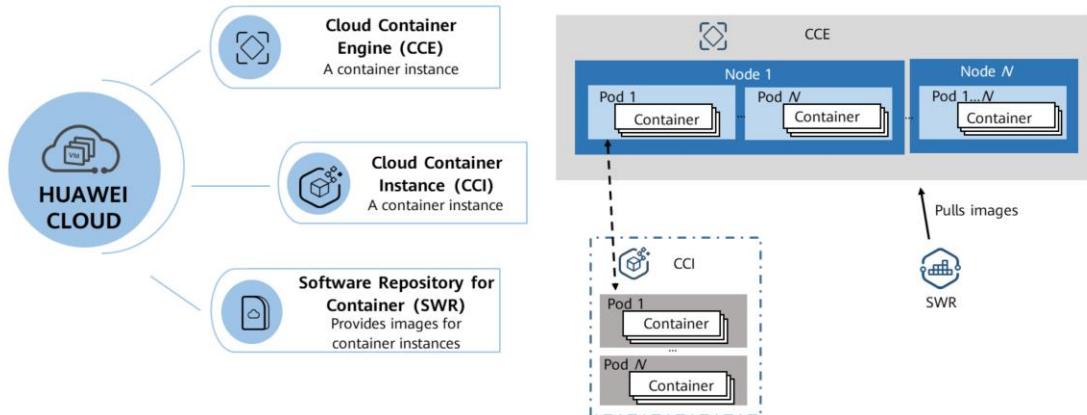
- Containers and VMs have similar advantages in resource isolation and allocation but different functions, because containers virtualize OSs instead of hardware. Containers are more portable and efficient.
- There is no virtualization layer in the container architecture. Therefore, containerization is called lightweight virtualization. Applications running in containers have better performance than those in VMs.
- Containers have become popular because of many benefits, including:
  - Agile building and deployment of applications: The creation of container images is easier and more efficient than that of VM images.
  - Consistent development, integration, and deployment: Containers can be quickly restored using images. You can customize new images for consistent container building and deployment.
  - Portability across clouds and OSs: Containers can run on Ubuntu, RHEL, CoreOS, Google Kubernetes Engine, physical servers, etc.
  - Application-centered management: The abstraction is improved from virtualizing hardware for OS isolation to virtualizing an OS for application isolation.
  - Loosely coupled, distributed, elastic, independent microservices: Applications are divided into independent, small units and can be deployed and managed separately instead of running on a single large server.
  - Isolated resources: Application performance can be predicted.
  - High resource utilization: Resources can be fully used.

## Difference Between Containers and VMs (2)

Item	Container	VM
Startup speed	Seconds	Minutes
Virtualization type	OS virtualization	Hardware virtualization
OS dependency	All containers share the host OS.	Each VM runs in its own guest OS.
Security	Process isolation with security risks	Complete isolation, which is more secure
Isolation strategy	Hypervisor	Cgroups
Image size	KB to MB	GB to TB
Virtualization performance	On par with physical servers	Limited
Per-machine capacity	Over 1,000 containers for each physical machine	Dozens of VMs

- Containers are an abstraction at the application layer. A container packages up code and its dependencies required for proper running of an application. Multiple containers can run on the same server with a shared OS kernel. Each container runs as an independent process in the user space. Containers take up less space than VMs, process more applications, and require less CPU and memory.
- Virtual Machines (VMs) are an abstraction of physical hardware and turn one server into multiple servers. The hypervisor allows multiple VMs to run on the same physical server. Each VM has its own OS, applications, necessary binaries, and libraries, taking up tens of GB. The startup speed of a VM may be slow.
- Container image: A container image is dedicated to running a specific service and usually contains only the resources required for running the service. Many widely used images are tens of MB or less in size.
- VM image: A VM image offers the operating environment (including the OS kernel) required by common processes and provides a complete collection of functions. The minimum size of a VM image is hundreds of MB.

## HUAWEI CLOUD Container Services



36      Huawei Confidential

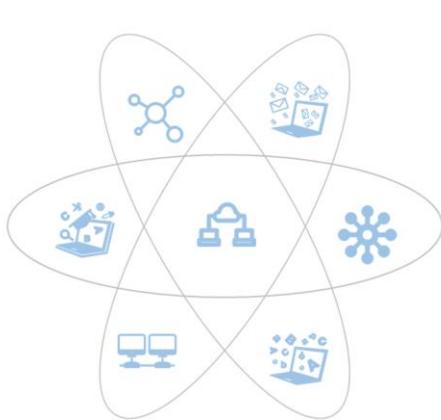


- Cloud Container Engine (CCE) is a high-performance, high-reliability service through which enterprises can manage containerized applications. CCE supports native Kubernetes applications and tools, allowing users to easily set up a container runtime environment on the cloud. CCE Turbo clusters run on the cloud native 2.0 infrastructure, accelerating compute, network, and scheduling.
- Cloud Container Instance (CCI) is a serverless container engine that allows users to run containers without creating or managing server clusters.
- SoftWare Repository for Container (SWR) allows users to easily manage the full lifecycle of container images and facilitates secure deployment of images for your applications. Users can upload, download, and manage container images through SWR Console, community CLI, or SWR APIs.
- SWR can either work with CCE and CCI or be used as an independent container image repository.

# Contents

1. Cloud Computing Basics
2. **Cloud Computing Technologies**
  - Compute
  - Network
  - Storage

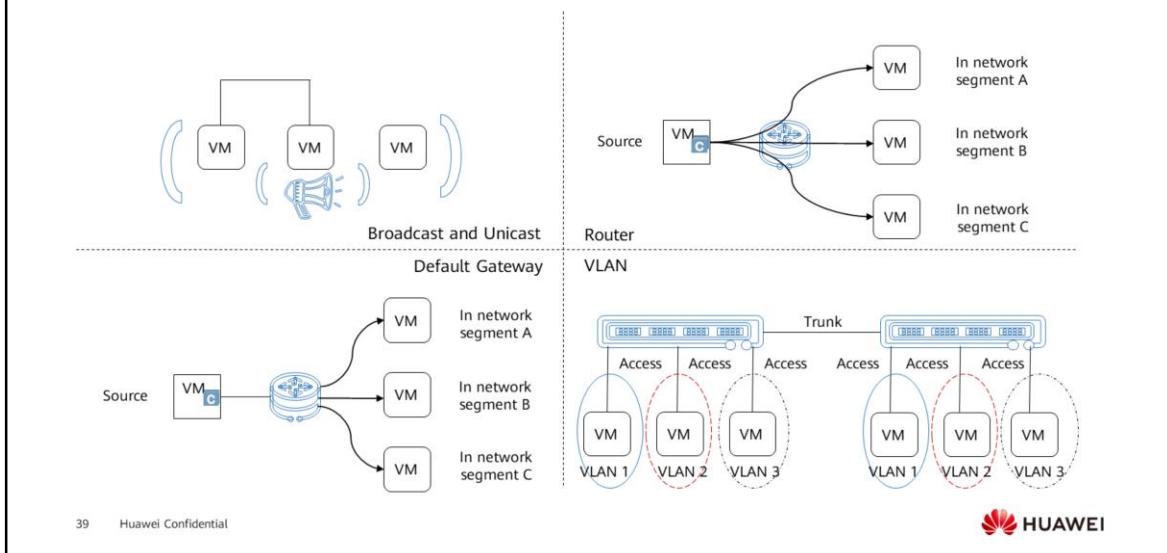
## What Does a Network Do?



 HUAWEI

- Networks bridge devices and VMs and allow them to communicate with each other. Therefore, networks are essential for ICT infrastructure.

## Basic Concepts of Conventional Networks



39      Huawei Confidential

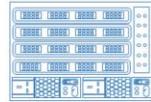
HUAWEI

- **Broadcast and unicast:** The communication between two devices is like that between people. The unicast, like one person talking to another, refers to that the information is sent and received between two nodes. The broadcast, like one person using a loudspeaker to talk to many people, has higher communication efficiency and ensures that the information can be sent to all related devices.
- **Router:** A router is a hardware device that connects two or more networks. It works as a gateway to read the address of each data packet and decide how to forward it.
- **Default gateway:** To understand the default gateway, we need to know what a gateway is. A gateway is a device that connects a subnet to an external network. When a device sends information to a host, a subnet mask determines whether the destination host is on the local subnet according to the destination address. If the host is on the local subnet, the device can directly send information to the host. If not, the device will first send the information to the default gateway or router, which then forwards the information to other networks to reach the host.
- **Virtual Local Area Network (VLAN):** VLAN is a group of logical devices and users, which are organized based on functions, departments, and applications, regardless of their physical locations. Such devices and users communicate with each other as if they are on the same network segment. VLANs can be used to isolate different services.

## Conventional Network Devices



Router



Layer 3 switch

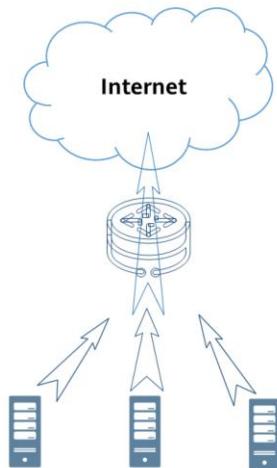


Layer 2 switch



Network interface card (NIC)

## What Does a Router Do?

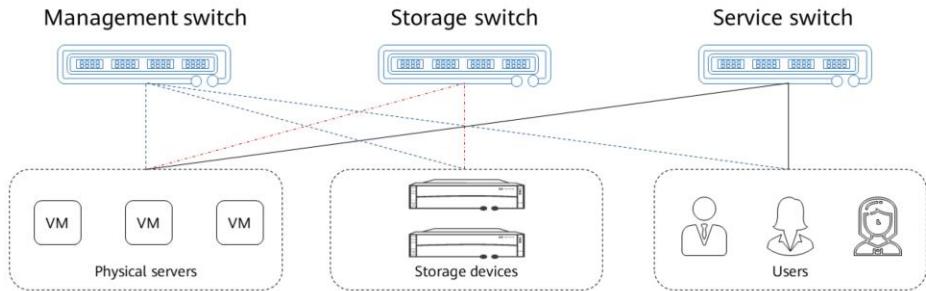


41      Huawei Confidential

 HUAWEI

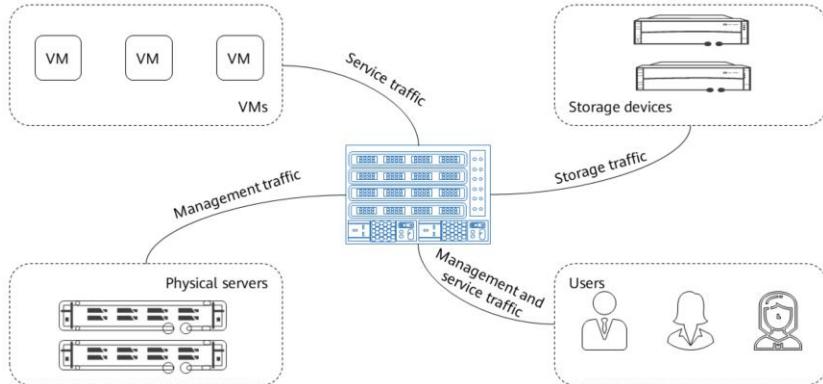
- Our PCs can access the Internet through a router. Likewise, servers can be connected to the Internet by using a router.
- A router is a gateway device that operates on the third layer of the OSI Model, the network layer. It stores and forwards packets between different networks, and routes data from one subnet to another. In network communications, routers can determine network addresses and select IP routes. Routers can flexibly set up connections for networks and send packets between them through different media access control mechanisms. Routers accept information only from the source and other related routers, functioning as interconnection devices on the network layer.

## What Does a Layer 2 Switch Do?



- A network switch is used to forward electrical signals, and establishes an exclusive electrical signal route for any two nodes connected to the switch. Ethernet switches are most commonly used. Other common switches include telephone voice switches and fiber switches. Switching allows devices to automatically or allows you to manually send information to an appropriate route, meeting the requirements of both communications ends. A switch has multiple ports, with each port providing the bridging function. A port can be connected to a local area network (LAN) or a high-performance server or workstation.
- On a conventional network, Layer 2 switches use VLANs to isolate network planes.

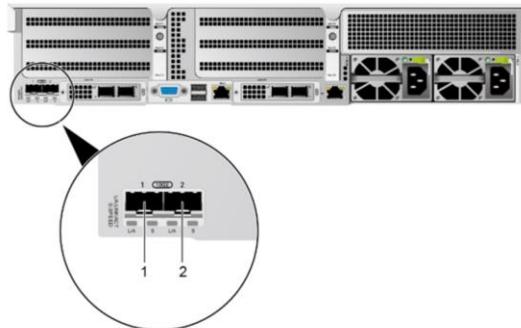
## What Does a Layer 3 Switch Do?



- For safety and management purposes, a large local area network (LAN) must be divided into several small LANs to reduce the impact of broadcast storms, so the virtual local area network (VLAN) technology is widely used. Communications between different VLANs are forwarded by routers. With the increase of access across networks, if only routers are used, the network scale and access speed are restricted because there are limited port quantity and the routing speed is slow. To address this, Layer 3 switches are developed. Layer 3 switches are designed for IP addresses. These switches provide simple APIs and are strong in processing Layer 2 packets, suitable for routing and switching data in large LANs. Layer 3 switches not only replace or partially complete the function of traditional routers in the third layer of the network model, but also have almost the same switching speed as the second layer. And the price of Layer 3 switches is cheaper.

## What Does a NIC Do?

- NICs are mainly used to connect different devices. Like a telephone card, they ensure devices can communicate. In addition, NICs can be bound to deliver higher reliability and better network performance.

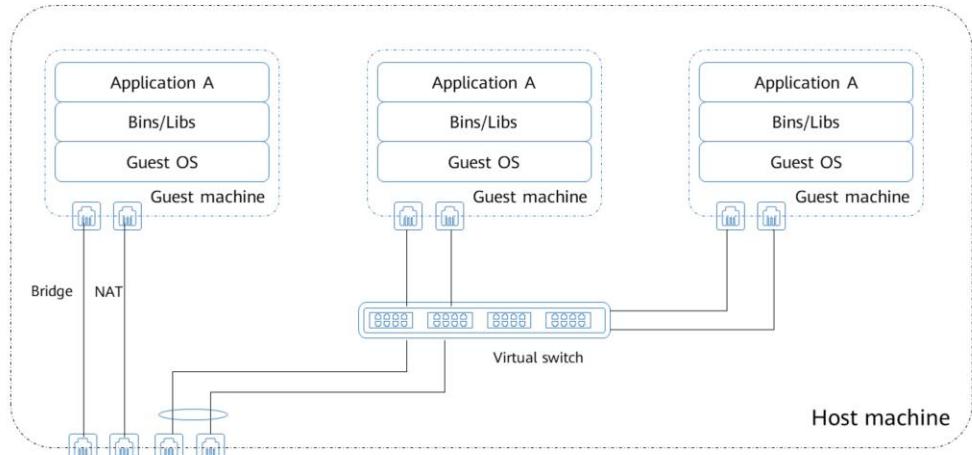


44      Huawei Confidential

 HUAWEI

- The onboard NIC provides network expansion capabilities. It transmits data from servers to other devices, providing application services externally.
- Commonly supported NIC speed rates include 100 Mbit/s, 1 Gbit/s, and 10 Gbit/s.

## Basic Concepts of Virtual Networks

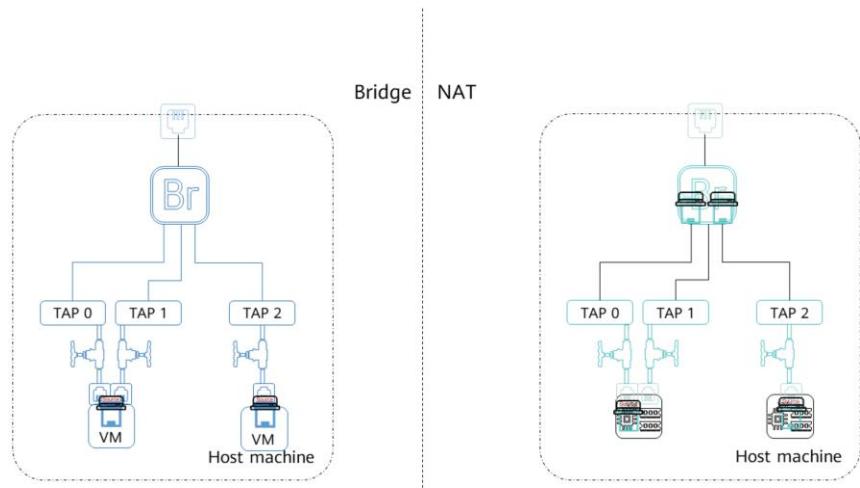


45 Huawei Confidential

 HUAWEI

- Why is a virtual network required?
  - VMs hosted on a physical machine may be in different IP address ranges, so these IP address ranges need to be isolated. In addition, VMs need to share the same physical NIC to access external networks. Therefore, virtual switches are used on servers to construct virtual networks.
- In network virtualization, the first problem to be solved is how to map virtual NICs of the VMs to the physical NICs of the physical server where the VMs are hosted. As shown in the figure, we can use network bridges, NAT and virtual switches to solve this problem.

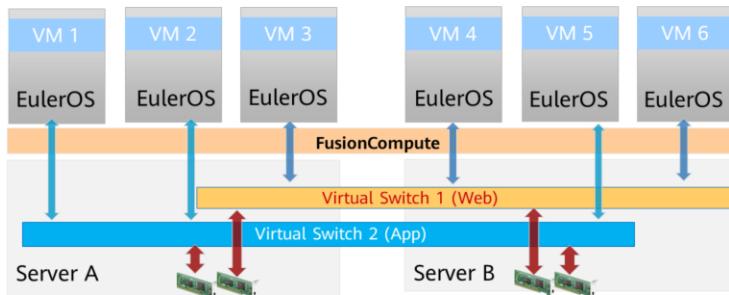
## What Do Bridge and NAT do?



- Both a bridge and NAT can forward the traffic of different VMs to physical NICs so that data packets can be routed from the server to the physical switch, implementing the communication between VMs and between VMs and external networks.
- Virtual switches also have the bridging function. A virtual switch has a table that defines mapping between MAC addresses and ports to isolate collision domains. Simply speaking, a bridge connects different physical LANs at the data link layer.
- NAT forwards the traffic to external networks through translating network addresses. NAT not only avoids the lack of IP addresses, but also protects computers on the private network from being attacked by other networks.

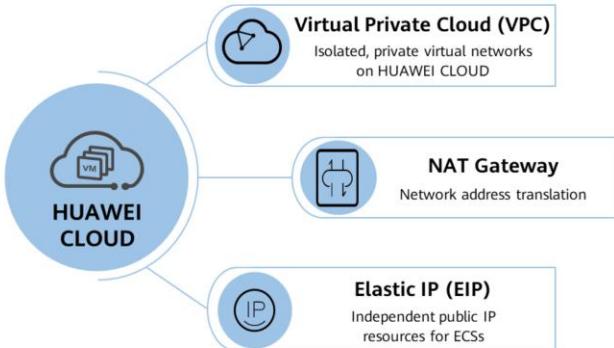
## What Does a Virtual Switch Do?

- Like the bridge and NAT, virtual switches are used to transmit the internal traffic of VMs to the external network through the network port of the physical server where the VMs reside. The common virtual switch models include OVS and EVS.



- Open vSwitch (OVS): An OVS is a software-based open source virtual switch. It supports multiple standards and protocols with additional support for the OpenFlow protocol, and can be integrated with multiple open-source virtualization platforms. An OVS can be used to transmit traffic between VMs and implement communication between VMs and external networks.
- Enhance vSwitch (EVS): An EVS is an enhanced OpenFlow-compliant virtual switch that improves the I/O performance based on the OVS forwarding technology. I/O performance is significantly improved by using the Intel DPDK technology and using user-mode processes rather than NICs to send and receive data.
- On an OVS, data is received and sent in the kernel mode, but on an EVS, data is processed in the user mode.
- Distributed Virtual Switch (DVS): Same as a physical switch does, a DVS constructs the network between VMs and connects VMs to external networks.
- A virtual NIC of a VM communicates with an external network by connecting to the DVS, then by connecting to the physical NIC of the host through the DVS uplink.
- Compared with traditional switches, using virtual switches reduces network devices and simplifies the network architecture, relieving the pressure of system management and maintenance.

## HUAWEI CLOUD Network Services

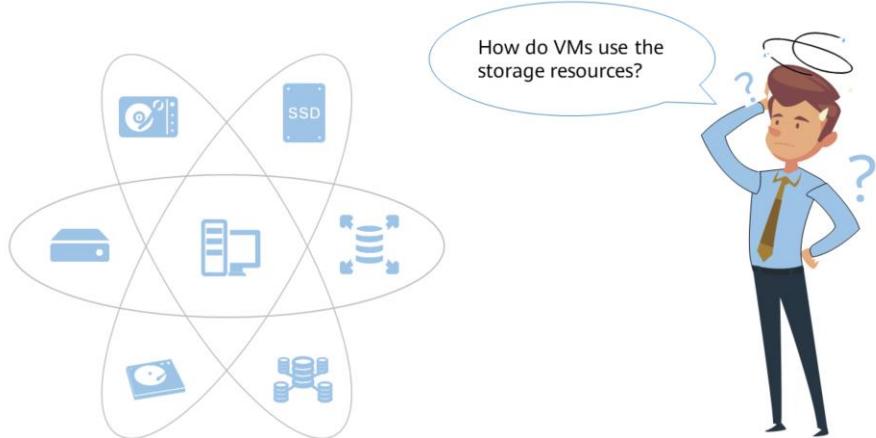


- A Virtual Private Cloud (VPC) is a private and isolated virtual network on HUAWEI CLOUD. Users can configure IP address ranges, subnets, and security groups, assign EIPs, and allocate bandwidths in a VPC.
- Public NAT gateways and private NAT gateways are used in different scenarios to provide the network address translation. A public NAT gateway provides SNAT and DNAT so that cloud servers in a VPC can share EIPs to access the Internet. A private NAT gateway provides the network address translation for servers in a VPC.
- The EIP service 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. This service provides various billing modes to meet diverse service requirements, and allows cloud servers in a VPC to share the same private IP address to access or provide services accessible from an on-premises data center or a remote VPC.

# Contents

1. Cloud Computing Basics
2. **Cloud Computing Technologies**
  - Compute
  - Network
  - Storage

## How Does Cloud Storage Work?



50      Huawei Confidential

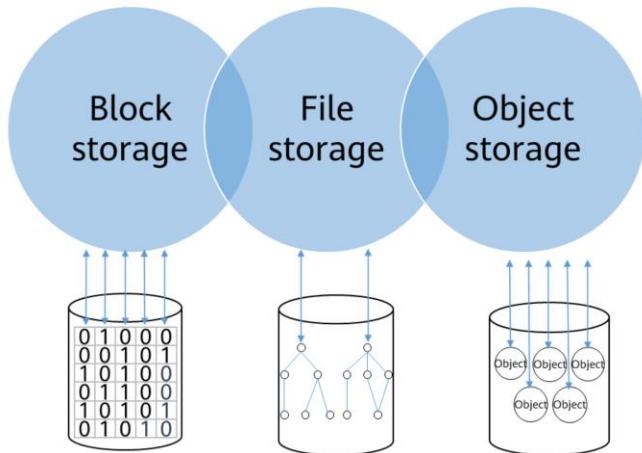
 HUAWEI

- How does storage media work?

A storage medium is any technology -- including devices and materials -- used to place, keep and retrieve electronic data.

- In terms of data storage, the existing cloud storage products can achieve higher efficiency at lower cost. Therefore, cloud storage will be an inevitable choice for individuals and enterprises.

## Mainstream Storage Types



51      Huawei Confidential



- Traditional servers have computing and storage coupled, and use their local physical disks to store data. This is what we call the traditional block storage, where a disk is connected to a server through a bus, delivering low latency. However, the number of disks attached to the server is limited, so traditional servers have poor performance in capacity, bandwidth, and reliability. The explosive data growth poses high requirements on data reliability, which requires decoupled compute and storage. To address this, storage arrays appear. Traditional disk arrays comprise controllers and disk enclosures. Two or more controllers can be used to provide high reliability. By adding disk enclosures, the capacity of disk arrays can be hundreds of thousands of times larger than that of local disks. Disk arrays independently connect to servers through FC switches or IP switches. This is today's block storage.
- As the IT system further develops, enterprises want their files to be shared among multiple hosts for concurrent access. This is shared file storage. Shared file storage shares data in the same data center or equipment room.
- As more and more Internet applications need to access data over the Internet using terminal devices, object storage that supports HTTP and HTTPS protocols is widely used. Object storage allows applications to access data by calling APIs and adopts a distributed architecture featuring large capacity and high reliability.

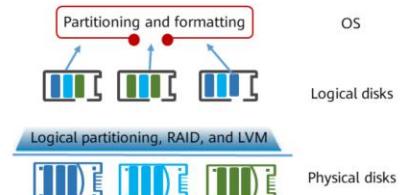
# Block Storage

## Definition

- Block storage maps the entire raw disk space to a server. For example, five disks in a disk array can be divided into several logical disks, which are then mapped to the server. After partitioning, formatting, and mounting on the server, the data is successfully stored.

## Application Scenario

- Block storage is ideal for most data storage scenarios.



- Block storage cannot be directly used in an operating system. Before using a block device, you must format it and create a file system on it. Data in the operating system is stored as files.
- Block storage has the lowest latency among the three types of storage and is ideal for mission-critical applications such as databases and ERP systems.

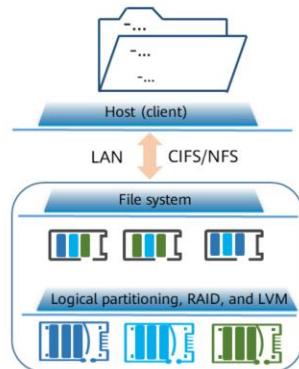
# File Storage



- File storage is like a shared folder in a file system. Users can directly store data on the file storage device over file system access protocols, such as NFS and CIFS.



- File storage is widely used in scenarios such as data backup and archive, image and video data storage, and file sharing.



- Network File System (NFS): NFS is a file sharing protocol between UNIX operating systems. It commonly applies to Linux clients.
- Common Internet File System (CIFS): CIFS is a protocol that allows programs to access files on remote computers over Internet. It mainly applies to Windows clients.
- File storage provides PB-level capacity and ms-level latency and is perfect for scenarios where data needs to be shared among multiple compute nodes , such as HPC and office automation.

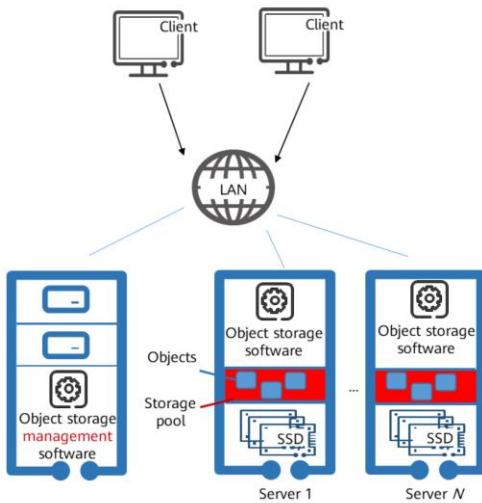
# Object Storage



- Object storage leverages both block storage and file storage. It offers fast, direct disk access, and distributed file sharing. Generally, object storage outperforms file storage.



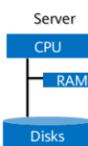
- Object storage is widely used in scenarios such as data backup, image and video data storage, and website hosting.



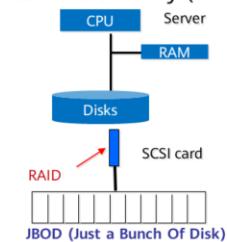
- Object storage has large-scale data management capability, which is its biggest advantage over file storage. File Storage uses a hierarchical structure to manage all files and directories. If there are too many files or directories stored, the search performance will be greatly reduced. Object storage provides a flat structure where all objects are stored at the same logical layer. This keeps the object search speed almost unchanged even if there are tens of billions of objects. However, object storage uses application-level APIs instead of system-level APIs. Traditional applications need to be redeveloped when being migrated to object storage systems, which makes the popularization of object storage difficult.
- Object storage is applicable to scenarios such as big data, IoT, backup and archive. It provides EB-level capacity and has the highest data durability among the three types of storage.

# Enterprise Storage

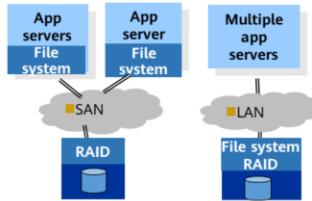
Disks in a server



External disk array (DAS)



Storage area network (SAN/NAS)



**Known issues:**

- Disks have become a system performance bottleneck.
- The number of disk slots is limited, resulting in small capacity.
- Data is stored on a single disk, lowering data reliability.
- Storage utilization is low.
- Data is scattered in local storage systems.

JBOD combines multiple physical disks into a logical unit to increase capacity, without providing any data protection.

**Resolved issues:**

- The number of disk slots is limited, resulting in small capacity.
- Data is stored on a single disk, lowering data reliability.

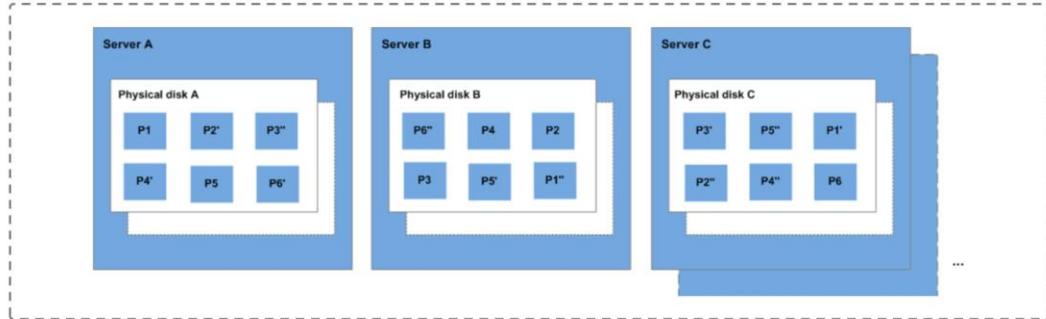
**Resolved issues:**

- Disks have become a system performance bottleneck.
- The number of disk slots is limited, resulting in small capacity.
- Data is stored on a single disk, lowering data reliability.
- Storage utilization is low.
- Data is scattered in local storage systems.

- Direct Attached Storage (DAS) connects an external storage device to an application server through the SCSI or FC interface, making the storage device part of the server. In this case, the data and operating system are not separated.
- Network Attached Storage (NAS) uses TCP/IP, ATM, and FDDI to connect storage devices, switches, and clients, and all these components form a private storage network. NAS integrates storage devices, network interfaces and Ethernet technology and stores data directly over Ethernet, which separates the storage function from the file server.
- Storage Area Network (SAN) is a private storage network that connects storage arrays and servers through switches.

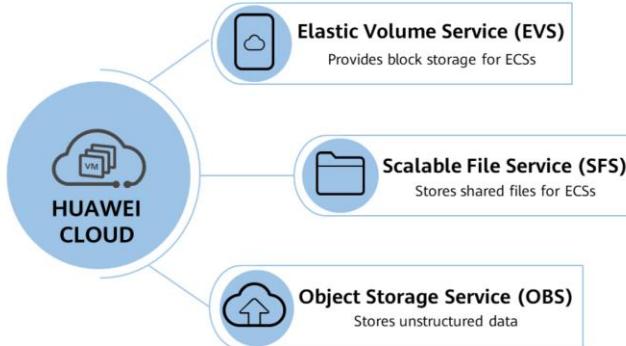
## Distributed Storage

- Distributed storage systems virtualize the available storage resources across all hosts of an enterprise to a virtual storage device. This way, data is stored in different locations on the storage network, improving system reliability, availability, and access efficiency.



- As data grows exponentially, storage of massive amount of data imposes great pressure on local storage and brings heavy burden to existing storage systems. To relieve the pressure, we have to adopt distributed storage and distributed file systems.
- How can we ensure high performance and high availability of distributed storage?
  - In addition to the backup, active-active, and multi-active architectures in the traditional architecture, multiple data copies are stored in the system to ensure high reliability and availability of the distributed storage system. If a storage node becomes faulty, the system can automatically switch the node's service to other nodes, achieving automatic fault tolerance. The distributed storage system leverages replication protocols to synchronize data to multiple storage nodes and ensures data consistency between copies. A piece of data has multiple copies, among which there is only one primary copy, and the rest are backup copies. Consistency is used to ensure data integrity when data is replicated from the primary copy to backup copies.

## HUAWEI CLOUD Storage Services



- Elastic Volume Service (EVS) provides persistent block storage for ECSs and BMSs. With data redundancy and cache acceleration techniques, EVS offers high availability, strong durability, and low latency. Users can format an EVS disk, create a file system on it, and store data persistently.
- SFS is a network attached storage (NAS) service that provides scalable, high-performance file storage. With SFS, you can enjoy shared file access spanning ECSs, BMSs, and containers created on CCE and Cloud Container Instance (CCI).
- Object Storage Service (OBS) provides a stable, secure cloud storage that is scalable, efficient, and easy-to-use. It offers REST APIs and allows users to store any amount of unstructured data in any format.

# Quiz

1. (Single-answer question) Do the bridge and NAT have the same working principles?
  - A. Yes
  - B. No
2. (Single-answer question) Which of the following is NOT a mainstream storage type?
  - A. Block storage
  - B. Object storage
  - C. Tape library
  - D. File storage

- B. The bridge only connects the network port of a VM to a physical network port to access external networks through a physical server. The NAT translates network addresses to enable VM traffic to access external networks through a physical network port.
- C. Currently, there are three mainstream storage types: block storage, file storage, and object storage. Tapes are also a type of storage media, but they are only used in some backup and archive scenarios.

# Summary

This chapter described:

- Cloud computing basics
- Compute, network, and storage technologies
- HUAWEI CLOUD services

# Recommendations

- Huawei iLearning
  - <https://e.huawei.com/en/talent/#/search?productTags=&productName=&navType=learningNavKey>
- Huawei Technical Support Website
  - <https://support.huaweicloud.com/intl/en-us/help-novicedocument.html>
- HUAWEI CLOUD Academy
  - <https://e.huawei.com/en/talent/#/ict-academy/home>

## Acronyms and Abbreviations

APP: Application

AS: Auto Scaling

CPU: Central Processing Unit

CCE: Cloud Container Engine

CCI: Cloud Container Instance

CIFS: Common Internet File System

ECS: Elastic Cloud Server

EIP: Elastic IP

EVS: Elastic Volume Service

## Acronyms and Abbreviations

GPU: Graphics Processing Unit

ICT: Information and Communications Technology

I/O: Input/Output

IaaS: Infrastructure as a Service

IBM: International Business Machines Corporation

KVM: Kernel-based Virtual Machine

IMS: Image Management Service

LXC: Linux Container

LVM: Logical Volume Manager

## Acronyms and Abbreviations

NAT: Network Address Translation

NFS: Network File System

NIST: National Institute of Standards and Technology

OS: Operation System

OBS: Object Storage Service

PC: Personal Computer

PaaS: Platform as a Service

RAID: Redundant Arrays of Independent Disks

SFS: Scalable File Service

## Acronyms and Abbreviations

SWR: SoftWare Repository for Container

SaaS: Software as a Service

TCO: Total Cost of Ownership

TAP: Test Access Point

VM: Virtual Machine

VLAN: Virtual Local Area Network

VPC: Virtual Private Cloud

# Thank you.

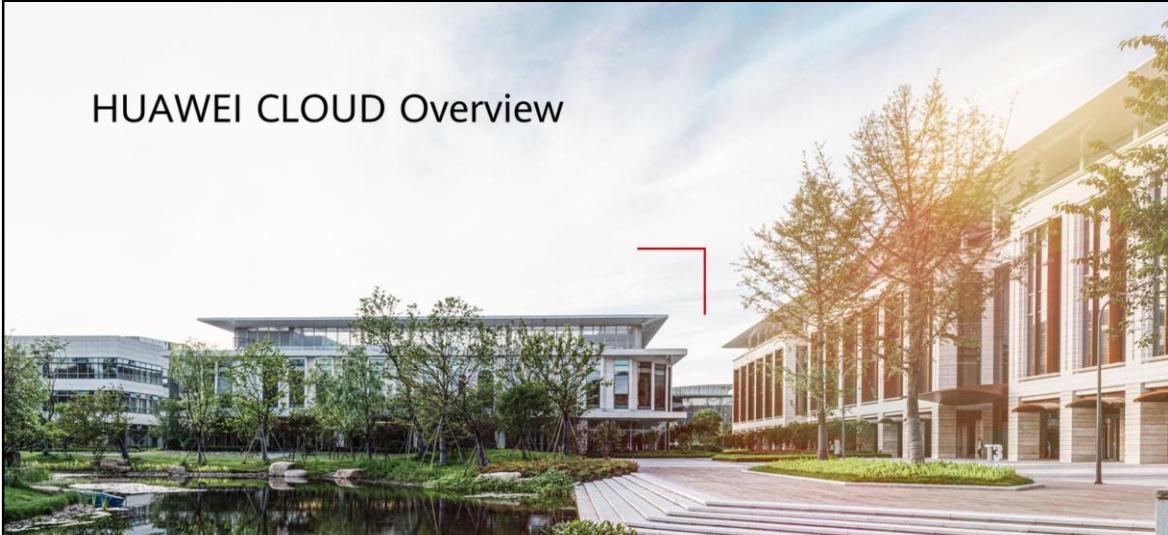
把数字世界带入每个人、每个家庭、  
每个组织，构建万物互联的智能世界。  
Bring digital to every person, home, and  
organization for a fully connected,  
intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.  
All Rights Reserved.

The information in this document may contain predictive  
statements including, without limitation, statements regarding  
the future financial and operating results, future product  
portfolio, new technology, etc. There are a number of factors that  
could cause actual results and developments to differ materially  
from those expressed or implied in the predictive statements.  
Therefore, such information is provided for reference purpose  
only and constitutes neither an offer nor an acceptance. Huawei  
may change the information at any time without notice.



# HUAWEI CLOUD Overview



 HUAWEI

## Foreword

- HUAWEI CLOUD is a leading cloud service provider, which brings Huawei's 30-plus years of expertise together in ICT products and solutions. We are committed to providing reliable, secure, and cost-effective cloud services to empower applications, harness the power of data, and help organizations of all sizes grow in today's intelligent world. HUAWEI CLOUD is also committed to bringing affordable, effective, and reliable AI services.
- In this chapter, we will have a brief overview of HUAWEI CLOUD.

# Objectives

- Upon completion of this course, you will:
  - Understand the positioning and application scenarios of HUAWEI CLOUD.
  - Understand the delivery modes, technical support, and ecosystem of HUAWEI CLOUD.
  - Understand concepts related to HUAWEI CLOUD.

# Contents

- 1. HUAWEI CLOUD Overview**
2. Application Scenarios
3. Delivery Modes
4. Technical Support
5. Ecosystem
6. Quick Start

## Start a Journey with HUAWEI CLOUD

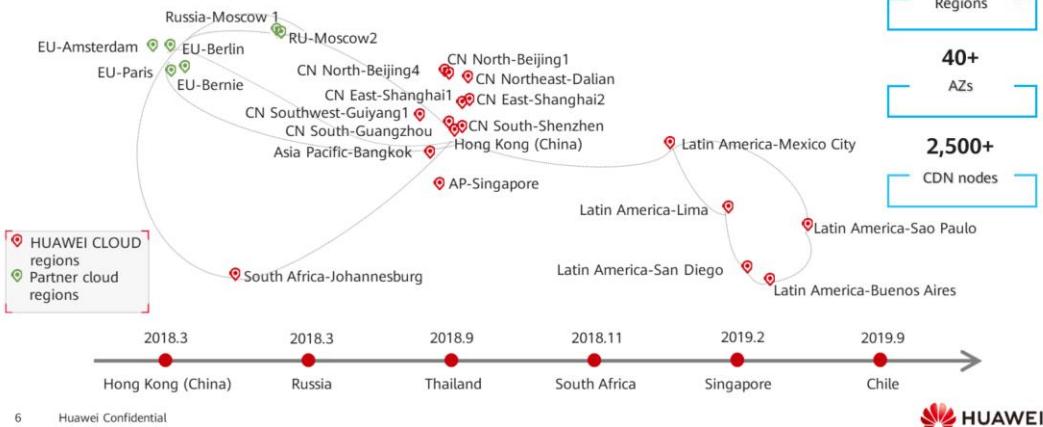


HUAWEI CLOUD official website: <https://www.huaweicloud.com/intl/en-us>

- HUAWEI CLOUD is a public cloud service brand that leverages Huawei's more than 30 years of expertise in the ICT field to provide innovative, secure, and cost-effective cloud services.
- Video: <https://www.huaweicloud.com/content/dam/cloudbu-site/archive/hk/en-us/about/index/images/huaweicloudvideo.mp4>.

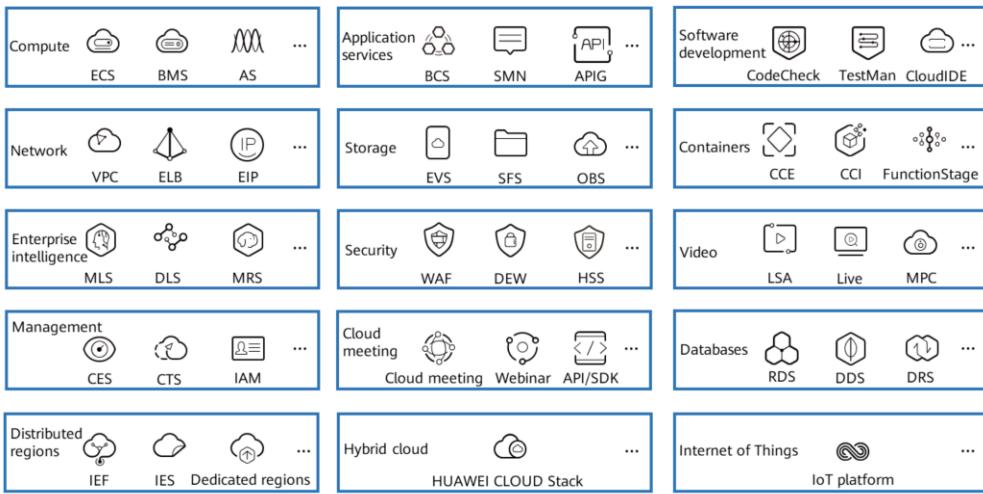
## Global Connectivity and Services

- HUAWEI CLOUD provides global connectivity. With HUAWEI CLOUD, Chinese enterprises can quickly expand into the international markets and international enterprises can also tap into fast-growing Chinese markets.



- HUAWEI CLOUD has been deployed in Singapore, Chile, Brazil, Mexico, and Peru. Together with our partner clouds, HUAWEI CLOUD is operating more than 40 AZs in over 20 geographic regions around the world. We can provide global public cloud services for multinational enterprises. We can help Chinese enterprises expand into the international markets and help international enterprises tap into the fast-growing Chinese markets. In Asia Pacific, we have local technical support teams in over 10 Asia Pacific countries. In Latin America, we have the most local data centers. HUAWEI CLOUD was officially launched in South Africa in 2019, becoming the first cloud service provider to operate local data centers there. Currently, we are providing cloud services for 12 African countries: Angola, Botswana, Ghana, Kenya, Mauritius, Mozambique, Namibia, Nigeria, South Africa, Tanzania, Zambia, and Zimbabwe. We have developed more than 200 partners in Africa and we are working together to serve customers in more than 30 African countries.

## 200+ Cloud Services Are Available and More Are Coming



7 Huawei Confidential



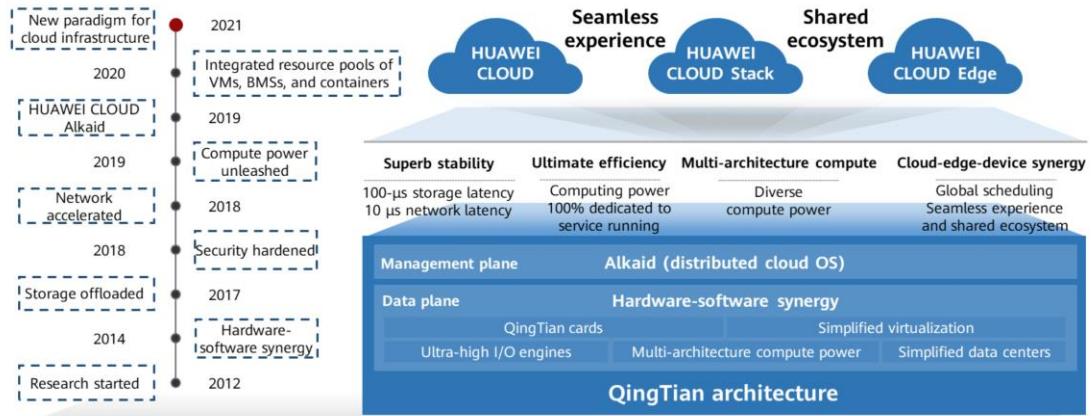
- HUAWEI CLOUD is continuously upgrading full-stack cloud native technologies. So far, we have launched more than 220 cloud services and more than 210 solutions.

# HUAWEI CLOUD Data Centers: Innovation from Chips

AI processors	Smart NICs	Faster and smarter SSDs	Chip-based root of trust
<b>Ascend</b> Ascend AI processor	<b>Hi1822</b> Industry-First 100 G iNIC	<b>Hi1812E</b> 4 <sup>th</sup> Gen SSD Controller	<b>DAEMON</b> Chip-Based Root of Trust
<ul style="list-style-type: none"><li>• 16 – 512 TOPS series</li><li>• Innovative DaVinci architecture</li><li>• Optimized AI instruction sets</li></ul>	<ul style="list-style-type: none"><li>• Programmable NICs, better performance than standard NICs</li><li>• Multi-protocol offloading, including VxLAN/RoCE/OVS</li><li>• 15 MPPS, 2.5 times higher than the industry's best</li></ul>	<ul style="list-style-type: none"><li>• IOPS: ↑ 75%+</li><li>• Bandwidth: ↑ 60%+</li><li>• Latency: ↓ about 15% (thanks to the intelligent multi-stream technology)</li></ul>	<ul style="list-style-type: none"><li>• Firmware security protection</li><li>• Strong ID security protection</li><li>• Trustworthiness management</li></ul>

- Chips are the core of and the most difficult part of R&D in the IT industry, which requires long-term investment.
- Huawei has over 20 years of experience in chip R&D and is constantly innovating chips for Cloud 2.0. We have launched a full series of chips for next-generation cloud data centers.
  - Compute chips: a full series of AI processors
  - Network chips: Huawei's next-generation network chips Hi822 use the NP-like programmable architecture and support offloading of multiple protocols.
  - Storage chips: The fourth generation of storage chips improved performance by over 75% and bandwidth by over 60%. Thanks to intelligent multi-stream technology, the latency was decreased by about 15%.
  - Security chips: Huawei has built security and trustworthiness into chips. We provide comprehensive protection for firmware, identities, software systems, and data management.

# Innovative QingTian Architecture

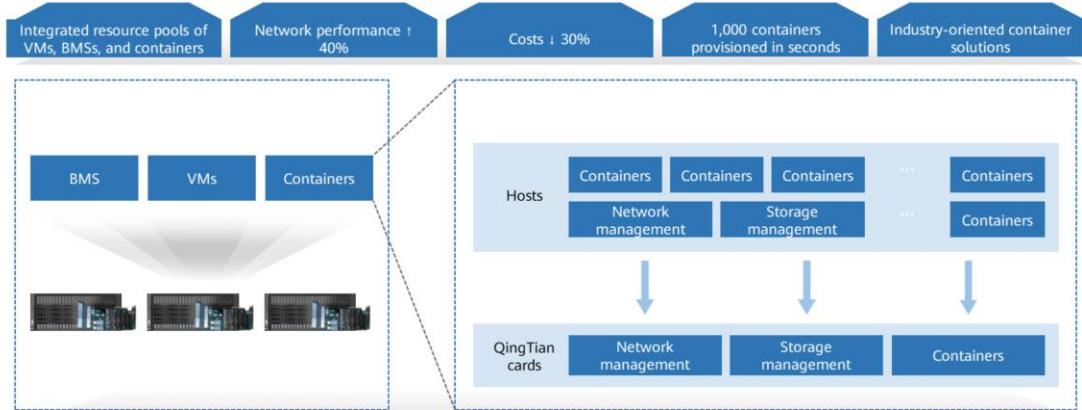


9      Huawei Confidential



- The QingTian architecture enables the Huawei cloud infrastructure to provide cloud services featuring unparalleled performance, ultimate stability, multi-architecture compute, and cloud-edge-device synergy. The QingTian architecture provides a consistent architecture, experience, and ecosystem for HUAWEI CLOUD, HUAWEI CLOUD Stack, and HUAWEI CLOUD edge.

## The Industry's Only BMS with No Loss of Performance

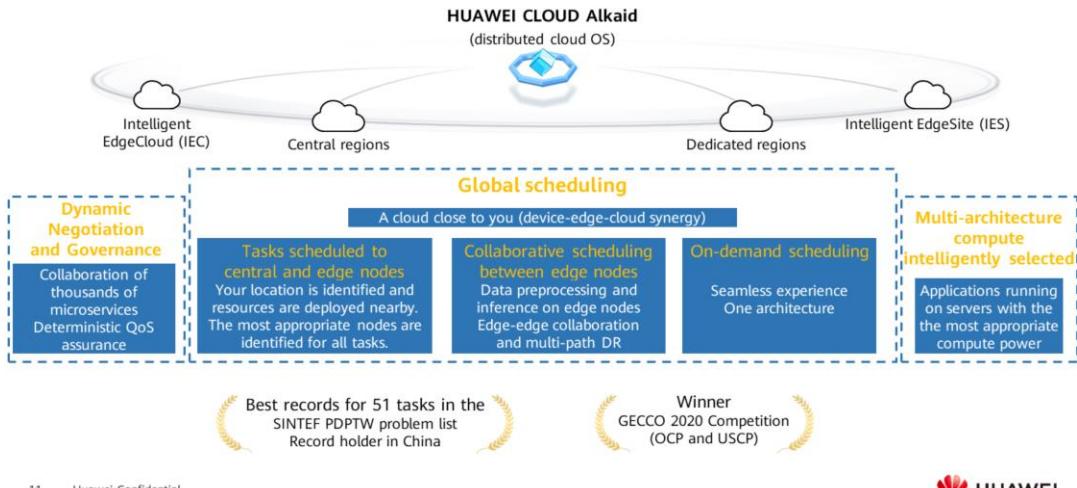


10     Huawei Confidential



- The QingTian architecture integrates the resource pools of BMSs, VMs, and containers, allowing resources to be 100% dedicated to processing services. This architecture powers BMS 2.0, the industry's only BMS with no loss of performance. The container engines have been offloaded to QingTian cards, improving the network performance by 40% and slashing costs by 30%.

# Alkaid Distributed Cloud OS Schedules Tasks Globally



11      Huawei Confidential



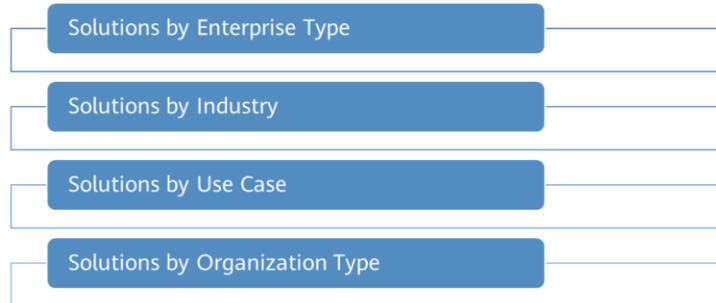
- The QingTian architecture offers hardware-software synergy for the data plane, and Alkaid is a distributed cloud OS for the management plane. QingTian is the infrastructure base of HUAWEI CLOUD, and Alkaid is a distributed cloud OS running on the QingTian architecture.
- The Alkaid distributed cloud OS provides three key capabilities to deliver an amazing simplified experience: dynamic negotiation and governance, global scheduling, and selecting the most appropriate compute power for your applications.
  - Global scheduling: Compute tasks are scheduled across over 100,000 distributed sites to deliver a superlative experience to users.
    - Scheduling across central and edge nodes: You access services from the node nearby that can best meet your requirements.
    - Edge-edge collaboration: You can complete data preprocessing and inference on edge nodes, for example, for vehicle-road collaboration scenarios.
    - On-demand scheduling: You run compute tasks on nodes with the most appropriate cloud services and operators (180+ cloud services and 40+ operators).

# Contents

1. HUAWEI CLOUD Overview
- 2. Application Scenarios**
3. Delivery Modes
4. Technical Support
5. Ecosystem
6. Quick Start

## Application Scenarios

- HUAWEI CLOUD provides cloud services and solutions to address your requirements, no matter which industry you are in.



- These four solutions will be described in detail later.

## Solutions by Enterprise Type

- HUAWEI CLOUD provides a full portfolio of cloud solutions to empower startups, marketing, management, and business expansion.

Startups	<ul style="list-style-type: none"><li>• Establishment, OA, marketing, intellectual property rights protection, and business expansion</li></ul>
Website building	<ul style="list-style-type: none"><li>• 3,000+ website templates, one-on-one expert advisory, free ICP filing</li></ul>
Enterprise management	<ul style="list-style-type: none"><li>• ERP on cloud</li></ul>
International expansion	<ul style="list-style-type: none"><li>• Solutions for helping Chinese-funded enterprises go international</li></ul>
Cloud migration	<ul style="list-style-type: none"><li>• Web applications on cloud</li></ul>
Transportation management	<ul style="list-style-type: none"><li>• Network freight platform on cloud</li></ul>

14      Huawei Confidential



- HUAWEI CLOUD provides a wide range of solutions to empower cloud migration, startup establishment, enterprise management, business expansion, and transportation management:
  - Cloud migration solutions for migrating Web applications to the cloud, cloud network interconnection, disaster recovery, and containers on cloud solutions.
  - Solutions oriented for startups
  - Enterprise management solutions, such as ERP on cloud and HUAWEI CLOUD Meeting
  - Solutions for helping cross-border e-commerce, gaming, and SaaS enterprises expand into international markets
  - Solutions for migrating network freight platforms to the cloud

## Solutions by Industry

- HUAWEI CLOUD provides solutions for a wide range of industries, so you can always find the cloud services you need.



- HUAWEI CLOUD provides the following industry-specific solutions (for details, visit the HUAWEI CLOUD official website):
  - Financial services: virtual banking, insurance services on cloud, securities quotes
  - Retail: e-commerce website building, smart retail, smart store, apparel & footwear websites
  - Media & entertainment: rendering, VR video, convergent media
  - Transportation and logistics: smart highway, smart airport, smart logistics, smart parking
  - Agriculture and environmental protection: smart meteorology, remote sensing
  - Gaming: cloud gaming, game deployment on cloud, game operations analysis, and game security.
  - Government and public utilities: remote sensing, smart meteorology, smart finance, and traffic intelligence.
  - Manufacturing: cloud MES, cloud simulation, intelligent coal blending, predictive maintenance
  - Industrial Internet: intelligent platforms for industrial transformation and upgrade
  - Energy: PV cloud network, smart charging, oil and gas exploration and development, and electric power data platform.
  - Automobile: IoV, automobile simulation, digital marketing, and self-driving development platform.
  - Smart city: e-Government cloud, e-Government big data, and city intelligent operation center.
  - Healthcare and life sciences: biomedicine, genomic sequencing, medical cloud, and medical images

- Education: higher education, talent cultivation, and online education

## Solutions by Use Case

- HUAWEI CLOUD pre-integrates products and capabilities to meet the requirements of running ICT businesses on the cloud.



- HUAWEI CLOUD provides the following general-purpose solutions (for details, visit the HUAWEI CLOUD official website):
  - Video: live video, VOD, video transcoding, 5G UHD production and broadcasting, cloud exhibition
  - Security: cloud security practices, compliance with graded protection requirements, website security, defense against brute force cracking attacks on cloud hosts, and general security.
  - Scientific computing: quantum computing, high-performance computing
  - Data platforms: data enablement platform DAYU, intelligent air traffic control data enablement platform
  - Business applications: ERP on cloud, core big data on cloud, Microsoft applications on cloud
  - Smart campus: industrial parks, office parks, logistics parks, education parks, smart communities, smart construction sites, and chemical parks.
  - Basic solutions: IPv6, cloud VR, application performance optimization, KYON enterprise-class cloud network, and BigData Pro.
  - Blockchain: e-Government services, financial innovation
  - DR and backup: cloud DR, backup and archiving.
  - Application platforms: ROMA
  - SAP on cloud: DeC, entire system on cloud, DR system on cloud, and development and test system on cloud.
  - Edge-cloud synergy: smart highway, smart logistics, smart campus security, and higher education data governance.
  - Mobile Internet: websites, mobile apps
  - DevOps: software training, game development, and e-commerce delivery
  - Enterprise office: enterprise cloud disk

## Solutions by Organization Type

- HUAWEI CLOUD provides solutions for helping enterprises, public welfare and non-profit organizations, and HMS partners to migrate to the cloud with ease.



- For more details, visit <https://www.huaweicloud.com/intl/en-us/solution/>.

## HUAWEI CLOUD Is Powering Every Industry



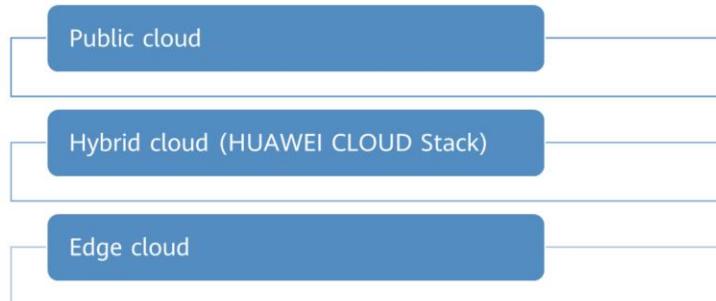
- Leveraging more than 30 years of experience in ICT infrastructure construction, HUAWEI CLOUD has been empowering enterprises in every industry and has established partnership with many customers. For more success stories, visit the HUAWEI CLOUD official website at <https://www.huaweicloud.com/intl/en-us/cases.html>.

# Contents

1. HUAWEI CLOUD Overview
2. Application Scenarios
- 3. Delivery Modes**
4. Technical Support
5. Ecosystem
6. Quick Start

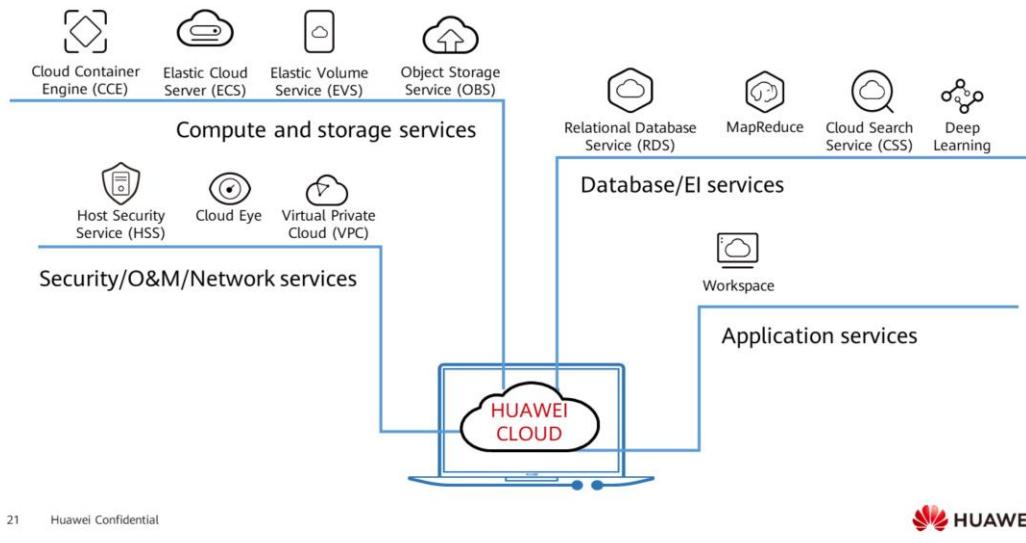
## Delivery Modes

- Different users may have different requirements for cloud service delivery. HUAWEI CLOUD provides the following delivery modes:



- The three delivery modes will be described in detail later.

## Public Cloud



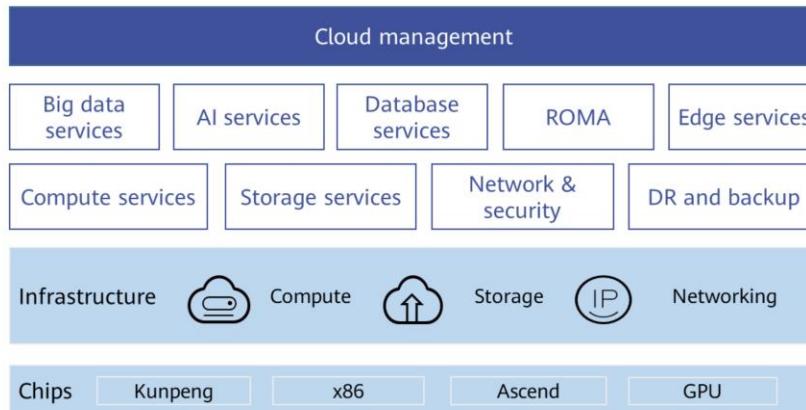
21 Huawei Confidential



- As an individual or representing an enterprise, you can use the IT infrastructure services from HUAWEI CLOUD without building up your own data center. HUAWEI CLOUD is always advancing our database and enterprise intelligence products and is always developing new cloud products to keep up with your requirements. You can quickly obtain infrastructure services and business applications from HUAWEI CLOUD. The public cloud ensures network security, system security, and data security. You can migrate IT systems to the cloud and set your mind at ease.
- You just need to connect to servers, storage devices, and platforms on the public cloud using an Internet connection. If security is an issue, you can use a carrier's private line, or use VPN, which may be a more cost-effective option.

## HUAWEI CLOUD Stack

- HUAWEI CLOUD Stack is a cloud infrastructure deployed at on-premises data centers.



22 Huawei Confidential



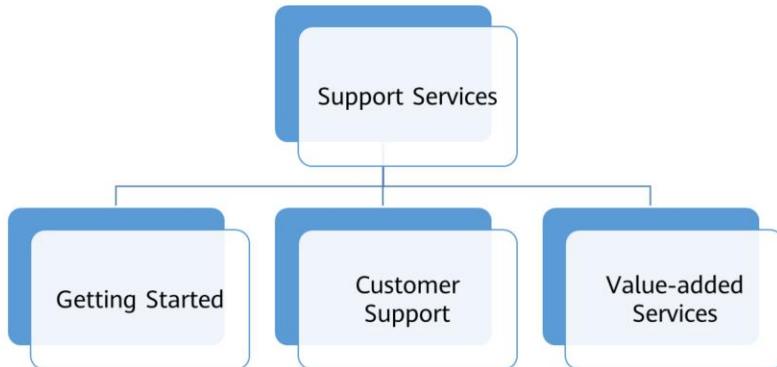
- HUAWEI CLOUD Stack combines the advantages of private and public clouds. You can quickly launch innovative services the way you do on a public cloud, and you can manage your resources the way you do on the private cloud. HUAWEI CLOUD Stack can adjust to your organizational structure and business processes, serving you as a single, unified cloud.
- The HUAWEI CLOUD Stack is ideal for medium- and large-sized enterprises that require local data storage or need physical devices to be physically isolated.
- HUAWEI CLOUD Stack can be used for cloud migration, cloud native transformation, big data analysis, AI applications, industry clouds, and city clouds.
- Advantages:
  - On-premises deployment of three types of enablement services: AI enablement, data enablement, and application enablement
  - Multi-level cloud management that perfectly matches the organizational architecture of governments and enterprises
  - AIOps and cloud federation
  - Cloud-edge synergy with intelligence extended to the edge; a unified edge framework, and quick access to AI video devices and IoT devices
  - Full-stack security and trustworthiness, leading functionality and performance
  - One unified cloud with two separate resource pools and a flourishing ecosystem

# Contents

1. HUAWEI CLOUD Overview
2. Application Scenarios
3. Delivery Modes
- 4. Technical Support**
5. Ecosystem
6. Quick Start

## HUAWEI CLOUD Technical Support

- HUAWEI CLOUD provides efficient service assurance and diverse support plans to walk you through the cloud migration process.



- HUAWEI CLOUD provides the following support services: getting started, customer support, and value-added services.

## Getting Started

### Getting started with cloud products

Learn about compute, network, storage, application, EI, database, security, and migration services. Familiarize yourself with common operations in minutes.

### Hands-on labs

Log in to the HUAWEI CLOUD website, register an account, and try out the cloud services.

### Online courses

Take the online courses designed for your role. These online courses can help sharpen your cloud skills and expertise.

- You can learn more about HUAWEI CLOUD services by reviewing infographics, taking online courses, and reading best practices. For more details, visit <https://support.huaweicloud.com/intl/en-us/help-novice.html>.

## Customer Support

Chatbot	<ul style="list-style-type: none"><li>Intelligent diagnosis, quick response, and fast troubleshooting.</li></ul>
Self-service	<ul style="list-style-type: none"><li>Convenient O&amp;M tools and answers to frequently asked questions.</li></ul>
Contact us	<ul style="list-style-type: none"><li>Professional pre-sales consulting and after-sales support services.</li></ul>
Service assurance	<ul style="list-style-type: none"><li>24/7 support and free ICP filing.</li></ul>
Product notices	<ul style="list-style-type: none"><li>HUAWEI CLOUD official bulletins.</li></ul>

- HUAWEI CLOUD provides multiple channels for customers to reach out. For details, visit <https://www.huaweicloud.com/intl/en-us/service/protection.html>.
- You can take advantage of self-service channels to:
  - Query arrears, spending details, invoices, and unsubscriptions in the Billing Center.
  - Associate or disassociate customers and assign discounts in the Partner Center.
  - Complete real-name verification, as an individual or as an enterprise.
  - Request ICP filing.
  - Create IAM users and assign permissions in the Identity and Access Management (IAM) console.

## Value-added Services



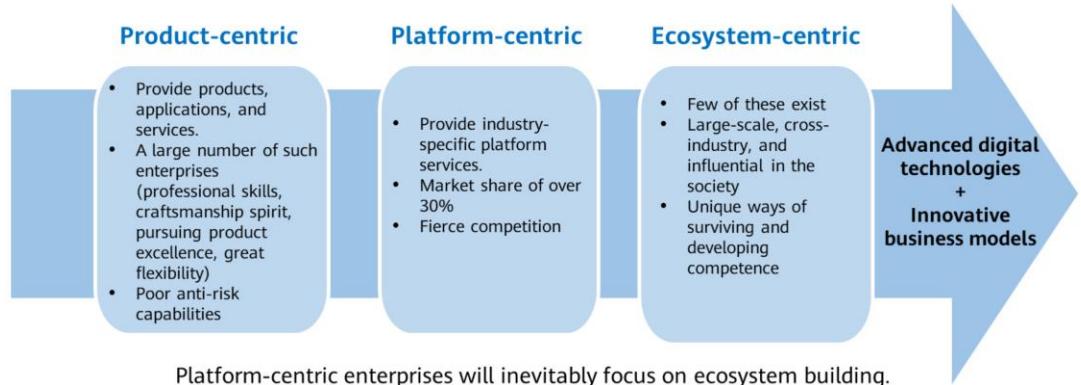
- HUAWEI CLOUD provides professional support services and customized training to address your requirements.
- Professional services cover cloud deployment, cloud management, expert services, and training.
- Training services cover cloud services, artificial intelligence, intelligent data & data warehouse, IoT, Kunpeng, agile DevOps, and WeLink.

# Contents

1. HUAWEI CLOUD Overview
2. Application Scenarios
3. Delivery Modes
4. Technical Support
- 5. Ecosystem**
6. Quick Start

## New Tech Gives Rise to New Products, and Ecosystems

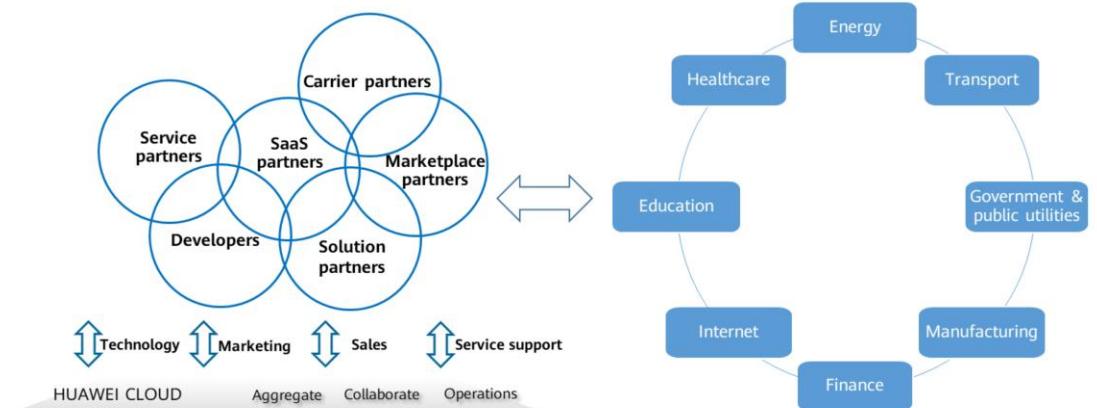
- Digital technologies are reshaping the business models of various industries.



- Enterprises pursue business growth and continuous profitability. To achieve such goals, they need to improve quality, increase efficiency, and keep innovating while simultaneously reducing costs and ensuring security and compliance. Reducing costs means not only reducing the total cost of ownership (TCO) of ICT, but also means reducing the overall production and operations costs through digital transformation.
- Enterprises start small and gradually grow big. How can they become stronger? The only way is to evolve into an ecosystem-centric enterprise. In today's Internet era, most enterprises start with a platform-centric strategy and then they gradually move towards an ecosystem-centric strategy to survive in highly competitive markets. Small enterprises will evolve into small ecological enterprises. This ecosystem is no longer an ecosystem of organizations but an ecosystem of teams.
- Huawei has always been helping enterprises grow bigger and stronger. After more than three years of hard work, HUAWEI CLOUD has attracted more than 19,000 partners, including more than 13,000 consulting partners and more than 6,000 technology partners. In addition, Huawei has established strategic cooperation with top consulting companies and carriers around the world. More than 4,000 applications have been released in the cloud market, and annual transactions exceed CNY1 billion. HUAWEI CLOUD is joining hands with partners around the world to create a prosperous cloud ecosystem and to benefit from

digital transformation.

## Multiple Parties Joining Hands for Digital Transformation



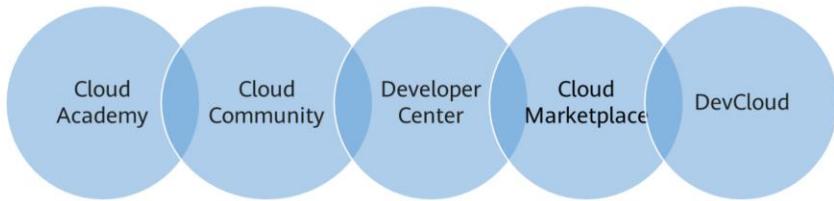
30 Huawei Confidential



- Service providers, marketplace partners, solution partners, SaaS partners, carrier partners, and developers are all important components of the HUAWEI CLOUD ecosystem. Only when they work together, can the public cloud develop faster. The only way to succeed is to partner up to better serve our customers.

## HUAWEI CLOUD Developer Support Platform

- A one-stop platform for developers to obtain resources and toolkits and interact with each other.



- HUAWEI CLOUD provides a professional support platform for developers to quickly get started with HUAWEI CLOUD. For details, visit <https://developer.huaweicloud.com/intl/en-us/>.

# HUAWEI CLOUD Marketplace Seller Learning Center

- HUAWEI CLOUD Marketplace provides end-to-end services for service providers, making registration and operations easier.

Negotiation	Registration	Operations and marketing	Monetization
 Advisory services <ul style="list-style-type: none"><li>• Professional consulting</li><li>• Business design</li><li>• Package design</li><li>• Registration</li></ul>	 Technical support <ul style="list-style-type: none"><li>• Resources for testing</li><li>• Solution testing</li><li>• Security testing</li><li>• Fault rectification</li></ul>	 Operations <ul style="list-style-type: none"><li>• Operations dashboard</li><li>• Recommended on homepage</li><li>• Displayed as popular products</li><li>• Brand pavilion</li></ul>	 Monetization <ul style="list-style-type: none"><li>• Online reconciliation</li><li>• Monthly settlement</li><li>• Quick payment</li></ul>
 Market analysis <ul style="list-style-type: none"><li>• Documentation</li><li>• Marketing strategy</li><li>• Development trends</li></ul>	 Training <ul style="list-style-type: none"><li>• Standard process</li><li>• New product launch</li><li>• Knowledge/Skill transfer</li><li>• On-site/Online training</li></ul>	 Marketing <ul style="list-style-type: none"><li>• Huawei exhibitions</li><li>• Brand events</li><li>• Forums &amp; Salons</li><li>• HUAWEI CLOUD partners</li></ul>	 Incentives <ul style="list-style-type: none"><li>• Partner benefits</li><li>• Annual MVPs</li></ul>

- HUAWEI CLOUD Marketplace Seller Learning Center:  
<https://marketplace.huaweicloud.com/intl/>.

# Contents

1. HUAWEI CLOUD Overview
2. Application Scenarios
3. Delivery Modes
4. Technical Support
5. Ecosystem
- 6. Quick Start**

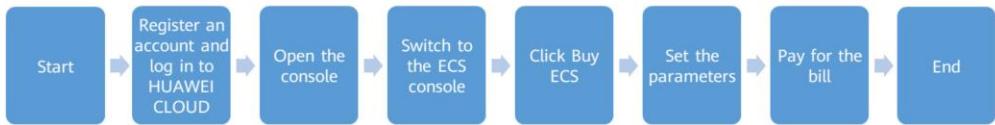
## How to Use HUAWEI CLOUD?

- How do we use HUAWEI CLOUD?
- What should we pay special attention to?



- Create an individual or enterprise account
- View service menus
- Access the console
- Buy resources
- View resource models

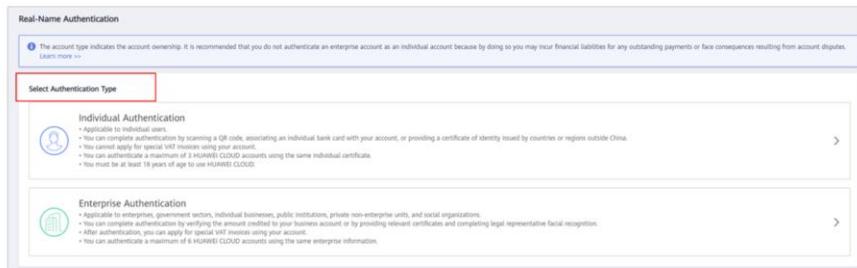
## Buy an ECS



- For details about how to purchase an ECS, see the following courses and exercises.
- Note:
  - Top up your account so that you can pay for your ECS.
  - After finishing using the ECS, delete the ECS and release resources to avoid unnecessary expenses.

## Register an Enterprise Account

- Log in to the HUAWEI CLOUD website at <https://www.huaweicloud.com/>.
- Register a HUAWEI CLOUD account.
- Verify the account.
  - After the registration is complete, use the registered account to log in to HUAWEI CLOUD. A message is displayed, indicating that the enterprise authentication needs to be completed or an email address needs to be added. Click OK.
  - Choose Real-Name Authentication > Enterprise Authentication and complete the real-name authentication.
- Add your Email address.



- The differences between an individual account and an enterprise account are as follows:
  - If the account type is **Individual**, the authenticated account belongs to an individual.
  - If the account type is **Enterprise**, the authenticated account belongs to an enterprise.
  - The account type determines its ownership. For example, after an enterprise user completes individual real-name authentication, the account belongs to the individual and this individual is responsible for any potential account arrears.

## What Is Console?

- The console is where you can request and manage cloud service resources.



## HUAWEI CLOUD Billing Modes

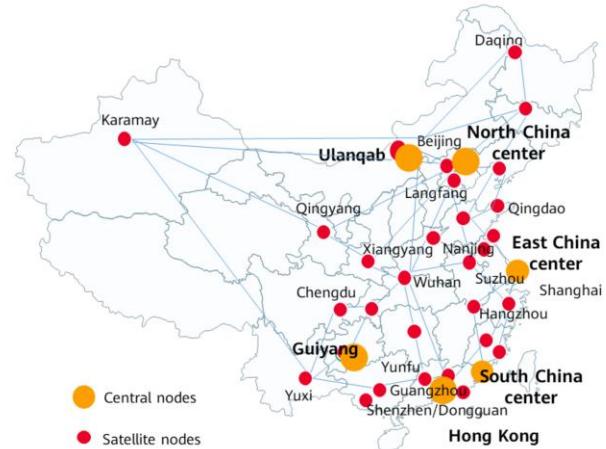
- The billing modes for cloud services vary. The following billing modes are available for ECS:



- Pay-per-use: a form of postpaid billing where you are billed for your service usage the month after you used the service
- Yearly/Monthly: a prepaid billing mode where you are charged for your service usage upfront. This cost-effective mode is ideal when the duration of ECS usage is predictable.
- Spot price: a postpaid billing mode where you are charged for your service usage at a discounted price compared to pay-per-use billing.

## What Is Region?

- Regions are determined based on geographical location and network latency. Public services, such as Elastic Cloud Server (ECS), Elastic Volume Service (EVS), Object Storage Service (OBS), Virtual Private Cloud (VPC), Elastic IP, and Image Management Service (IMS), are shared within the same cloud region.
- Regions are classified as either universal or dedicated. A universal region provides cloud services for all tenants. A dedicated region provides only services of a certain type or only for certain tenants.

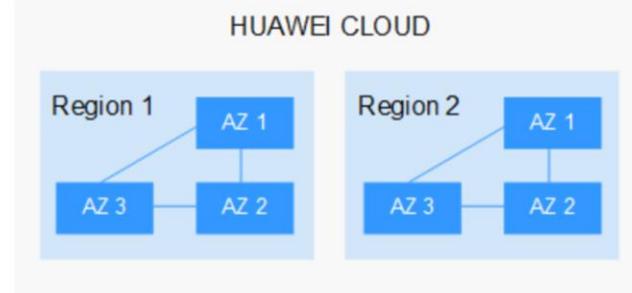


- A region represents an independent physical data center. Select the region nearest your customers for the best responsiveness.
- In China, "2+7+N" centers and nodes are connected through high-speed networks to provide cloud services nationwide.
- The "2" in "2+7+N" refers to the data centers in Ulanqab and Guiyang, the level-1 centers built by HUAWEI CLOUD. In each level-1 center, three AZs are built to achieve high availability. The distance between any two AZs is 30 to 50 kilometers.
- The "7" in "2+7+N" refers to main HUAWEI CLOUD regions, including North China, East China, South China, and Hong Kong.
- The "N" in "2+7+N" refers to satellite nodes. Each node works as an e-Government cloud and a node of HUAWEI CLOUD. Currently, there are five satellite nodes: Ulanqab, Xiangyang, Yuxi, and Karamay.
- When selecting a region, consider the following factors:
  - Latency: Select a region closest to you or your target users to ensure low network latency and quick access. The regions in Chinese mainland use basically the same infrastructure, BGP networks, and resources. If you or your target users are in Chinese mainland, select any region as the latency to these regions is almost the same.

- Resource prices: The prices may vary by region.

## What Is AZ?

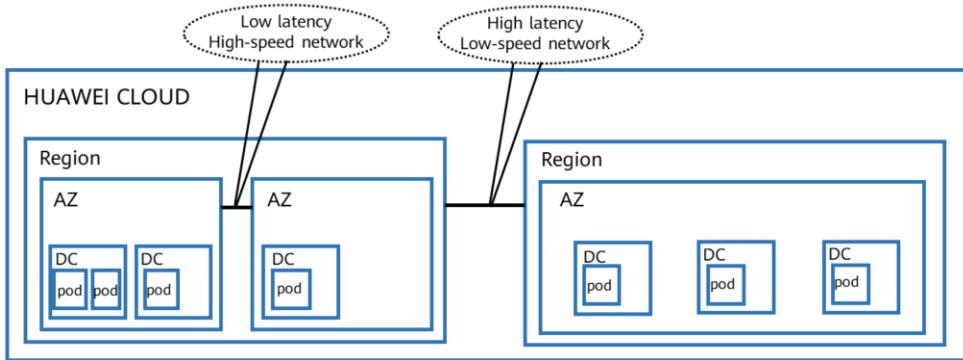
- An AZ contains one or more physical data centers. Each AZ has independent cooling, fire extinguishing, moisture-control, and electrical facilities. Within an AZ, compute, networking, storage, and other resources are logically divided into multiple clusters. AZs within a region are interconnected using high-speed optical fibers to ensure high availability.



- Which AZ should I select? When selecting your AZs, consider your requirements for disaster recovery (DR) and network latency.
  - For high DR, deploy applications across different AZs.
  - For low latency, deploy applications in the same AZ.

## Differences Between Regions and AZs

- An AZ is a subdivision of a region. A region can contain multiple AZs.



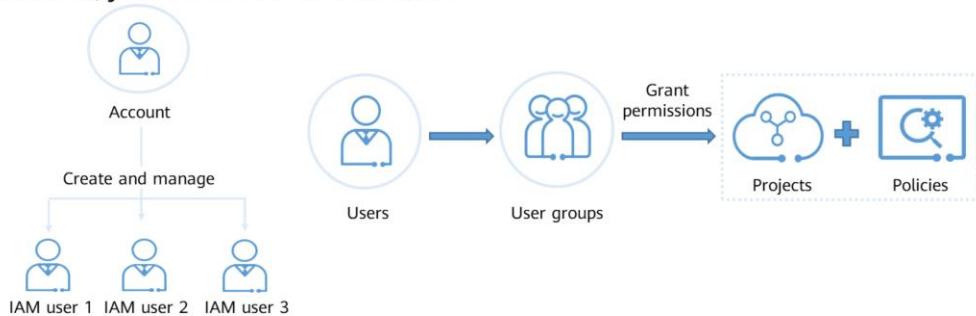
41 Huawei Confidential



- A DC refers to a data center or an equipment room. A Pod is a resource pool managed by a virtualization platform or an independent cloud software instance. The latency to a pod is equivalent to that to an AZ.
- Select the nearest region to obtain the lowest latency. The region cannot be changed after the resources are provisioned. If intra-city or remote DR is required, you can deploy applications across regions.
- If high availability is required, you can deploy applications across AZs.

## What Is IAM?

- Identity and Access Management (IAM) is used to manage user access to cloud resources. If you want to share resources with others but do not want to share your account, you can create an IAM user.



## What Is a Project?

### Definition

- A project is used by IAM to group and isolate resources in the same region.

### Features

- A project is used for physical isolation. Resources cannot be transferred between IAM projects. They can only be deleted and then provisioned again.

## What Is an Enterprise Project?



- The Enterprise Management Service also supports projects, which are called enterprise projects. Enterprise projects group and manage resources across regions.

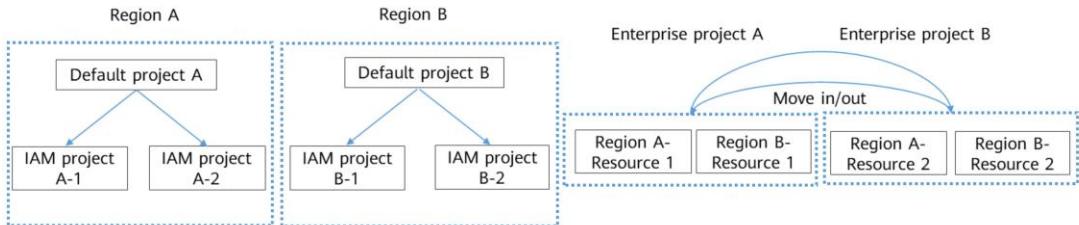


- An enterprise project can contain resources of multiple regions, and resources can be added to or removed from enterprise projects. If you have enabled enterprise management, you cannot create an IAM project. You can only manage existing projects.

- IAM projects will be replaced by enterprise projects.

## Differences Between IAM Projects and Enterprise Projects

- IAM projects group and physically isolate resources in the same region. Resources cannot be transferred between IAM projects. They can only be deleted and then provisioned again.
- Enterprise projects group and logically isolate resources. An enterprise project can contain resources from multiple regions, and resources can be added to or removed from enterprise projects. Enterprise projects can be used to grant permissions to use specific cloud resources.



## Quiz

1. (True or False) Can HUAWEI CLOUD services be billed by minute?
  - A. T
  - B. F
2. (Multiple-answer) Which of the following statements about regions and AZs are incorrect?
  - A. An AZ is larger than a region.
  - B. An AZ belongs to a region.
  - C. A region can contain multiple AZs.
  - D. A region can contain only one AZ.

- F. Currently, only the pay-per-use, monthly/yearly, and spot price are available.
- A, D. A region is larger than an AZ and can contain multiple AZs.

## Summary

- We provided a brief introduction to HUAWEI CLOUD, including the advantages of nodes, cloud products, solutions, ecosystems, and technical support systems, the positioning of HUAWEI CLOUD, and how to better use HUAWEI CLOUD.

# Recommendations

- Huawei Learning
  - <https://e.huawei.com/en/talent/#/>
- HUAWEI CLOUD technical support
  - <https://support.huaweicloud.com/intl/en-us/help-novice.html>
- HUAWEI CLOUD Academy
  - <https://edu.huaweicloud.com/intl/en-us/>

## Acronyms and Abbreviations

- AI: Artificial intelligence
- AS: Auto Scaling
- APM: Application Performance Management
- AOM: Application Operations Management
- AZ: Availability Zone
- API: Application Programming Interface
- BMS: Bare Metal Server
- BCS: Hyperledger Fabric
- CCE: Cloud Container Engine
- CDN: Content Delivery Network

## Acronyms and Abbreviations

- CBH: Cloud Bastion Host
- CPTS: Cloud Performance Test Service
- CAE: Computer Aided Engineering
- CES: Cloud Eye Service
- CTS: Cloud Trace Service
- CCS: Cloud Catalog Service
- CRS: Cloud Record Service
- CDM: Cloud Data Migration
- CMC: Cloud Migration Center
- DES: Data Express Service

## Acronyms and Abbreviations

- DNS: Domain Name Service
- DDS: Document Database Service
- DDM: Distributed Database Middleware
- DAS: Data Admin Service
- DBSS: Database Security Service
- DMS: Distributed Message Service
- DWS: Data Warehouse Service
- DevOps: Development and Operations
- ECS: Elastic Cloud Server
- EVS: Elastic Volume Service

## Acronyms and Abbreviations

- ELB: Elastic Load Balance
- EI: Enterprise Intelligence
- ERP: Enterprise Resource Planning
- GES: Graph Engine Service
- HMS: Huawei Mobile Service
- HSS: Host Security Service
- ICT: Information and Communications Technology
- IMS: Image Management Service
- IAM: Identity and Access Management
- IOPS: Input/Output Operations Per Second

## Acronyms and Abbreviations

- I/O: Input/Output
- LTS: Log Tank Service
- MVP: Most Valuable Player
- MRS: MapReduce Service
- NIC: Network Interface Controller
- OBS: Object Storage Service
- OCR: Optical Character Recognition
- OMS: Object Storage Migration Service
- OVS: Open Virtual Switch
- QoS: Quality of Service

## Acronyms and Abbreviations

- RoCE: RDMA over Converged Ethernet (a network protocol that allows remote direct memory access (RDMA) over the Ethernet.)
- RDS: Relational Database Service
- SDK: Software Development Kit
- SFS: Scalable File Service
- SA: Situation Awareness
- SMN: Simple Message Notification
- SWR: SoftWare Repository for Container
- SMS: Server Migration Service
- SSD: Solid State Disk

## Acronyms and Abbreviations

- SAP: System Applications and Products, an enterprise management software company in Germany
- TMS: Tag Management Service
- TTS: Text-To-Speech
- VBS: Volume Backup Service
- VPC: Virtual Private Cloud
- VPN: Virtual Private Network
- VxLAN: Virtual Extensible Local Area Network
- WAF: Web Application Firewall

# Thank you.

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

Copyright©2021 Huawei Technologies Co., Ltd.  
All Rights Reserved.

The information in this document may contain predictive  
statements including, without limitation, statements regarding  
the future financial and operating results, future product  
portfolio, new technology, etc. There are a number of factors that  
could cause actual results and developments to differ materially  
from those expressed or implied in the predictive statements.  
Therefore, such information is provided for reference purpose  
only and constitutes neither an offer nor an acceptance. Huawei  
may change the information at any time without notice.



## Compute Cloud Services



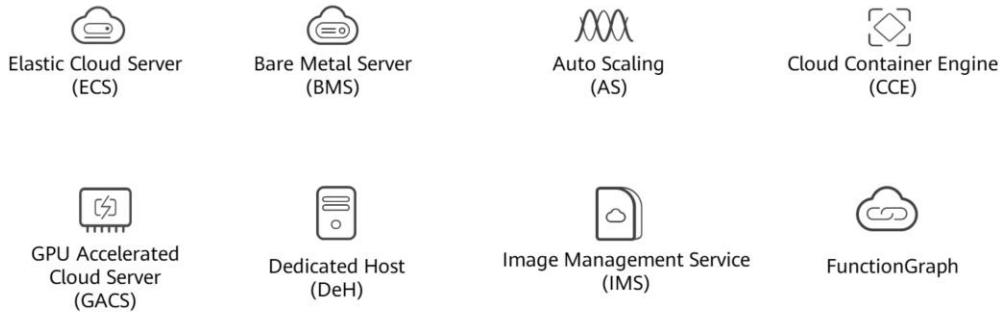
## Foreword

- Compute resources are essential to the development of enterprise service systems. For cloud computing, compute services are the most important cloud services.
- In this section, we will learn about the compute services on HUAWEI CLOUD.

# Objectives

- Upon completion of this course, you will:
  - Understand common compute services available on HUAWEI CLOUD.
  - Understand the positioning, technical details, and usage of these compute services.

## Compute Cloud Services

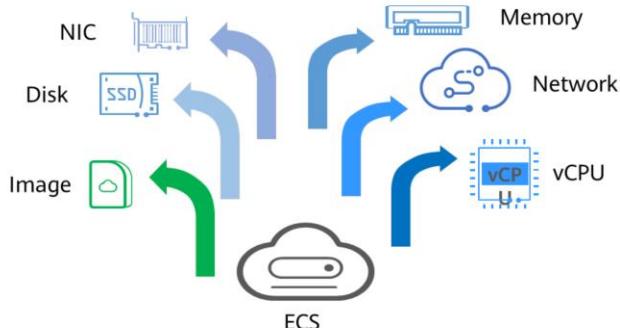


# Contents

- 1. Elastic Cloud Server (ECS)**
2. Bare Metal Server (BMS)
3. Image Management Service (IMS)
4. Auto Scaling (AS)
5. Cloud Container Engine (CCE)
6. Other Compute Services

## What Is ECS?

- An ECS is a basic computing unit that consists of vCPUs, memory, an OS, and Elastic Volume Service (EVS) disks. After an ECS is created, you can use it on the cloud similarly to how you would use your local computer or physical server.



6      Huawei Confidential

HUAWEI

- ECS has the following advantages:

- A variety of specifications to choose from: Different ECS types are available for different applications. There are multiple, customizable specifications for each type.
- A wide range of available images: Public, private, and shared images can be selected.
- Different types of EVS disks: Common I/O, high I/O, general-purpose SSD, and ultra-high I/O EVS disks are available for different service requirements.
- Flexible billing: Yearly/monthly and pay-per-use billing modes are available for different applications. You can purchase and release resources as service levels fluctuate.
- Reliable data: Virtual block storage based on a distributed architecture provides robust throughput that is scalable and reliable.
- Security: The network is isolated from viruses and Trojans by security group rules. Security services, such as Anti-DDoS, Web Application Firewall and Vulnerability Scan Service are also available to protect your ECSs.
- Flexible, easy-to-use: Elastic computing resources are automatically adjusted based on service requirements and policies to efficiently meet service requirements.
- Highly efficient O&M: Multi-choice management via the management console, remote access, and APIs with full management permissions.
- Cloud monitoring: Cloud Eye monitors your ECSs in real time, generating alarms and sending notifications when it detects abnormal metrics.
- Load balancing: Elastic Load Balance automatically distributes traffic to multiple ECSs to keep the loads on the servers balanced. It improves the fault tolerance of your applications and enhances application capabilities.

## Why ECS?

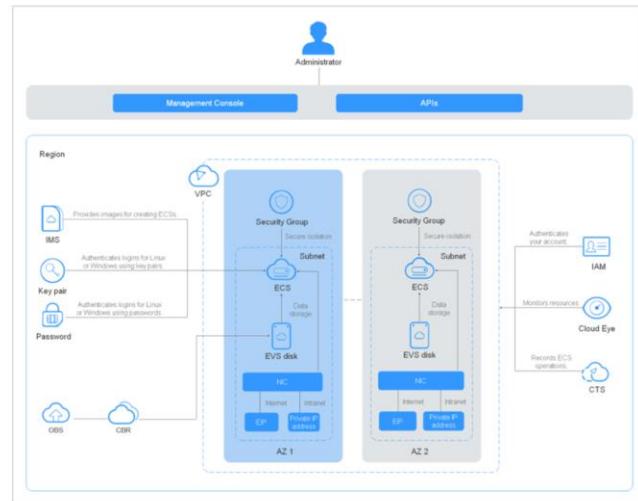


7      Huawei Confidential



- **Stability and Reliability**
  - A variety of EVS disk types: Common I/O, high I/O, ultra-high I/O, general purpose SSD, and extreme SSD disks are available for different service requirements.
  - Reliable data: Scalable, reliable, high-throughput virtual block storage is based on a distributed architecture. This architecture ensures that data can be quickly migrated or restored, if necessary, which means you will not lose your data as the result of a single hardware fault.
  - Backup and restoration of ECSs and EVS disks: You can configure backup policies on the management console or use an API to back up ECSs and EVS disks periodically or at a specified time.
- **Security Protection**
  - A range of security services provide multi-dimensional protection: Security services, such as Web Application Firewall and Vulnerability Scan Service, are available to protect your ECSs.
  - Security evaluation: The security of cloud environments is evaluated to help you quickly detect security vulnerabilities and threats. Security configurations are reviewed and suggestions provided on how to improve system security. Actions will be recommended to reduce or avoid altogether potential losses resulting from viruses or other malicious attacks.

## ECS Architecture

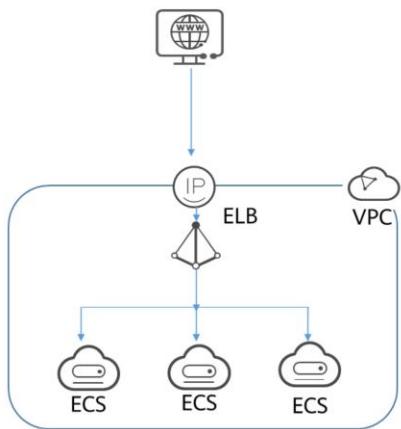


9      Huawei Confidential



- ECS works with other products and services to provide computing, storage, network, and image installation functions.
  - ECSs are deployed in multiple Availability Zones (AZs) connected with each other through an intranet. If an AZ becomes faulty, other AZs in the same region will not be affected.
  - With the Virtual Private Cloud (VPC) service, you can build a dedicated network, configure subnets and security groups, and allow the VPC to communicate with the external network through an EIP with bandwidth assigned.
  - With the Image Management Service (IMS), you can create images for ECSs, or create ECSs using private images for rapid service deployment.
  - EVS provides storage and Volume Backup Service (VBS) provides data backup and recovery functions.
  - Cloud Eye is a key service to help ensure ECS performance, reliability, and availability. You can use Cloud Eye to monitor ECS resource usage.
  - Cloud Backup and Recovery (CBR) backs up data for EVS disks and ECSs and creates snapshots in case you need to restore them.

## Scenarios – Internet



### Application Scenarios

Website R&D and testing, and small-scale databases

### Recommended ECS

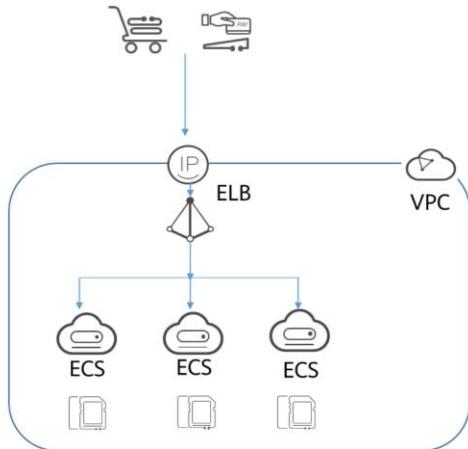
General-computing ECSs

### Recommendation Reasons

- Requirements: To minimize upfront deployment and O&M costs, applications need to be deployed on only one or just a few servers, but there are no special requirements for CPU performance, memory, disk capacity, or bandwidth, strong security and reliability.
- Solution: General-computing ECSs provide a balance of compute, memory, and network resources. They are appropriate for medium-workload applications and meet the cloud service needs of both enterprises and individuals.

- General-computing ECSs provide a balance of compute, memory, and network resources and a baseline level of vCPU performance with the ability to burst above the baseline. These ECSs are suitable for many applications, such as web servers, enterprise development, and small databases.

## Scenarios – E-Commerce



### Application Scenarios

Precision marketing, E-Commerce, and mobile apps

### Recommended ECS

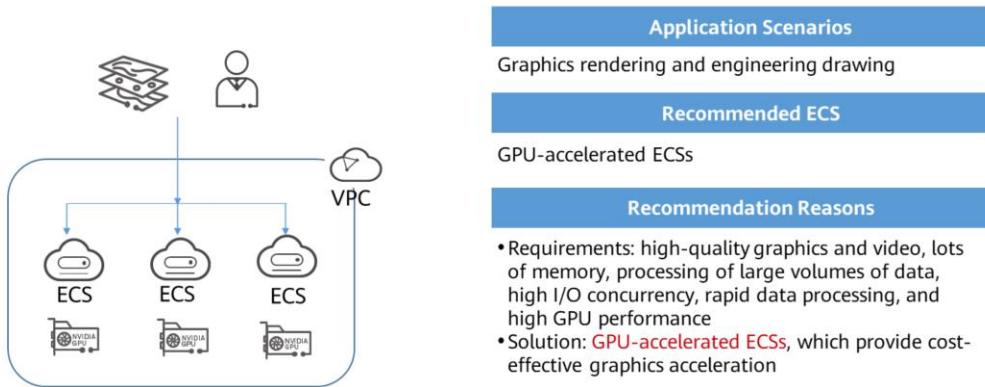
Memory-optimized ECSs

### Recommendation Reasons

- Requirements: large amount of memory, rapid processing of large volumes of data, and fast network access
- Solution: **memory-optimized ECSs**, which feature a large amount of memory, ultra-high I/O EVS disks, and appropriate bandwidths

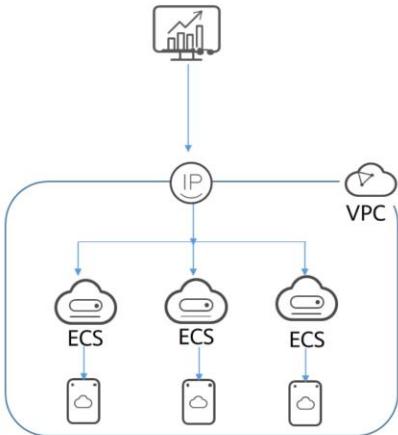
- Memory-optimized ECSs have a large memory size and provide high memory performance. They are designed for memory-intensive applications that involve a large amount of data, such as precision advertising, e-commerce big data analysis, and IoV big data analysis.
- E-commerce presents special challenges.
  - Sudden Traffic Surges: Access traffic can surge to hundreds of times normal levels during promotions, flash sales, and sweepstakes. Servers become overloaded and e-commerce platforms may even crash.
  - Poor User Experience: Massive amounts of static data, such as product pictures and videos content, is usually stored on servers, resulting in slow loading, time-consuming and costly. Users in different network environments may experience delayed access to such data, resulting in poor user experience.
  - Lack of Proper Analytics: Due to the lack of big data platforms and analysis tools, existing customers, financial products, and transaction data cannot be effectively analyzed. As a result, there are problems such as high promotion investment and low second-order rate.
  - Security: E-commerce enterprises have to deal with risks in various processes, such as traffic diversion, registration and login, browsing and comparison, preference obtaining, ordering, payment, delivery, and evaluation. The vulnerabilities may come from credential stuffing, scalpers, web page tampering, DDoS attacks, data breaches, and Trojans.

## Scenarios – Graphics Rendering



- GPU-accelerated ECSs provide outstanding floating-point computing capabilities. They are suitable for applications that require real-time, highly concurrent massive computing.
- GPU-accelerated ECSs are classified as G series and P series ECSs.
  - G series: Graphics-accelerated ECSs, which are suitable for 3D animation rendering and CAD
  - P series: Computing-accelerated or inference-accelerated ECSs, which are suitable for deep learning, scientific computing, and CAE

## Scenarios – Data Analysis



### Application Scenarios

MapReduce and Hadoop

### Recommended ECS

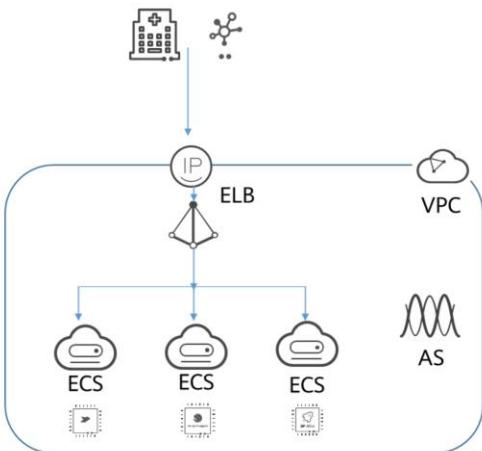
Disk-intensive ECSS

### Recommendation Reasons

- Requirements: processing of large volumes of data; high I/O performance and rapid data switching and processing
- Solution: **disk-intensive ECSS**, which are suitable for applications requiring high-performance sequential read/write on ultra-large datasets in local storage

- Disk-intensive ECSS are delivered with local disks for high storage bandwidth and IOPS. In addition, local disks are more cost-effective in massive data storage scenarios. Disk-intensive ECSS use local disks to provide high sequential read/write performance and low latency, improving file read/write performance.
- They provide powerful and stable computing capabilities, ensuring efficient data processing.
- They provide high intranet performance, including robust intranet bandwidth and PPS, for data exchange between ECSS during peak hours.
- Disk-intensive ECSS are suitable for distributed Hadoop computing, large-scale parallel data processing, and log processing. Disk-intensive ECSS use hard disk drives (HDDs) and a default network bandwidth of 10GE, providing high PPS and low network latency. Each disk-intensive ECS supports up to 24 local disks, 48 vCPUs, and 384 GiB of memory.

## Scenarios – High-Performance Computing



### Application Scenarios

Computing and storage systems for scientific computing, genetic engineering, games, animations, and biopharmaceuticals

### Recommended ECS

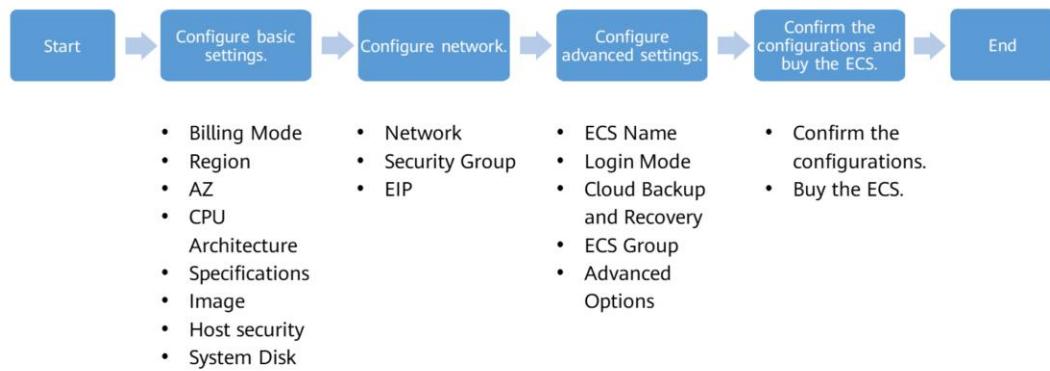
High-performance computing ECSs

### Recommendation Reasons

Solution: **high-performance computing ECSs**, which meet the computing, storage, and rendering needs of high-performance infrastructure services and applications that require a large number of parallel computing resources.

- Each vCPU of a high-performance computing ECS corresponds to the hyper-thread of an Intel® Xeon® Scalable processor core. High-performance computing ECSs are suitable for high-performance computing scenarios. They provide massive parallel computing resources and high-performance infrastructure services to meet the requirements of high-performance computing and massive storage and ensure rendering efficiency.

## Purchasing an ECS



- Select a billing mode, yearly/monthly or pay-per-use.
  - You can purchase a yearly/monthly ECS subscription and enter your required duration. Yearly/monthly subscriptions are pre-paid, using a single, lump sum payment.
  - If you choose pay-per-use billing, you do not need to choose a required duration. Pay-per-use usage is postpaid.
- Select required specifications: HUAWEI CLOUD provides various ECS types for you to select based on different applications. You can view the available ECS types and specifications in the list. Alternatively, you can enter a flavor (such as c3) or search for a flavor by vCPU and memory.
- Set Network by selecting an available VPC and subnet from the drop-down list, and specifying a private IP address assignment mode. You can also create a VPC if needed. VPC provides a network, including subnets and security groups, for an ECS.
- Set EIP. If you want the ECS to connect to the Internet, it needs to have an EIP bound.
- Set Login Mode. Key pair is recommended because key pair authentication is more secure than using a password.

# Configuring Basic Settings

- Set Billing Mode, Region, AZ, CPU Architecture, and Specifications.

The screenshot shows two panels of the Huawei Cloud ECS configuration interface. The top panel is for 'Basic Settings' and includes tabs for 'Yearly/Monthly', 'Pay-per-use' (selected), and 'Spot price'. It shows a dropdown for 'Region' set to 'AP-Singapore', with other options like 'CN Southwest...' and 'CN North-Beijing4'. Below this is a note about selecting the nearest region for low latency. The bottom panel is for 'CPU Architecture' and shows tabs for 'x86' (selected) and 'Kunpeng'. It includes dropdowns for 'Latest generation', 'vCPUs', 'Memory', and a 'Flavor Name' search bar. Both panels have red boxes highlighting the 'Billing Mode' section and the 'CPU Architecture' section.

16      Huawei Confidential



- Configure basic settings.
  - **Billing Mode:** An ECS can be billed on a pay-per-use, yearly/monthly, or spot price basis. For yearly/monthly subscriptions, the longer the subscription, the more you save.
  - **Region and AZ:** ECSSs in different regions cannot communicate with each other over an intranet. Select a region closest to your target users to ensure low network latency and quick access.
  - **CPU Architecture:** x86-based CPUs use Complex Instruction Set Computing (CISC). Kunpeng CPUs use Reduced Instruction Set Computing (RISC).
  - **Specifications:** Select a flavor and image based on service requirements.
- Select an ECS type.
  - General computing-plus ECSSs are suitable for governments, enterprises, and the financial industry, where there are strict requirements for security and privacy; for Internet applications, which demand excellent network performance; for big data and HPC, which require a lot of vCPUs; and for website setups and e-Commerce, which need to be cost-effective.
  - Memory-optimized ECSSs are designed for memory-intensive applications that process a large amount of data, such as precision advertising, e-commerce big data analysis, and IoT big data analysis.
  - Ultra-high I/O ECSSs are designed for high-performance relational databases, NoSQL databases (such as Cassandra and MongoDB), and ElasticSearch.
  - GPU-accelerated ECSSs are suitable for applications that require real-time, highly concurrent massive computing.
  - FPGA-accelerated ECSSs are suitable for video processing, machine learning, genomics research, and financial risk analysis.
  - AI1-accelerated ECSSs are used for general technologies, such as machine vision, voice recognition, and natural language processing to support smart retail, smart campus, robot cloud brain, and safe city scenarios.

## Configuring Network

- Select a VPC, subnet, and security groups for the ECS.

The screenshot shows two panels of the Huawei Cloud network configuration interface. The top panel is titled 'Network' and displays dropdown menus for 'vpc-default(192.168.0.0/16)', 'subnet-default(192.168.0.0/24)', and 'Automatically-assigned IP address'. It also shows 'Available private IP addresses: 250'. Below these are buttons for 'Create VPC' and 'Add NIC', with a note 'NICs you can still add: 1'. The bottom panel is titled 'Security Group' and shows a list of security groups, with one entry 'Sys-WebServer (6b832bf5-6804-4a8b-972e-31b2d3374c...)' selected. It includes a 'Create Security Group' button and tabs for 'Security Group Rules' (selected), 'Inbound Rules', and 'Outbound Rules'. A note states: 'Similar to a firewall, a security group logically controls network access.'

- Configure networks for the ECS.
  - A subnet is a network used to manage ECSs. It provides a set of IP addresses and DNS. The IP addresses of all ECSs in a subnet belong to the subnet.
  - A security group provides access control for ECSs that have the same security protection requirements within a given VPC. It enhances security for ECSs.
  - Extension NICs are optional for configuration.

# Configuring Advanced Settings

- Set ECS Name, Login Mode, Cloud Backup and Recovery, ECS Group, and Advanced Options.

The screenshot shows the 'Create New ECS' configuration page. It includes fields for:

- ECS Name:** ecs-dc78 (highlighted with a red box)
- Login Mode:** Password (highlighted with a red box)
- Cloud Backup and Recovery:** Create new (highlighted with a red box)
- ECS Group (Optional):** Anti-affinity (highlighted with a red box)

Other visible elements include 'Allow duplicate name', 'Key pair', 'Set password later', 'Username' (root), 'Password' instructions, 'Confirm Password', 'Advanced Options' (checkbox), and 'Configure now'.

- The advanced configurations of the ECS include:
  - **ECS Name:** It can be customized but must comply with the naming rules. If multiple ECSSs are purchased at a time, the system automatically sequences these ECSSs.
  - **Login Mode:**
    - **Key pair:** You use a key pair for login authentication.
    - **Password:** You use a username and its initial password for ECS authentication. For Linux ECSSs, the initial password is the root password. For Windows ECSSs, it is the Administrator password.
  - **Cloud Backup and Recovery:** With CBR, you can back up data for EVS disks and ECSSs, and use backups to restore the EVS disks and ECSSs if something happens.
  - **ECS Group (Optional):** An ECSS group applies the anti-affinity policy to the ECSSs in it so that the ECSSs are automatically allocated to different hosts.
  - **Advanced Options** is optional.

## Access Methods

- HUAWEI CLOUD provides a web-based management platform. You can access ECSs through the management console or HTTPS-based REST APIs.

API

Management Console



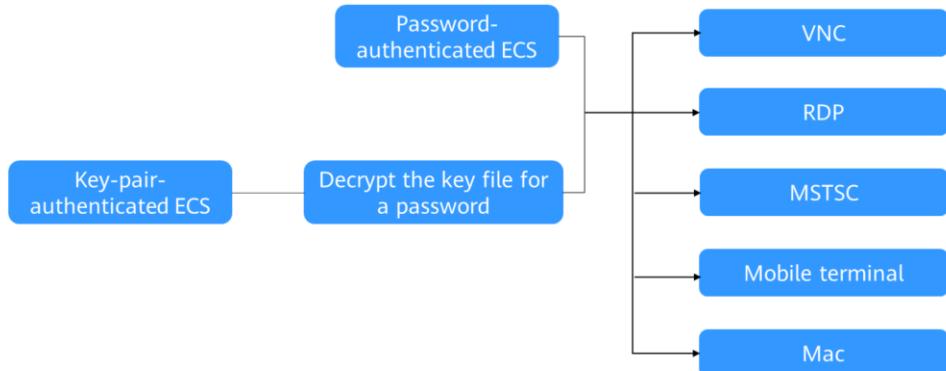
Use an API if you need to integrate the ECSs into a third-party system for secondary development.

After registering on HUAWEI CLOUD, log in to the management console and click Elastic Cloud Server under Compute on the homepage.

- You can also access your ECSs using SDKs.

## Logging In to a Windows ECS

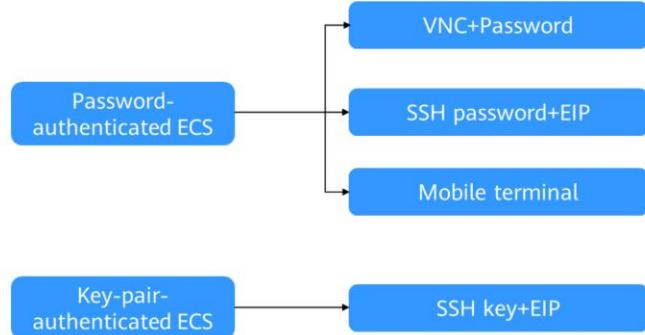
- Select a login method and log in to the ECS.



- Select a login method and log in to the Windows ECS.
  - Through the management console (VNC): The login username is Administrator.
  - Using the RDP file provided on the management console: The login username is Administrator, and the ECS must have an EIP bound.
  - Using MSTSC: The login username is Administrator, and the ECS must be bound with an EIP.
  - From a mobile terminal: The login username is Administrator, and the ECS must have an EIP bound.
  - From a Mac: The login username is Administrator, and the ECS must have an EIP bound.
- For more login methods, visit [https://support.huaweicloud.com/intl/en-us/usermanual-ecs/en-us\\_topic\\_0092494943.html](https://support.huaweicloud.com/intl/en-us/usermanual-ecs/en-us_topic_0092494943.html).

## Logging In to a Linux ECS

- The method of logging in to an ECS varies depending on the login authentication configured when you purchased the ECS.



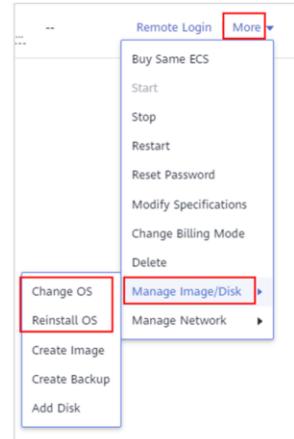
- To log in to Linux ECS using a password for the first time, you can log in as root:
  - Through the management console (VNC).
  - Using an SSH password, as long as the ECS has an EIP bound.
  - From a mobile terminal, as long as the ECS has an EIP bound.

## Reinstalling/Changing an ECS OS

- Scenarios: If the OS of an ECS fails to start, requires optimization, or cannot meet service requirements, reinstall or change the OS.

### Notes

- Only the original image of the ECS can be used to reinstall the OS.
- Changing the OS will change the system disk of the ECS. After the change, there will be a new system disk ID, and the original system disk will be gone.



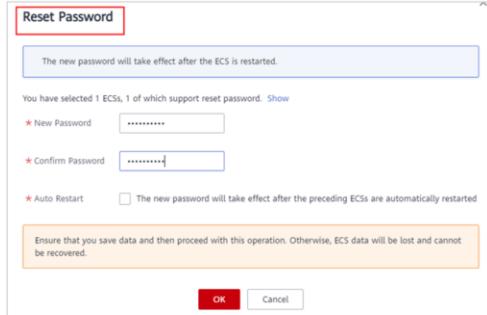
- Procedure
  - Log in to the management console.
  - Click the map icon in the upper left corner and select the desired region and project.
  - Under **Compute**, select **Elastic Cloud Server**.
  - Locate the row containing the target ECS. Click **More** in the **Operation** column and select **Manage Image/Disk > Reinstall OS**. Before reinstalling the OS, stop the ECS or select **Automatically stop the ECSs and then reinstall OSs**.
  - Configure the login mode. If the target ECS used key pair authentication, you can replace the original key pair.

## Modifying ECS Specifications

- If the specifications of an existing ECS cannot meet service requirements, modify the ECS specifications as needed, for example, by increasing the number of vCPUs or adding memory.
- Notes
  - To modify the specifications of a yearly/monthly ECS, select the target specification, pay the difference in price or claim the refund, and restart the ECS.
  - There is no need to make an additional up front payment and there are no refunds if you modify the specifications of a pay-per-use ECS.

## Resetting the ECS Login Password

- Scenarios: The ECS password is lost or has expired.
- Prerequisites: One-click password reset plug-ins have been installed on the ECS.
- Notes: ECSs created using a public image have the one-click password reset plug-in installed by default.



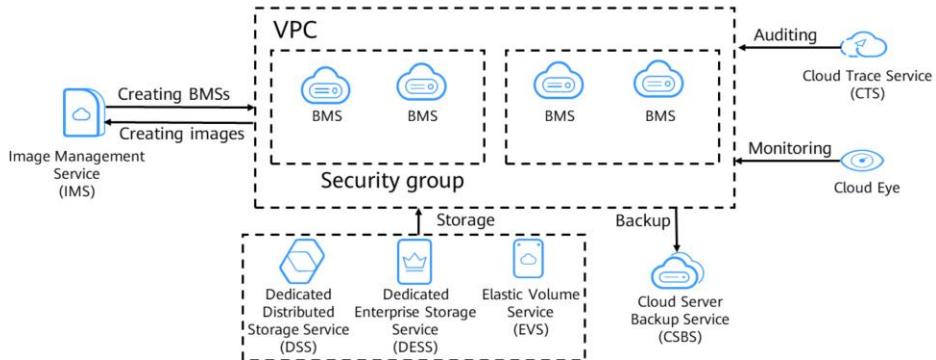
- Plug-in name: CloudResetPwdAgent and CloudResetPwdUpdateAgent.
- After installing the one-click password reset plug-ins, do not delete the CloudResetPwdAgent or CloudResetPwdUpdateAgent process, or one-click password reset will not be supported.

# Contents

1. Elastic Cloud Server (ECS)
- 2. Bare Metal Server (BMS)**
3. Image Management Service (IMS)
4. Auto Scaling (AS)
5. Cloud Container Engine (CCE)
6. Other Compute Services

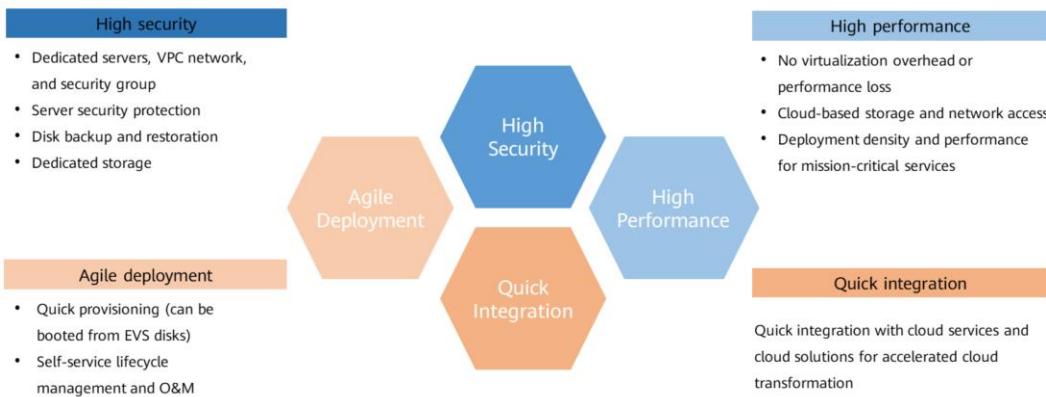
## What Is BMS?

- Bare Metal Server (BMS) combines the scalability of VMs with the high performance of physical servers. It provides dedicated servers on the cloud, delivering the performance and security required by core databases, critical applications, high-performance computing (HPC), and Big Data.



- Essentially, a BMS is a physical server. The difference is that BMSs can be easily configured and purchased on the cloud platform, but traditional physical servers can only be configured and purchased in person.
- BMSs support automatic provisioning, automatic O&M, VPC connection, and interconnection with shared storage. You can provision and use BMSs as easily as ECSs and enjoy excellent computing, storage, and network performance of physical servers.

## Why BMS?



- Advantages of BMS:

- High security: BMS allows you to use dedicated compute resources, add servers to VPCs and security groups for network isolation, and integrate related components for server security. The BMSs running on the QingTian architecture can use EVS disks, which can be backed up for restoration. BMS interconnects with Dedicated Storage Service (DSS) to ensure the data security and reliability required by enterprise services.
- High performance: BMS has no virtualization overhead, allowing compute resources to be dedicated to running services. Running on QingTian, an architecture from Huawei that is designed with hardware-software synergy in mind, BMS supports high-bandwidth, low-latency storage and networks on the cloud, meeting the deployment density and performance requirements of mission-critical services such as enterprise databases, big data, containers, HPC, and AI.
- Agile deployment: The hardware-based acceleration provided by the QingTian architecture enables EVS disks to be used as system disks. The required BMSs can be provisioned within minutes when you submit your order. You can manage your BMSs throughout their lifecycle from the management console or using open APIs with SDKs.
- Quick integration of cloud services and solutions: Within a given VPC, cloud services and cloud solutions (such as databases, big data applications, containers, HPC, and AI solutions) can be quickly integrated to run on BMSs, accelerating cloud transformation.

# Configuring Basic Settings

- Set Billing Mode, Region, AZ, CPU Architecture, and Specifications.

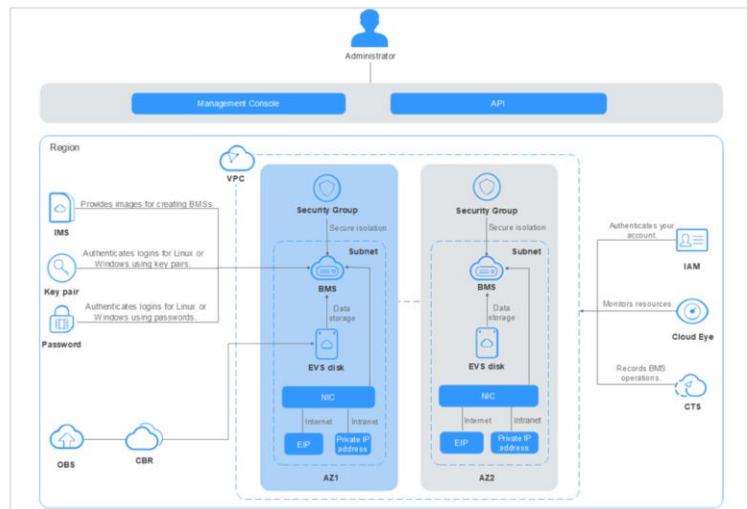
The screenshot shows two panels of the Huawei Cloud ECS configuration interface. The top panel is for 'Basic Settings' and includes tabs for 'Yearly/Monthly', 'Pay-per-use' (selected), and 'Spot price'. It shows a dropdown for 'Region' set to 'AP-Singapore' with a note: 'For low network latency and quick resource access, select the region nearest to your target users. Learn how to select a region.' Below are buttons for 'Random', 'AZ1', 'AZ2', and 'AZ3'. The bottom panel is for 'CPU Architecture' and shows tabs for 'x86' (selected) and 'Kunpeng'. It includes dropdowns for 'Latest generation', 'vCPUs', 'Memory', and a 'Flavor Name' search bar.

28      Huawei Confidential



- Configure basic settings.
  - **Billing Mode:** An ECS can be billed on a pay-per-use, yearly/monthly, or spot price basis. For yearly/monthly subscriptions, the longer the subscription, the more you save.
  - **Region and AZ:** ECSSs in different regions cannot communicate with each other over an intranet. Select a region closest to your target users to ensure low network latency and quick access.
  - **CPU Architecture:** x86-based CPUs use Complex Instruction Set Computing (CISC). Kunpeng CPUs use Reduced Instruction Set Computing (RISC).
  - **Specifications:** Select a flavor and image based on service requirements.
- Select an ECS type.
  - General computing-plus ECSSs are suitable for governments, enterprises, and the financial industry, where there are strict requirements for security and privacy; for Internet applications, which demand excellent network performance; for big data and HPC, which require a lot of vCPUs; and for website setups and e-Commerce, which need to be cost-effective.
  - Memory-optimized ECSSs are designed for memory-intensive applications that process a large amount of data, such as precision advertising, e-commerce big data analysis, and IoT big data analysis.
  - Ultra-high I/O ECSSs are designed for high-performance relational databases, NoSQL databases (such as Cassandra and MongoDB), and ElasticSearch.
  - GPU-accelerated ECSSs are suitable for applications that require real-time, highly concurrent massive computing.
  - FPGA-accelerated ECSSs are suitable for video processing, machine learning, genomics research, and financial risk analysis.
  - AI1-accelerated ECSSs are used for general technologies, such as machine vision, voice recognition, and natural language processing to support smart retail, smart campus, robot cloud brain, and safe city scenarios.

## BMS Architecture

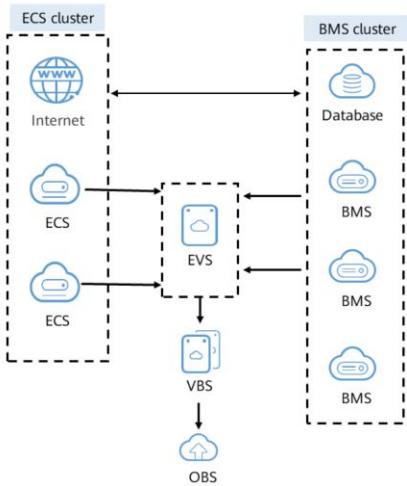


29    Huawei Confidential



- BMS works together with other cloud services to provide compute, storage, network, and imaging.
  - BMSs are deployed in multiple availability zones (AZs) connected with each other through an internal network. If an AZ becomes faulty, other AZs in the same region will not be affected.
  - With the Virtual Private Cloud (VPC) service, you can build a dedicated network for BMS, configure subnets and security groups, and allow resources deployed in the VPC to communicate with the Internet through an EIP (with bandwidth assigned).
  - With the Image Management Service (IMS), you can install OSs on BMSs or create BMSs using private images for rapid service deployment.
  - The Elastic Volume Service (EVS) provides storage, and Volume Backup Service (VBS) provides data backup and restoration.
  - Cloud Eye is a key tool to monitor BMS performance, reliability, and availability. Using Cloud Eye, you can monitor BMS resource usage in real time.
  - Cloud Backup and Recovery (CBR) backs up data for EVS disks and BMSs, and uses snapshot backups to restore the EVS disks and BMSs when necessary.

## Scenarios - Core Database



### Application Scenarios

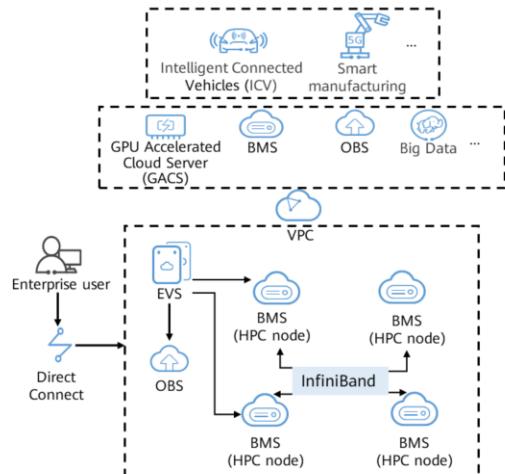
Core database. Multiple BMS flavors are available and shared EVS disks can be attached to BMSs, providing the performance and security required by core databases.

### Recommendation Reasons

- Requirements: Some critical database services cannot be deployed on VMs and must be deployed on physical servers that have dedicated resources, isolated networks, and assured performance.
- Solution: The BMS service meets these database service requirements by providing high-performance servers dedicated to individual users.

- High security and performance: Each BMS is dedicated to a single tenant and provides ultra-high computing performance without virtualization overhead. In addition, three-copy backup ensures data security and reliability.
- Quick provisioning: A BMS can be provisioned within minutes after you submit an order. The system automatically installs an OS, configures the network, and attaches disks for the BMS when receiving the order.
- Real Application Cluster (RAC): Shared EVS disks address the storage limitations faced by local disks. RAC deployment is available for core enterprise systems.
- Flexible deployment: BMSs can be deployed together with ECSs to meet diverse computing needs. They use VPC to communicate securely with other cloud resources and use EIPs to make themselves accessible from the Internet.

## Scenarios - High Performance Computing (HPC)



### Application Scenarios

Supercomputing centers and DNA sequencing. For high performance and high throughput scenarios, BMSs with the latest CPUs, coupled with a 100 Gbit/s network, provide low latency and high performance services.

### Recommendation Reasons

- Requirements:** In HPC scenarios, such as supercomputer centers and DNA sequencing, massive volumes of data need to be processed and the computing performance, stability, and real-time responsiveness need to be stellar.
- Solution:** HPC node (BMS)
  - Low latency: 100 Gbit/s, isolated, microsecond-level latency network
  - High performance: the latest Intel CPUs
  - Convenient scale-up: open APIs for easy ecosystem integration



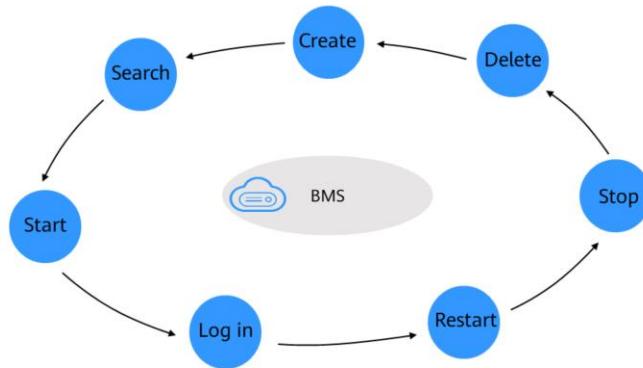
- **High-performance ECS:** Compute-intensive ECSs, such as general computing-plus (C6) and memory-optimized (M6) ECSs, use 2nd Gen Intel® Xeon® scalable processors to provide robust, stable computing performance, and Huawei-developed intelligent high-speed NICs to provide networks with ultra-high bandwidth and ultra-low latency.
- **High-performance BMS:** High-performance computing (H2) BMSs with 100 Gbit/s EDR InfiniBand NICs provide excellent computing performance with no virtualization overhead. You can apply for BMSs on demand through the management console.
- **Excellent network performance:** Secure, isolated virtual networks are provided for HPC users on the public cloud. The networks communicate with each other through intelligent high-speed NICs that deliver excellent bandwidth.

## Comparisons Between a BMS, ECS, and Physical Server

Item	BMS	ECS	Physical Server
Physical resources	Exclusive	Shared	Exclusive
Application scenarios	Mission-critical applications or services that require high performance	General-purpose and specific services	Traditional services
Provisioning	Flexible	Flexible	Inflexible
Advanced features	Automatic provisioning, automatic O&M, VPC interconnection, and interconnection with shared storage	Automatic provisioning, automatic O&M, VPC interconnection, and interconnection with shared storage	Traditional features

- A lack of flexibility is the main problem with physical servers. Although cloud computing is super popular right now, some enterprises may still choose physical servers for absolute best possible performance. The only reason is that physical servers do not have performance loss due to no virtualization overhead.
- However, it takes a long time to deploy physical servers, the O&M is complex, and the architecture cannot be reconstructed easily. When physical servers break down, it takes a lot of time, effort, and money to fix them.
- When Enterprises choose to avoid VMs (ECSs), it is typically because VMs are not able to provide the performance required by their core databases. Additionally, they do not want to adjust their core applications to adapt to VM deployment. These enterprises are faced with a dilemma.
- BMS is designed to address this dilemma. It provides physical servers exclusive to a particular enterprise's use, so they do not have to compromise on performance or resource isolation.
- Meanwhile, it delivers cloud capabilities such as online delivery, automatic O&M, VPC interconnection, and interconnection with shared storage. You can provision and use BMSs as easily as ECSs and enjoy excellent computing, storage, and network performance of physical servers.

## BMS Lifecycle Management



Self-service application, simple configuration, provisioning in minutes, and full-lifecycle management

## Creating a BMS - Basic Configuration

- Configure the region, AZ, flavor, and image.

The screenshot shows the configuration steps for creating a BMS instance:

- Region:** AP-Singapore (selected)
- AZ:** AZ1 (selected)
- Flavor:** physical.d2.large (selected)
- Image:** CentOS 7.4 64bit for BareMetal (selected)

Below the flavor table, there is a note: "BMSs in different regions cannot communicate with each other over an intranet. For low network latency and quick resource access, select the nearest region."

- The basic configuration of a BMS includes:
  - Region and AZ: BMSs in different regions cannot communicate with each other over VPC. For low network latency and quick resource access, select the region nearest to your target users.
  - Flavor/Image: Select a flavor and image based on service requirements.

## Creating a BMS - Network Configuration

- Configure the VPC, NICs, enhanced high-speed NICs, security groups, and the EIP.

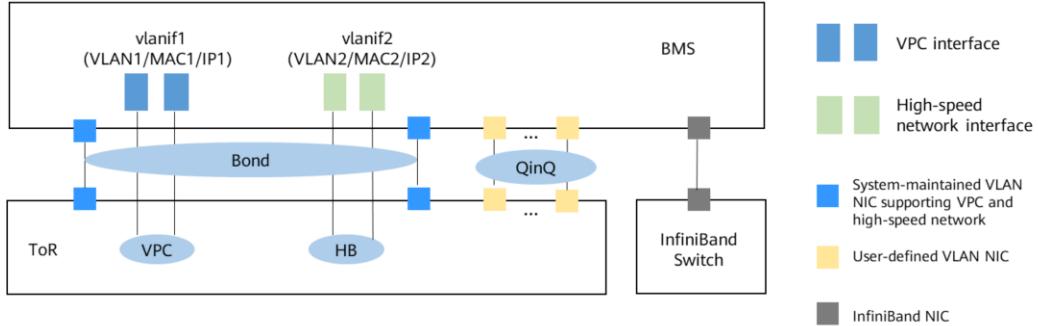
The screenshot shows the network configuration section of a BMS creation wizard. Key fields include:

- VPC:** vpc-default (selected)
- NIC:** Primary NIC: subnet-default(192.168.0.0/24); Enhanced High-Speed NIC: None selected; Security Group: default (Inbound:TCP/3389, 22 | Outbound: -); EIP: None required.
- Enhanced High-Speed NIC:** None selected.
- Security Group:** default (Inbound:TCP/3389, 22 | Outbound: -); Inbound: TCP/3389, 22; Outbound: -.
- EIP:** Not required.

- When you use VPC for the first time, the system automatically creates a VPC for you, including a security group and NIC, and enables DHCP for subnets by default.
- Security groups are used to control access to BMSs. You can define different access control rules for a security group, and these rules take effect for all BMSs added to this security group.
- When creating a BMS, you can only select a single security group, but after the BMS is created, you can associate it with additional groups.

## BMS Network

- Five types of networks are available for BMS: VPC, high-speed network, enhanced high-speed network, user-defined VLAN, and InfiniBand network. They are isolated from each other.



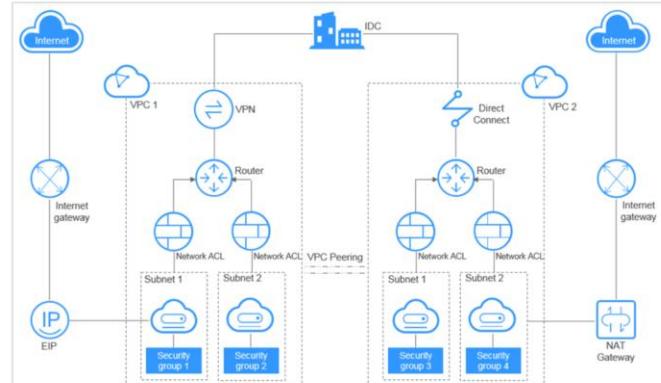
37      Huawei Confidential



- In the figure, ToR means "top-of-rack". It refers to how the server cabinet is cabled up. The access switch is placed on top of the rack and the server is placed beneath it. HB indicates a high-speed network. QinQ represents an 802.1Q tunnel.
- VPC and high-speed network interfaces are generated by the system and cannot be changed. NIC bonding is used to group multiple interfaces together.
- BMSs can communicate with ECSs through VPCs or InfiniBand networks (if any).
- Only VPC supports security groups, EIPs, and ELB.
- For a high-speed network and user-defined VLAN, BMSs in the same network can only communicate with each other through layer-2 connections.

## BMS Network - VPC

- A VPC is a logically isolated, configurable, and manageable virtual network. It helps to improve the security of BMSs in the cloud system and simplifies network deployment.



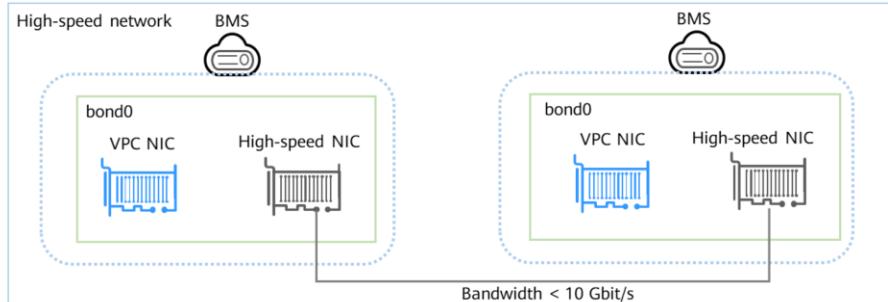
38    Huawei Confidential



- Advantages of VPC:
  - Flexible configuration: You can configure security groups, VPNs, IP address segments, and bandwidth in a VPC.
  - High security: VPCs are logically isolated from each other. By default, different VPCs cannot communicate with each other. Network ACLs protect subnets, and security groups protect ECSSs.
  - Seamless Interconnection: By default, a VPC cannot communicate with the Internet, but you can use EIP, ELB, NAT Gateway, VPN, and Direct Connect to enable access to or from the Internet. By default, two VPCs cannot communicate with each other, but you can create a VPC peering connection to enable the two VPCs in the same region to communicate with each other using private IP addresses.
  - High-speed access: More than 20 dynamic BGP connections to multiple carriers can be established. Dynamic BGP provides automatic failover in real time, automatically choosing the best alternative path when a network connection fails.

## BMS Network - High-Speed Network

- A high-speed network is an internal network between BMSs. It provides high bandwidth for connecting BMSs in the same AZ. If you want to deploy services requiring high throughput and low latency, you can create high-speed networks.
- High-speed networks share the same physical plane with VPCs. A high-speed network carries only east-west traffic and supports only layer-2 communication because it does not support layer-3 routing.



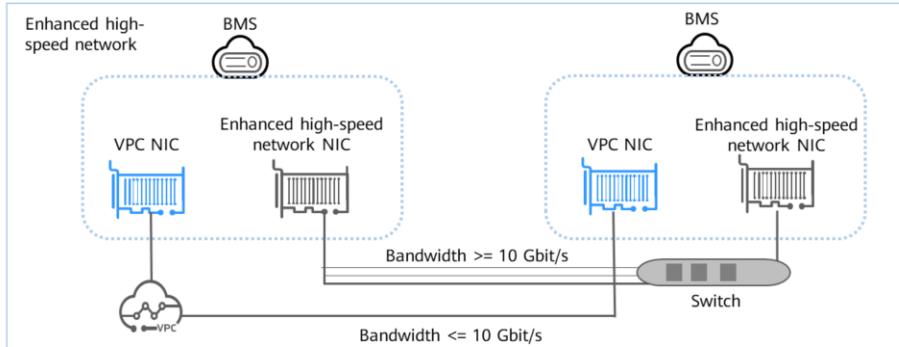
39      Huawei Confidential



- Restrictions on using high-speed networks:
  - When creating a BMS, the network segment used by standard NICs cannot overlap with that used by high-speed NICs.
  - A high-speed network does not support security groups, EIPs, DNS, VPNs, or Direct Connect connections.
  - You must select different high-speed networks for different high-speed NICs configured for a BMS.
  - Once a BMS is provisioned, you cannot then later configure a high-speed network for it.

## BMS Network - Enhanced High-Speed Network

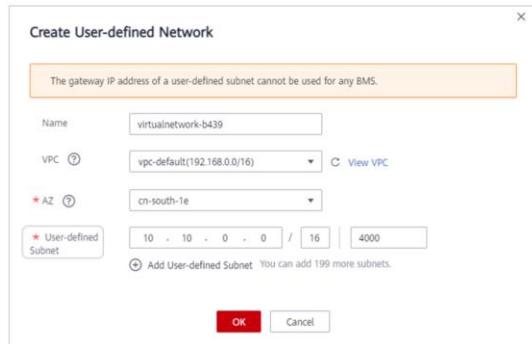
- An enhanced high-speed network is a high-quality, high-speed, low-latency internal network for BMSs to communicate with each other.



- Enhanced high-speed networks use upgraded hardware and software to allow BMSs in different PODs to communicate with each other. An enhanced high-speed network has the following advantages over a high-speed network:
  - The bandwidth is at least 10 Gbit/s.
  - The number of network planes can be customized and up to 4,000 subnets are supported.
  - VMs on a BMS can access the Internet.

## BMS Network - User-defined VLAN

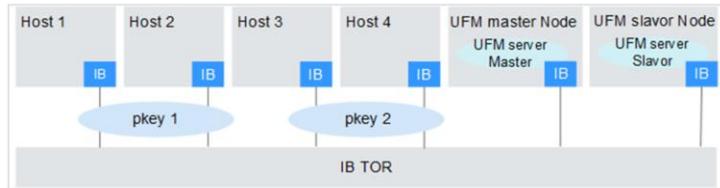
- You can allocate VLAN subnets to isolate traffic in scenarios such as SAP HANA and virtualization. User-defined VLAN NICs are in pairs. You can configure NIC bonds to achieve high availability.



- User-defined VLANs in different AZs cannot communicate with each other.

## BMS Network - InfiniBand Network

- An InfiniBand network features low latency and high bandwidth, and is good for high performance computing (HPC) projects. An InfiniBand network supports two communication modes: RDMA and IPoIB.
- To create an InfiniBand network, you must select a flavor that supports InfiniBand NICs during BMS creation.



- After an InfiniBand network is provisioned, BMSs can communicate with each other using RDMA. When IPoIB communication is used, you need to configure IP addresses on the InfiniBand network port. You can use static IP addresses or IP addresses assigned using DHCP.
- InfiniBand is widely used for communication between servers (for example, replication and distributed working), between a server and a storage device (for example, SAN and DAS), and between a server and a network (for example, LAN, WAN, and the Internet).
- InfiniBand highlights:
  - A standard protocol
  - High bandwidth, low latency
  - RDMA
  - Offloaded transmission

## Creating a BMS - Advanced Configuration

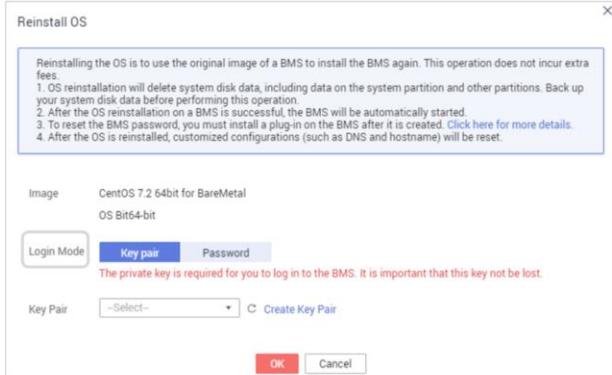
- Configure the BMS name, login mode, and advanced settings.

The screenshot shows the 'Advanced Configuration' section of a BMS setup interface. At the top, there are three tabs: 'Login Mode' (highlighted with a red box), 'Key pair' (selected), and 'Password'. A note below the tabs states: 'The private key is required for you to log in to the BMS. It is important that this key not be lost.' Under the 'Key pair' tab, there is a dropdown menu labeled '--Select--' and a button to 'Create Key Pair'. A note below the dropdown says: 'To click Remote Login to log in to a Linux BMS in key pair login mode, you must set a login password after the BMS is created. Learn how to set the password.' Below this is another tab 'Advanced Settings' with 'Do not configure' and 'Configure now' buttons. At the bottom, there is a 'BMS Name' input field containing 'bms-837d' and a note: 'If you buy more than one BMS at a time, the system automatically adds a suffix to the name of each BMS, for example, bms-0001, bms-0002...'. The entire interface is contained within a light gray box.

- You can choose to use a key pair or password for remote login authentication. For a Linux BMS, you are advised to choose key pair authentication. You can create a key pair and download the private key for remote login authentication. To ensure BMS security, a private key can be downloaded only once, so take care not to lose your downloaded private keys. You can also import the public keys of your existing key pairs to HUAWEI CLOUD, and then use the corresponding private keys to authenticate remote logins.

## Using a BMS - Reinstalling the OS

- If the OS of a BMS fails to start, gets infected by a virus, or requires optimization, reinstall the OS.



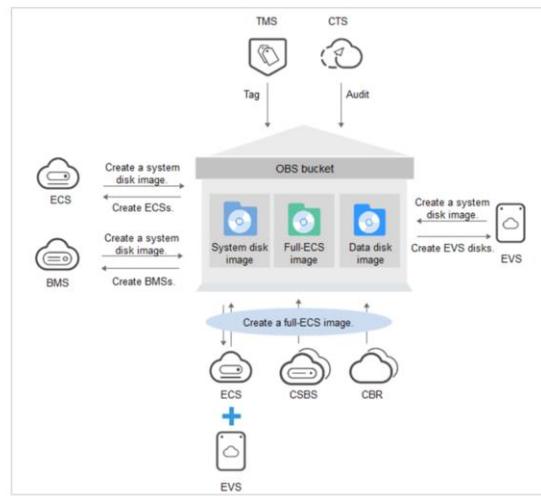
- Precautions for reinstalling a BMS OS:
  - Reinstalling the OS will interrupt services running on the BMS.
  - Reinstalling the OS destroys all of the data on all of the partitions of the system disk. Back up data before performing this operation.
  - Do not stop or restart the BMS during the reinstallation, or the reinstallation may fail.
  - After the OS is reinstalled, custom configurations that had previously existed, such as a DNS and hostname, will be gone. These customizations will have to be repeated.

# Contents

1. Elastic Cloud Server (ECS)
2. Bare Metal Server (BMS)
- 3. Image Management Service (IMS)**
4. Auto Scaling (AS)
5. Cloud Container Engine (CCE)
6. Other Compute Services

## What Is IMS?

- Image Management Service (IMS) allows you to manage the entire lifecycle of your images. You can create ECSs or BMSs from public, private, or shared images. You can also create a private image from a cloud server or an external image file to make it easier to migrate workloads to the cloud or on the cloud.

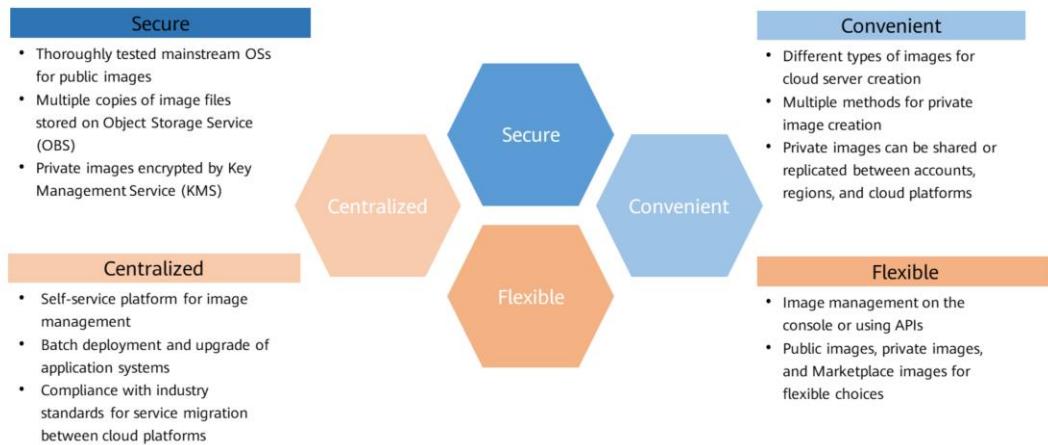


46      Huawei Confidential



- An image is a server or disk template that contains an operating system (OS), service data, and necessary application software, such as database software. IMS provides public, private, Marketplace, and shared images.

## Why IMS?



- Convenient: You can create a private from an ECS or external image file, or batch create ECSs from an image.
- Flexible: You can manage images through the management console or using APIs.
- Centralized: IMS provides a self-service platform to simplify image management and maintenance.
- Secure: Public images come with multiple mainstream OSs such as Windows Server, Ubuntu, and CentOS, which have been thoroughly tested to provide secure and stable services.

## Image Types

- A public image is a standard image provided by the cloud platform. It contains an OS and various preinstalled applications, and is available to all users.
- A private image is created by users and is visible only to the user who created it.
- A shared image is a private image another user has shared with you.
- A Marketplace image is a third-party image published in the Marketplace. It has an OS, various applications, and custom software preinstalled.



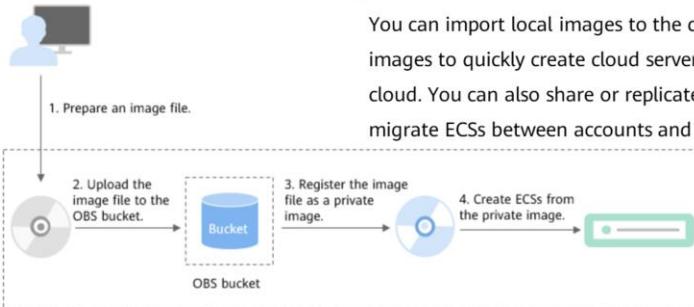
- **Public image:** A public image is a standard image provided by the cloud platform and is available to all users. It contains an OS and various preinstalled public applications. If a public image does not contain the application environment or software you need, you can use a public image to create an ECS and then install the software you need. Public images include the following OSs to choose from: Windows, CentOS, Debian, openSUSE, Fedora, Ubuntu, EulerOS, and CoreOS. When you use certain public images, the system recommends the Host Security Service (HSS) and server monitoring. HSS supports two-factor authentication for logins, defense against account cracking, and weak password detection to protect your ECSs against brute force attacks.
- **Private image:** A private image is only available to the user who created it. It contains an OS, service data, preinstalled public applications, and custom applications that the image creator added. A private image can be a system disk image, data disk image, or full-ECS image.
  - A system disk image contains an OS and pre-installed software for various services. You can use a system disk image to create ECSs and migrate your services to the cloud.
  - A data disk image contains only service data. You can use a data disk image to create EVS disks and use them to migrate your service data to the cloud.

# Scenarios - Migrating Servers to the Cloud or in the Cloud

## Application Scenarios

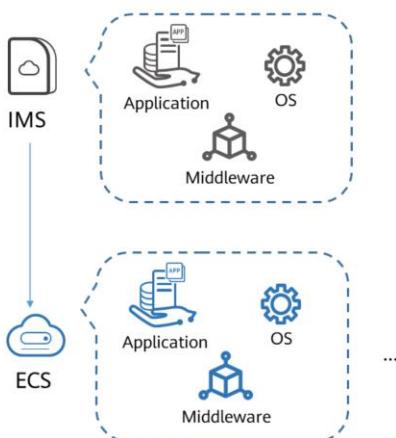
Migrating servers to the cloud or in the cloud

## Recommendation Reasons



- A variety of image formats can be imported, including VMDK, VHD, QCOW2, RAW, VHDX, QED, VDI, QCOW, ZVHD2, and ZVHD. Image files in other formats need to be converted to one of these formats before being imported. You can use the open-source tool **qemu-img** or the Huawei tool **qemu-img-hw** to convert the image.

## Scenarios - Deploying a Specific Software Environment



### Application Scenarios

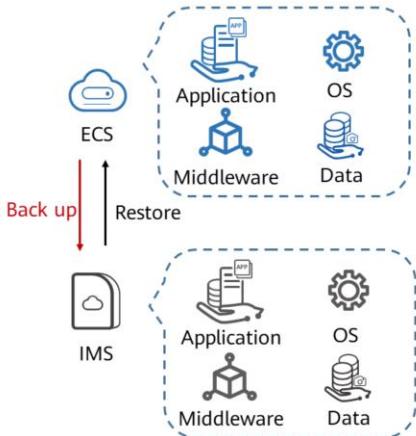
Deploying a specific software environment

### Recommendation Reasons

You can use shared or Marketplace images to quickly build custom software environments without having to manually configure environments or install any software. This is especially useful for Internet startups.

- In traditional batch service deployment, you need to evaluate different service scenarios, select an OS, database, and software, and install them. The deployment quality depends on the skills of R&D and O&M personnel.
- On the cloud platform, you can quickly create ECSs by using public, private, Marketplace, or shared images. You only need to identify sources of shared images. Public, private, and Marketplace images have been thoroughly tested to ensure security and stability.

## Scenarios - Backing Up Server Environments



### Application Scenarios

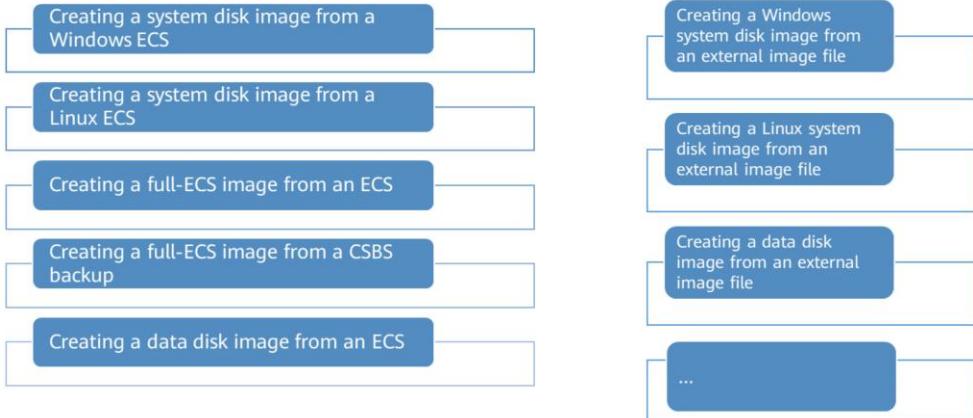
Backing up server environments

### Recommendation Reasons

You can create an image from an ECS to back up the ECS. If the ECS breaks down for some reason, you can use the image to restore it.

- This is similar to system restoration with Ghost. You can create a Ghost recovery point for your PC. If the PC is infected with a virus or the system breaks down for some reason, you can restore it to the recovery point you created.
- On the public cloud, you can create a private image to back up an ECS. If periodic backup is required, you are advised to use cloud services such as Cloud Server Backup Service (CSBS) and Volume Backup Service (VBS) for the backup.

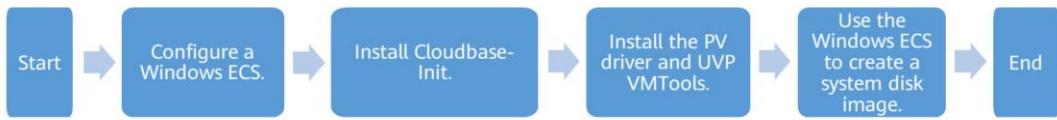
# Creating a Private Image



- You can use an ECS or external image file to create an ECS private image.
- You can also:
  - Use an ISO file to create an ECS system disk image.
  - Use a CBR backup to create a full-ECS image.
  - Use a BMS to create a system disk image.

# Creating a System Disk Image from a Windows ECS

This course will show how to create a system disk image from a Windows ECS as an example.



## Configuring a Windows ECS

- Prepare a Windows ECS and check whether the ECS NIC is configured to use DHCP.



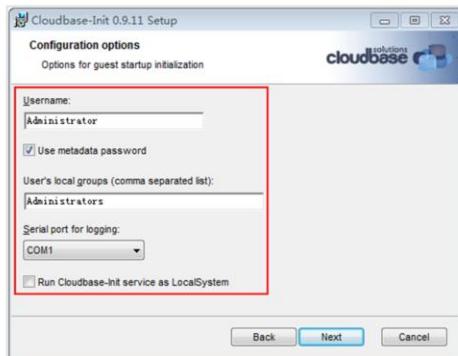
55      Huawei Confidential



- If a Windows ECS is using a static IP address, you will have to log in to the ECS and change the network settings to use DHCP. The procedure is as follows:
  - Log in to the Windows ECS. Choose **Start > Control Panel > Network and Internet > Network and Sharing Center** > *[Connection with the static IP address]* > **Properties** > **Internet Protocol Version 4 (TCP/IPv4)** > **Properties** > **General**.
  - On the **General** tab page, select **Obtain an IP address automatically** and **Obtain DNS server address automatically**, and click **OK**.

## Installing Cloudbase-Init

- To ensure that ECSs created from a private image are configurable, you are advised to install Cloudbase-Init on the ECS before using it to create a private image.



56      Huawei Confidential



- Cloud-Init/Cloudbase-Init is a cloud initialization program, which initializes specific configurations, such as the host name, key, and user data, of a newly created ECS.
  - For Windows, download and install Cloudbase-Init.
  - For Linux, download and install Cloud-Init.
- Before you download Cloud-Init/Cloudbase-Init from the official website to the ECS, bind an EIP to the ECS so that the ECS can connect to the Internet.

## Installing the PV Driver and UVP VMTools

- To ensure that ECSs created from a private image support both Xen and KVM virtualization, install the PV driver and UVP VMTools on the ECS before using it to create a private image.



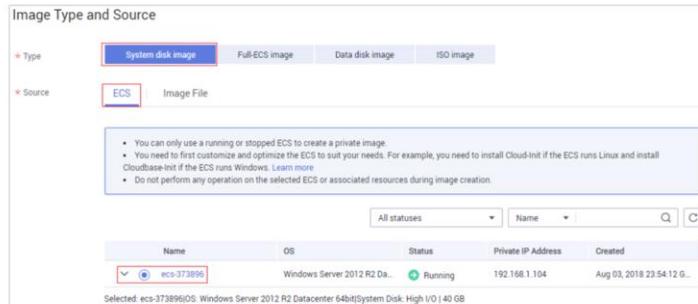
57      Huawei Confidential



- Xen and KVM are open-source virtualization technologies. They require, respectively, the PV driver and UVP VMTools.

## Using a Windows ECS to Create a System Disk Image

- On the Image Management Service page, click Create Image.
- In the Image Type and Source area, select System disk image for Type.
- By default, ECS is selected for Source. Select an ECS from the list.



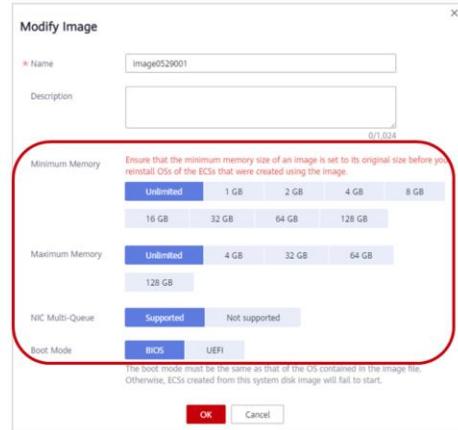
58      Huawei Confidential



- You do not need to stop an ECS when you use it to create a private image. However, to enhance data consistency in the production environment, stopping the ECS is recommended.

## Image Management - Modifying Image Information

- You can modify the image name, description, minimum and maximum memory, NIC multi-queue, and SR-IOV driver.



- Only private images that are in the **Normal** state can be modified.
- NIC multi-queue enables multiple CPUs to process NIC interruptions for load balancing.
- After the SR-IOV driver is installed for an image, the network performance of ECSs created from the image will be greatly improved.

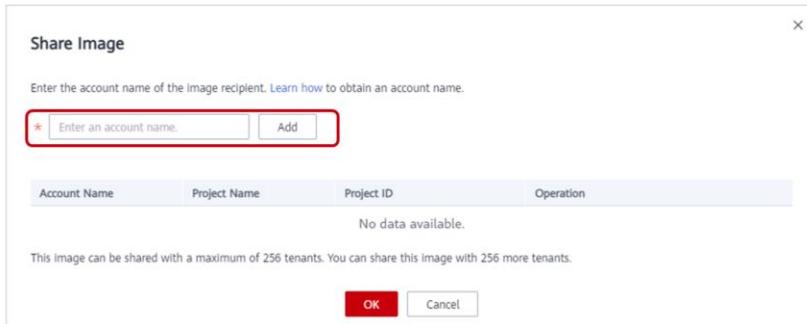
# Image Management - Deleting an Image

- Note that:
  - Deleted private images cannot be retrieved. Perform this operation only when absolutely necessary.
  - After a private image is deleted, it cannot be used to create cloud servers or EVS disks.
  - After a private image is deleted, ECSs created from the image can still be used and are still billed. However, the OS cannot be reinstalled for the ECSs and an ECS with the same configuration cannot be recreated.
- Deleting the source image of a replicated image has no effect on the replicated image. Similarly, deleting a replicated image has no effect on its source.

The screenshot shows the 'Image Management Service' interface. At the top, there are tabs for 'Public Images' (disabled), 'Private Images' (selected), and 'Images Shared with Me'. Below the tabs, a message says 'IMI is now commercially available. Any private image stored will be billed according to IMI pricing.' There is a note about optimizing private images for ECS creation. The main area displays a table of images. One row is selected, showing details: Name: Image020001, ID: 1, Status: Normal, OS Type: Linux, OS: CentOS 7.6 (64bit). The 'Image Type' column shows 'ECS system disk image(s)'. The 'Disk Capacity (GB)' is 40, and 'Encrypted' is No. The 'Created' date is Jul 27, 2021 10:53:13 GMT+0800. On the right, there are buttons for 'Apply for Server', 'Modify', and a context menu with options: Delete (highlighted), Share, Export, and Replicate.

## Image Management - Sharing an Image

- You can share your private images.



61      Huawei Confidential



- You can share images, stop sharing images, and add or delete tenants that can use the shared images.
- The recipient can choose to accept or reject the shared images, or remove images they have previously accepted.

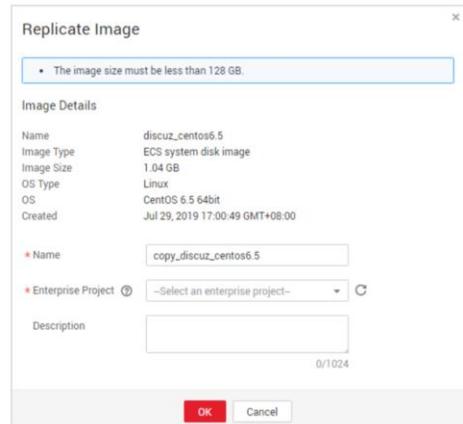
## Image Management - Encrypting an Image

- You can create an encrypted image to securely store data.
- Encrypted images cannot be shared with other users or published in the Marketplace.
- The system disk of an ECS created from an encrypted image is also encrypted, and its key is the same as the image key.
- If an ECS has an encrypted system disk, private images created from the ECS are also encrypted.

The screenshot shows the 'Image Information' configuration page. It includes sections for 'Function' (set to 'ECS system disk image'), 'Architecture' (set to 'X86'), 'Boot Mode' (set to 'UEFI'), 'OS' (with dropdowns for 'Select OS...' and 'Select OS version...'), 'System Disk (GB)' (set to 40-1,024), 'Name' (input field), and 'Encryption' (checkbox). The 'Encryption' checkbox is highlighted with a red border.

## Image Management - Replicating an Image Within a Region

- You may need to replicate an image in the following scenarios:
- Creating an unencrypted version of an encrypted image
- Replicating an encrypted image
- Creating an encrypted version of an unencrypted image



- You may need to replicate an image in the following scenarios:

- Creating an unencrypted version of an encrypted image

Encrypted images cannot be shared with other users or published in the Marketplace. If you want to publish or share an encrypted image, you need to create an unencrypted version.

- Replicating an encrypted image

The key used for encrypting an image cannot be changed directly. If you want to change the key of an encrypted image, you can replicate this image and encrypt the new image using a different key.

- Creating an encrypted version of an unencrypted image

If you want to encrypt an unencrypted image, you can replicate the image and encrypt the new image using a key.

## Image Management - Replicating an Image Across Regions

- You can replicate an image from one region to another and use the replicated image to create identical ECSs. This allows you to more quickly migrate services across regions.

The dialog box displays the following information:

Replicate Image	
Image Size	9.7 GB
OS Type	Windows
OS	Windows-Server-2012-R2-Datacenter-64bit
Created	Dec 02, 2020 11:22:41 GMT+08:00
Replication Mode	<input checked="" type="radio"/> Within Region <input type="radio"/> Across Regions
★ Name	copy_lr-iaas-odin1_1202-new-2012-g5r
★ Destination Region	—Select—
★ Destination Project	—Select—
★ IAM Agency	—Select— <small>C View Agency ?</small>
<input type="button"/> OK <input type="button"/> Cancel	

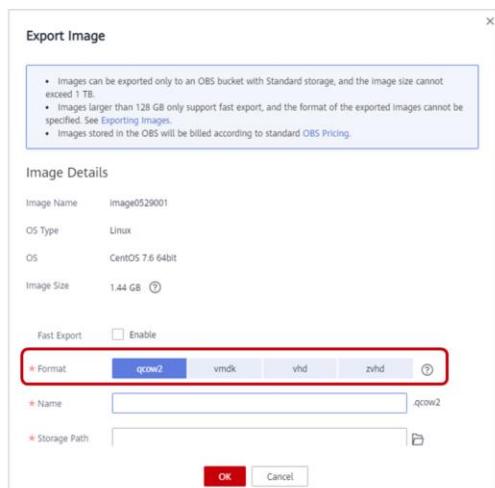
64      Huawei Confidential



- Cross-region image replication is required for system HA typically when your system is deployed in multiple regions. In most cases, ECSs are deployed in multiple regions. If you want to clone an ECS across regions, you can replicate its image across the regions and then use the image to create the identical ECSs in the target region.
- Cross-region replication is applicable to cross-region server deployment or data backup. It is often used together with image sharing for cross-region, cross-account image replication.
- You can select multiple images for cross-region replication at once. However, you are not allowed to select an ISO image, encrypted image, full-ECS image, frozen image, or an image that is currently being created.

## Image Management - Exporting an Image

- You can export an image if you want to:
  - Store the image on specified storage devices.
  - Use the image to create servers on other cloud platforms.



 HUAWEI

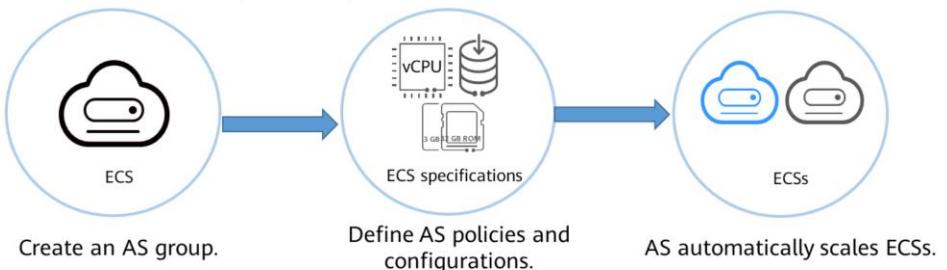
- You can export private images to OBS buckets in a specified format and then download the images from the buckets to specified storage devices. Images exported in different formats may vary in size. You will be charged for the OBS storage occupied by exported images.

# Contents

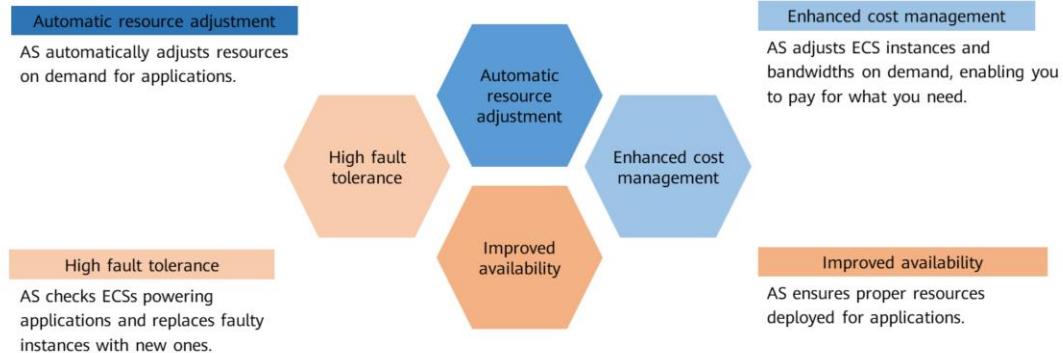
1. Elastic Cloud Server (ECS)
2. Bare Metal Server (BMS)
3. Image Management Service (IMS)
- 4. Auto Scaling (AS)**
5. Cloud Container Engine (CCE)
6. Other Compute Services

## What Is AS?

- Auto Scaling (AS) automatically adjusts resources to keep up with changes in demand based on pre-configured AS policies. You can specify AS configurations and policies based on service requirements. These configurations and policies free you from having to repeatedly adjust resources to keep up with service changes and spikes in demand, helping you reduce the resources and manpower required.



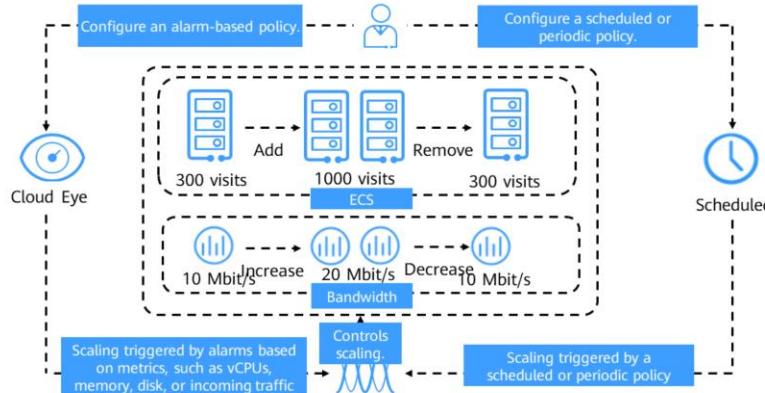
## Why AS?



- AS advantages:
  - Automatic resource adjustment: AS adds ECS instances and increases bandwidth for your applications when the access volume increases and reduces unneeded resources when the access volume drops, ensuring system stability.
  - Enhanced cost management: AS enables you to use instances and bandwidth on demand by automatically adjusting system resources, so utilization goes up and costs go down.
  - Improved availability: AS ensures there are always enough resources deployed for your applications. When working with ELB, AS automatically associates a load balancing listener with any instances newly added to the AS group. Then, ELB automatically distributes access traffic to all instances in the AS group through the listener, which improves system availability.
  - High fault tolerance: AS monitors the status of instances in an AS group, and replaces any unhealthy instances it detects.

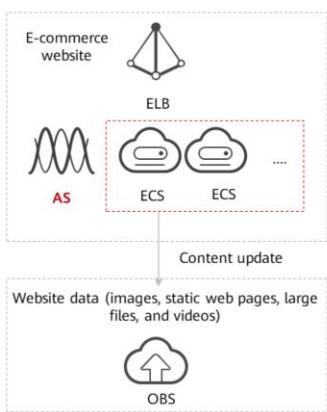
## AS Architecture

- AS automatically adjusts compute resources based on service demands and configured AS policies. The number of ECS instances changes to match service demands, ensuring service availability.



- AS allows you to adjust the number of ECSSs in an AS group and EIP bandwidths bound to the ECSSs.
  - Scaling control: You can specify thresholds and schedule when different scaling actions are taken. AS will trigger scaling actions on a repeating schedule, at a specific time, or when configured thresholds are reached.
  - Policy configuration: You can configure alarm-based, scheduled, and periodic policies as needed.
  - Alarm-based: You can configure alarm metrics such as vCPU, memory, disk, and inbound traffic.
  - Scheduled: You can schedule actions to be taken at a specific time.
  - Periodic: You can configure scaling actions to be taken at scheduled intervals, a specific time, or within a particular time range.
  - When Cloud Eye generates an alarm for a monitoring metric, for example, CPU usage, AS automatically increases or decreases the number of instances in the AS group or the EIP bandwidth.
  - When the configured triggering time arrives, a scaling action is triggered to increase or decrease the number of ECS instances or the bandwidth.

## Scenarios – Web Applications



### Application Scenarios

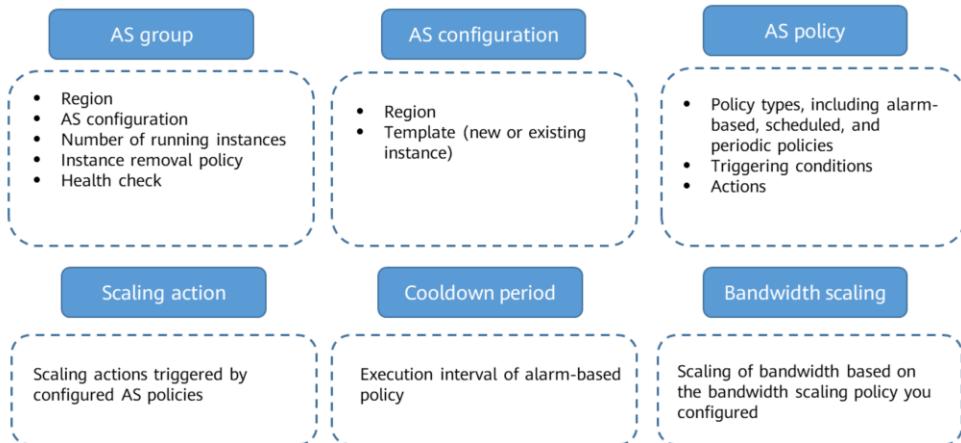
- E-commerce websites
- Heavy-traffic web portals

### Recommendation Reasons

- E-commerce: During big promotions, E-commerce websites need more resources. AS automatically scales out ECS instances and bandwidth within minutes to ensure that promotions go smoothly.
- Heavy-traffic portals: Service load changes are difficult to predict for heavy-traffic web portals. AS dynamically scales in or out ECS instances based on monitored ECS metrics, such as vCPU usage and memory usage.

- Using ELB with AS
  - Working with ELB, AS automatically increases or decreases resources based on changes in demand while ensuring that the load of all the ECS instances in the AS group stays balanced.
  - After ELB is enabled in an AS group, AS automatically associates a load balancing listener with instances newly added to the AS group. Then, ELB automatically distributes access traffic to all instances in the AS group through the listener, which improves system availability. If the instances in the AS group are running a range of different types of applications, you can bind multiple load balancing listeners to the AS group to listen to each of these applications, improving scalability.

## AS Basic Concepts



71 Huawei Confidential



- AS Basic Concepts
  - AS group: An AS group consists of a collection of instances **and AS policies** that have similar attributes and apply to the same scenario. It is the basis for enabling or disabling AS policies and performing scaling actions.
  - AS configuration: An AS configuration is a template specifying specifications for the instances to be added to an AS group. The specifications include the ECS type, vCPUs, memory, image, disk, and login mode.
  - AS policy: An AS policy can trigger scaling actions to adjust the number of instances in an AS group. An AS policy defines the condition to trigger a scaling action and the operations to be performed. When the triggering condition is met, the system automatically triggers a scaling action.
  - Scaling action: A scaling action adds instances to or removes instances from an AS group. It ensures that the number of instances in an application system is the same as the expected number of instances by adding or removing instances when the triggering condition is met, which improves system stability.
  - Cooldown period: To prevent an alarm policy from being repeatedly triggered for the same event, we use a cooldown period. The cooldown period specifies how long any alarm-triggered scaling action will be disallowed after a previous scaling action is complete. The cooldown period is not used for scheduled or periodic scaling actions.
  - Bandwidth scaling: AS automatically adjusts a bandwidth based on the configured bandwidth scaling policy. AS can only adjust the bandwidth of pay-per-use EIPs and shared bandwidths. It cannot adjust the bandwidth of

yearly/monthly EIPs.

## Getting Started with AS



## Creating an AS Configuration

- Configuration Template options

### Create a specifications template

If you have special requirements on the specifications of the ECSs used for capacity expansion, specify the specifications in a template and use it to create an AS configuration. Then, the specifications will be applied to the ECSs added to the AS group in scaling actions.

### Use specifications of an existing ECS

You can use an existing ECS to quickly create an AS configuration. Then, the specifications of this ECS, such as the vCPUs, memory, image, disk, and ECS type, will be applied to ECSs added to the AS group in scaling actions.

## Creating an AS Group

- An AS group consists of a collection of instances and AS policies that have similar attributes and apply to the same scenario. It is the basis for enabling or disabling AS policies and performing scaling actions.
- AS automatically scales in or out instances or maintains a fixed number of instances in an AS group through scaling actions triggered by configured AS policies.
- When creating an AS group, you need to configure parameters, such as Max. Instances, Min. Instances, Expected Instances, and Load Balancing.

The screenshot shows the 'Create AS Group' page. At the top, there's a policy selection section with 'Multi-AZ Extension Policy' (selected), 'Load-balanced' (radio button), and 'Sequenced' (radio button). Below this, the 'Name' field is set to 'as-group-b29f'. The 'Max. Instances' field is highlighted with a red box and contains the value '1'. The 'Expected Instances' field is also highlighted with a red box and contains the value '0'. The 'Min. Instances' field contains the value '0'. In the 'Advanced Policies' section (labeled '2'), there are three main categories: EIP, Data Disk, and Health Check Method. The 'EIP' section has a dropdown menu showing 'Oldest instance created from oldest AS config...' with 'Release' (selected) and 'Do not release' options. The 'Data Disk' section has a similar dropdown for 'Release' (selected) and 'Do not release'. The 'Health Check Method' section has a dropdown menu showing 'ECS health check' with 'Health Check Interval' set to '5 minutes'.

74      Huawei Confidential



- Main parameters for creating an AS group
  - Multi-AZ Expansion Policy:** This parameter is required only when two or more AZs are selected.
  - Max./Min. Instances:** Specifies the minimum or maximum number of ECS instances in an AS group.
  - Expected Instances:** Specifies the number of ECSs that are expected to run in an AS group. It is between the minimum and maximum numbers of instances. Generally, when the service peak is about to arrive, **Expected Instances** enables you to quickly provision a large number of ECS instances.
  - Instance Removal Policy:** When instances are automatically removed from your AS group, the instances that are not in the currently used AZs will be removed first. Additionally, AS will check whether instances are evenly distributed in the currently used AZs. If the load among AZs is unbalanced, AS balances the load among AZs when removing instances. If the load among AZs is balanced, AS removes instances following the instance removal policy you configured here.

## Creating an AS Policy

- Main parameters: Policy Type and Cooldown Period

The screenshot shows the 'Add AS Policy' configuration page. Key fields include:

- Policy Name:** as-policy-8da5
- Policy Type:** Alarm (highlighted with a red box)
- Rule Name:** as-alarm-8db0
- Monitoring Type:** System monitoring (highlighted with a red box)
- Trigger Condition:** CPU Usage > Max (highlighted with a red box)
- Monitoring Interval:** 5 minutes (highlighted with a red box)

A note at the bottom states: "To determine if an OS supports metrics Memory Usage, Inband Outgoing Rate, and Inband Incoming Rate, see Elastic Cloud Server User Guide. Before using Agent to monitor metrics, make sure that the Agent plug-in has been installed on all instances in the AS group. Learn how to install the Agent plug-in."

75      Huawei Confidential



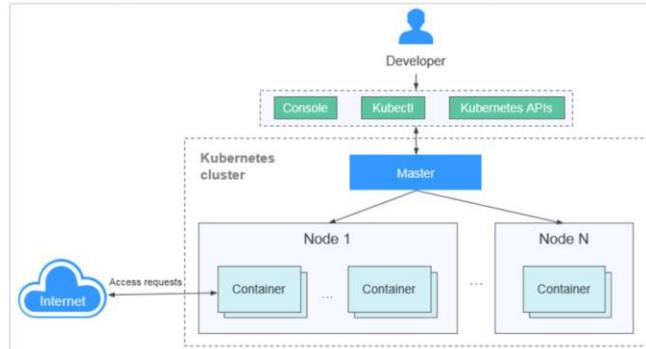
- If the service workloads are unpredictable, you can configure alarm-based AS policies. These policies are used to trigger scaling actions based on real-time monitoring data (such as CPU usage) to dynamically adjust the number of instances in the AS group. AS restarts the cooldown period after a scaling action is complete. During the cooldown period, scaling actions triggered by alarms will be denied. Scheduled and periodic scaling actions are not affected.

# Contents

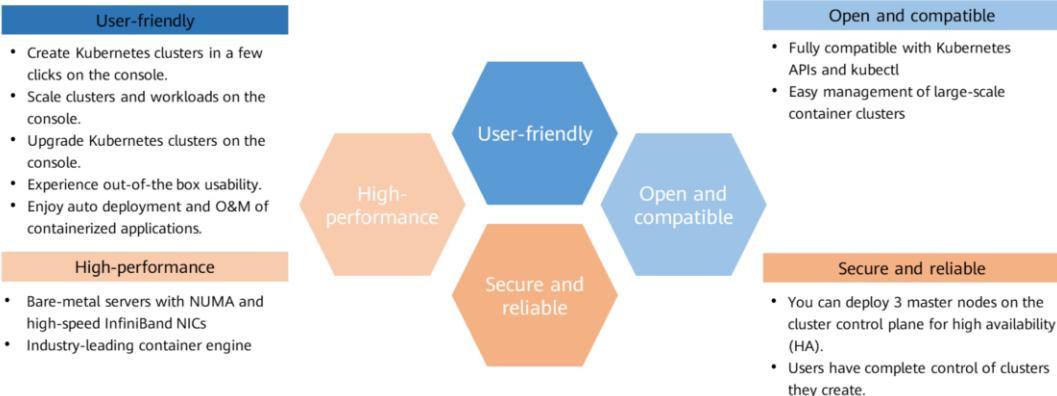
1. Elastic Cloud Server (ECS)
2. Bare Metal Server (BMS)
3. Image Management Service
4. Auto Scaling (AS)
- 5. Cloud Container Engine (CCE)**
6. Other Compute Services

## What Is CCE?

- Cloud Container Engine (CCE) is a highly scalable, high-performance, enterprise-class Kubernetes service for you to run containers and applications. With CCE, you can easily deploy, manage, and scale containerized applications on HUAWEI CLOUD.



## Why CCE?

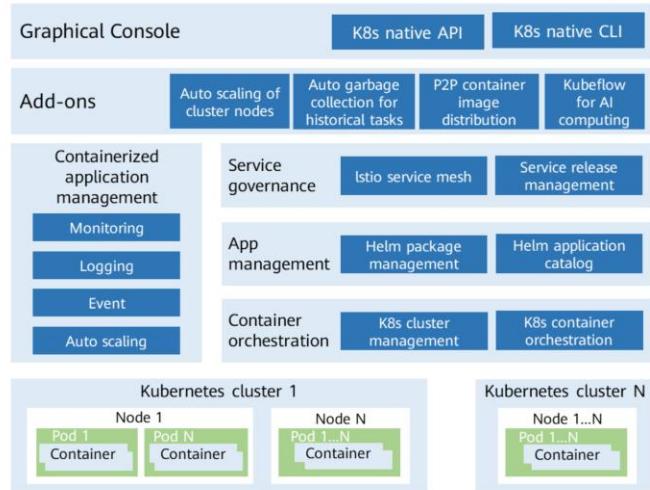


78      Huawei Confidential



- User-friendly:
  - Creating a Kubernetes cluster is as easy as a few clicks on the console. You can create either VM nodes or bare-metal nodes, or both, in a cluster.
  - From auto deployment to O&M, you can manage your containerized applications all in one place throughout their lifecycle.
  - You can also scale your clusters and workloads in just a few clicks on the console. Auto scaling policies can be flexibly combined to deal with in-the-moment load spikes.
  - The console enables you to easily upgrade your clusters.
  - Application Service Mesh (ASM) and Helm charts are pre-integrated, delivering out-of-the-box usability.
- High-performance:
  - CCE draws on Huawei's years of field experience in computing, network, storage, and heterogeneous infrastructure. You can concurrently launch containers at scale.
  - The bare-metal NUMA architecture and high-speed InfiniBand network cards yield a three- to five-fold improvement in computing performance.

## CCE Architecture

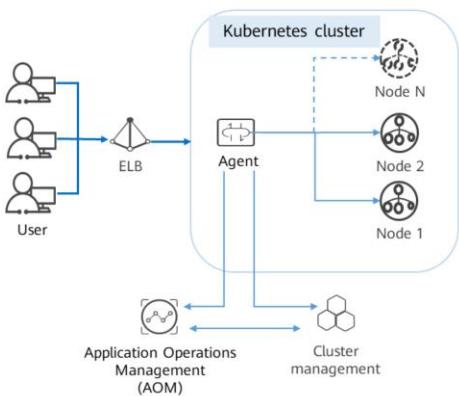


80 Huawei Confidential



- CCE is deeply integrated with high-performance HUAWEI CLOUD services, including compute (ECS/BMS), network (VPC/EIP/ELB), and storage (EVS/OBS/SFS) services. It supports heterogeneous computing architectures such as GPU, NPU, and Arm. By using multi-AZ and multi-region disaster recovery, CCE ensures high availability of Kubernetes clusters.

## Scenario - Auto Cluster Scaling



### Function Description

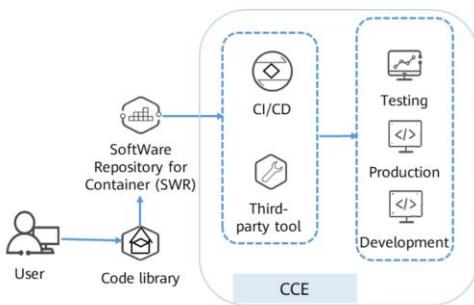
CCE adjusts compute resources based on auto scaling policies to handle fluctuating service loads. Specifically, CCE automatically adds or reduces cloud servers for your cluster or containers for your workload.

### Benefits

- Flexible: Multiple scaling policies are supported and containers can be provisioned within seconds when specific conditions are met.
- Highly available: Pods are automatically monitored and unhealthy pods will be replaced with new ones to ensure high service availability.
- Low cost: You are billed only for the cloud servers you use.

- Application scenarios:
  - Traffic surges brought by promotions and flash sales on online shopping apps and websites
  - Fluctuating service loads of live streaming that require real-time scaling based on CPU or memory usage
  - Increase in the number of game players that go online in certain time periods

## Scenario - DevOps



### Function Description

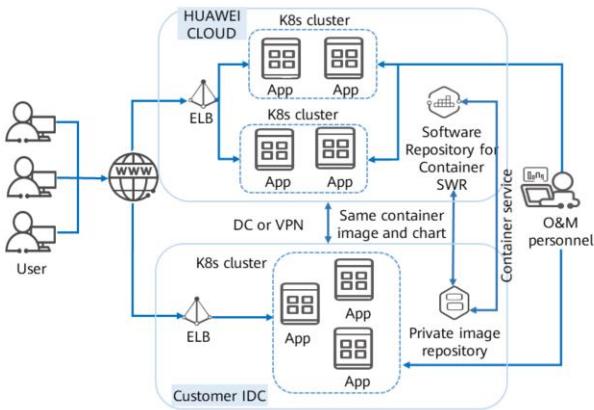
CCE automatically completes code compilation, image build, grayscale release, and container-based deployment based on code sources. CCE can interconnect with your CI/CD systems. You can containerize traditional applications and deploy them in the cloud.

### Benefits

- **Efficient CI/CD management:** Reduces scripting workload by more than 80% through streamlined process interaction.
- **Flexible integration:** Provides various APIs to integrate with existing CI/CD systems, facilitating customization.
- **High performance:** Allows for flexible task scheduling with a fully containerized architecture.

- Development and Operations (DevOps) is a set of processes, approaches, and systems for collaboration between software development, O&M, and quality assurance (QA) teams.
- Scenario description: Your applications and services may receive a lot of feedback and requirements. To release new features and improve user experience, you need fast continuous integration (CI). An efficient tool to support CI is container. By deploying containers, you can streamline the development, testing, and release process, realizing continuous delivery (CD).
- Continuous integration (CI), continuous delivery (CD), and continuous deployment

## Scenario - Hybrid Cloud



### Function Description

Environment-independent containers allow you to seamlessly migrate applications and data between private and public clouds. You can achieve efficient resource usage and realize disaster recovery (DR).

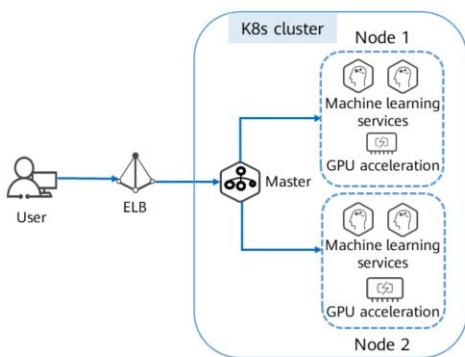
### Benefits

- Lower costs:** Resource pools on HUAWEI CLOUD support rapid service scaling during peak hours, for only a fraction of the cost involved in building private clouds from scratch.
- On-cloud DR:** Your services can be deployed both on-premises and in the cloud. The on-premises system provides services while the cloud system ensures DR.
- Shared base:** The on-premises system shares the technical base with the cloud system. HUAWEI CLOUD resources are available on the on-premises system whenever required.



- Detailed description of application scenarios:
  - Multi-cloud deployment for DR and backup:** To achieve high service availability, you can deploy applications on container services from multiple cloud providers. When a cloud is down, application loads will be automatically distributed to other clouds.
  - Load balancing and auto scaling:** Large enterprise systems often span cloud facilities in different regions. They also need to be automatically resizable — they can start small and then scale up as system load grows. This frees enterprises from the costs of planning, purchasing, and maintaining more cloud facilities than needed and transforms large fixed costs into much smaller variable costs.
  - Migration to clouds and database hosting:** Finance, security, and other industries whose top concern is data confidentiality want to keep critical systems in local IDCs while moving other systems to the cloud. They want to manage these systems, no matter in local IDCs or in the cloud, in a unified manner.
  - Decoupling development and deployment:** To ensure IP security, you can set up the production environment on a public cloud and the development environment in your local IDC.

## Scenario - AI Computing



### Use Case

AI computing

### Benefits

- **Outstanding performance:** The bare-metal NUMA architecture and high-speed InfiniBand NICs drive a three- to five-fold improvement in AI computing performance.
- **Efficient computing:** GPUs are shared and scheduled among multiple containers, greatly reducing computing costs.
- **Proven success:** AI containers are compatible with all mainstream GPU models and have been used at scale in HUAWEI CLOUD's Enterprise Intelligence (EI) products.

- By integrating Volcano, CCE has the following advantages in running high-performance computing, big data, and AI jobs:
  - **Hybrid deployment:** HPC, big data, and AI jobs can be run together.
  - **Optimized multi-queue scheduling:** Multiple queues can be used for multi-tenant resource sharing and group planning based on priorities and time periods.
  - **Advanced scheduling policies:** Gang scheduling, fair scheduling, resource preemption, and GPU topology are supported.
  - **Multi-task template:** You can use a template to define multiple tasks in a single Volcano Job, beyond the limit of Kubernetes native resources. Volcano Jobs can describe multiple job types, such as TensorFlow, MPI, and PyTorch.
  - **Job extension plugins:** The Volcano Controller allows you to configure plugins to customize environment preparation and cleanup in stages such as job submission and pod creation. For example, before submitting a common MPI job, you can configure the SSH plugin to provide the SSH information of pod resources.

## CCE Concepts

Cluster	A cluster is a collection of cloud resources required for running containers, such as cloud servers and load balancers.
Pod	A pod consists of one or more related containers that share the same storage and network space.
Node	A node is a server (a VM or PM) on which containerized applications run.
Service	A Service is an abstraction which defines a logical set of pods and a policy by which to access them (sometimes this pattern is called a microservice).
Container	A container is a running instance of a Docker image. Multiple containers can run on the same node.
Image	An image is a binary that includes all of the requirements for running a container.

- CCE concepts:

- A cluster is a combination of cloud resources required for container running, such as cloud servers and load balancers. In a cluster, one or more elastic cloud servers (ECSs, also called nodes) are deployed in the same subnet to provide compute resources for container running.
- Pods are the smallest and most basic deployable objects in Kubernetes. A pod encapsulates an application container (or, in some cases, multiple containers), storage resources, a unique network IP address, and options that govern how the containers should run.
- A node is a server (a VM or PM) on which containerized applications run. The node agent (kubelet) runs on each node to manage containers on the nodes. The number of nodes in a cluster can be scaled.
- A Service is an abstract method that exposes a group of applications running on a pod as network services.
- A container is a running instance of a Docker image. Multiple containers can run on the same node. Containers are basically software processes but have separate namespaces and do not run directly on a host.
- Docker creates an industry standard for packaging containerized applications. Docker images are like templates that include everything needed to run containers, and are used to create Docker containers. In other words, a Docker image is a special file system that includes the required programs, libraries, resources, and configuration files to make an application run. It also contains parameters you can configure for your application, such as anonymous volumes, environment variables, and users.

- For details, see [https://support.huaweicloud.com/intl/en-us/productdesc-cce/cce\\_productdesc\\_0011.html](https://support.huaweicloud.com/intl/en-us/productdesc-cce/cce_productdesc_0011.html).

## CCE Configuration Process



- Register a HUAWEI CLOUD account and log in to the CCE console.
- Select a cluster type and create a cluster.
- Deploy a workload (application) using an existing or newly created image or orchestration template.

## Creating a Cluster

- When creating a CCE cluster, set the billing mode, region, cluster version, management scale, and number of master nodes.

The screenshot shows the configuration interface for creating a CCE cluster. It includes fields for Billing Mode (Yearly/Monthly, Pay-per-use), Region (AP-Singapore), Cluster Name (cce-vivi), Version (v1.17.17 selected), Management Scale (50 nodes selected), and Number of master nodes (3 selected). A note at the bottom indicates that 2,000 nodes are sold out. A red box highlights the Management Scale and Number of master nodes sections.

Billing Mode: Yearly/Monthly, Pay-per-use

Region: AP-Singapore

Cluster Name: cce-vivi

Version: v1.17.17

Management Scale: 50 nodes

Number of master nodes: 3

## Scaling a Cluster

- CCE automatically scales a cluster (adding or releasing worker nodes) according to the scaling policies you configure. For example, when workloads cannot be scheduled into the cluster due to insufficient cluster resources, scale-out will be automatically triggered.

Events | Auto Scaling | Kubectl | Resource Tags | Istioctl

**Scale-out Settings** Scale-out Policies

**Edit**

Maximum Nodes: 10      Cooldown Period (s): 900

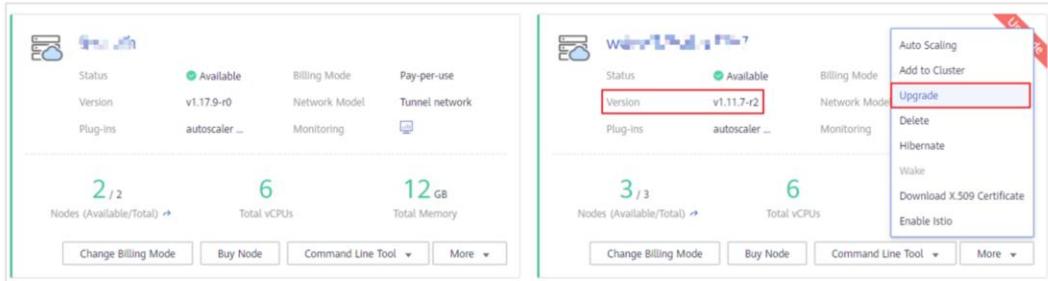
**Node Configuration**

Billing Mode	Pay-per-use	AZ	Specifications
Node Quantity	1	System Disk	Data Disk
EIP		Subnet subnet-479e	Network Model Tunnel network
Cloud Server Fee	¥0.00 /hour		

Currently automatic scale-in is not supported. Manual scale-in can be performed according to resource usage.

## Upgrading a Cluster

- Currently, you can upgrade only CCE clusters containing VM nodes. CCE clusters consisting of BMS nodes or nodes created from private images, CCE Turbo clusters, and Kunpeng clusters cannot be upgraded.



89      Huawei Confidential

HUAWEI

- Precautions:

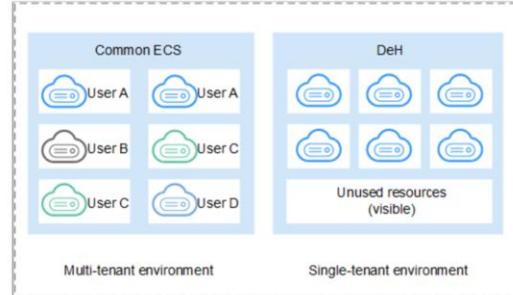
- Upgraded clusters cannot be rolled back. Therefore, perform the upgrade during off-peak hours to minimize the impact on your services.
- Before upgrading a cluster, get familiar with the features and differences of each cluster version in the Kubernetes Release Notes. Exceptions may occur after the upgrade if applications are incompatible with the new cluster version.
- Do not shut down or restart nodes during cluster upgrade. Otherwise, the upgrade will fail.
- Before upgrading a cluster, disable auto scaling policies to prevent node scaling during the upgrade. Node scaling will cause the upgrade to fail.
- If you locally modify the configurations of a cluster node, the cluster upgrade may fail or the configuration may be lost after the upgrade. You are advised to modify the configurations on the CCE console (cluster or node pool list page) so that they will be automatically inherited during the upgrade.
- During the cluster upgrade, the running workloads will not be interrupted, but access to the API server will be temporarily interrupted.

# Contents

1. Elastic Cloud Server (ECS)
2. Bare Metal Server (BMS)
3. Image Management Service
4. Auto Scaling (AS)
5. Cloud Container Engine (CCE)
- 6. Other Compute Services**

## What Is DeH?

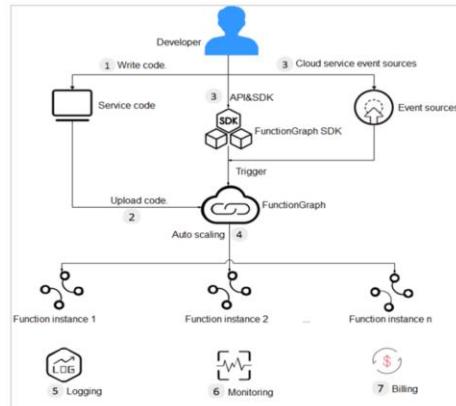
- A Dedicated Host (DeH) is a physical server fully dedicated for your own services. DeH allows you to ensure performance by keeping compute resources isolated. DeH also allows you to use your existing software licenses, so you can leverage existing investments to save money.



- Application scenarios:
  - Industries that have high requirements for regulation compliance and security: You can exclusively use a physically isolated host to meet compliance and security requirements.
  - Tenants that need to use their existing licenses (BYOL): If you have a licensed OS or software (licensed based on the number of physical sockets or cores), you can bring your own license and migrate your services to the cloud platform.
  - Industries that are extremely sensitive to performance and stability: DeH is ideal for service scenarios with higher requirements on server performance and stability such as finance, securities and gaming applications. DeH guarantees the stability of CPUs and network I/O, ensuring smooth running of applications.
  - Independent resource deployment and flexible management: You can create ECSs on a specified DeH and specify your ECS specifications based on the type of DeH. You can also migrate ECSs between DeHs or migrate ECSs from public resource pool to a specified DeH.

## What Is FunctionGraph?

- FunctionGraph allows you to run your code without provisioning or managing servers, while ensuring high availability and scalability. All you need to do is upload your code and set execution conditions, and FunctionGraph will take care of the rest. You pay only for what you use and you are not charged when your code is not running.



- FunctionGraph is designed for real-time file and data stream processing, web and mobile app backends, and artificial intelligence (AI) applications.
  - FunctionGraph processes files in real time by triggering a function once a client uploads a file to OBS. Functions can generate image thumbnails, convert video formats, and aggregate and filter data files.
  - FunctionGraph also works with Data Ingestion Service (DIS) to process data streams in real time. It supports application activity tracking, sequential transaction processing, data stream analysis, data sorting, metric generation, log filtering, indexing, social media analysis, and IoT device data telemetry and metering.
  - FunctionGraph also interconnects with your VMs or other services to build highly available and scalable web and mobile app backends.
  - Finally FunctionGraph also works with Enterprise Intelligence (EI) services for text recognition and illicit image identification. For example, build a function to identify pornographic and terrorism-related images.

## Quiz

1. (True or False) There is a hypervisor layer in containerization, just like the traditional virtualization featuring VMs.
  - A. True
  - B. False
2. (True or False) The functions of an IMS image are the same as those of an ISO image.
  - A. True
  - B. False

- False. Containerization has no virtualization layer.
- False. An ISO image is used to install an OS. An IMS image is more like a template that is generated after an ISO image is modified. It is mainly used to batch create cloud servers instead of just installing cloud server OSs.

## Summary

- This chapter described compute cloud services. After completing this course, you will be able to understand each phase of technical transformation, from hardware, virtualization, cloud platform, and to cloud services. In this process, many new products, such as Elastic Cloud Server (ECS) and Cloud Container Engine (CCE) will be used. Both of these products can be used to deploy application systems, but the technical architectures are different. Therefore, to better help enterprises migrate their service systems to the cloud, you need to clearly understand the technical details of each cloud service.

# Recommendations

- Huawei iLearning
  - <https://e.huawei.com/en/talent/>
- HUAWEI CLOUD Help Center
  - <https://support.huaweicloud.com/intl/en-us/help-novice.html>
- HUAWEI CLOUD Academy
  - <https://edu.huaweicloud.com/intl/en-us/>

## Acronyms and Abbreviations

- AI: Artificial intelligence
- API: Application Programming Interface
- AS: Auto Scaling
- BMS: Bare Metal Server
- CCE: Cloud Container Engine
- CI/CD: Continuous Integration/Continuous Delivery
- CISC: Complex Instruction Set Computer
- CPH: Cloud Phone
- CPU: Central Processing Unit
- DeH: Dedicated Host

## Acronyms and Abbreviations

- DevOps: Development and Operations
- DHCP: Dynamic Host Configuration Protocol
- ECS: Elastic Cloud Server
- EI: Enterprise Intelligence
- GPU: Graphics Processing Unit
- HPC: High Performance Computing
- HTTPS: Hypertext Transfer Protocol over Secure Sockets Layer
- IB: InfiniBand
- IMS: Image Management Service
- K8s: Kubernetes

## Acronyms and Abbreviations

- IPoIB: Internet Protocol over Infiniband
- NUMA: Non-Uniform Memory Access
- RDMA: Remote Direct Memory Access
- RISC: Reduced Instruction Set Computer
- SR-IOV: Single Root Input/Output Virtualization
- VLAN: Virtual Local Area Network
- VPC: Virtual Private Cloud

# Thank you.

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

Copyright©2021 Huawei Technologies Co., Ltd.  
All Rights Reserved.

The information in this document may contain predictive  
statements including, without limitation, statements regarding  
the future financial and operating results, future product  
portfolio, new technology, etc. There are a number of factors that  
could cause actual results and developments to differ materially  
from those expressed or implied in the predictive statements.  
Therefore, such information is provided for reference purpose  
only and constitutes neither an offer nor an acceptance. Huawei  
may change the information at any time without notice.



# Network Cloud Services



 HUAWEI

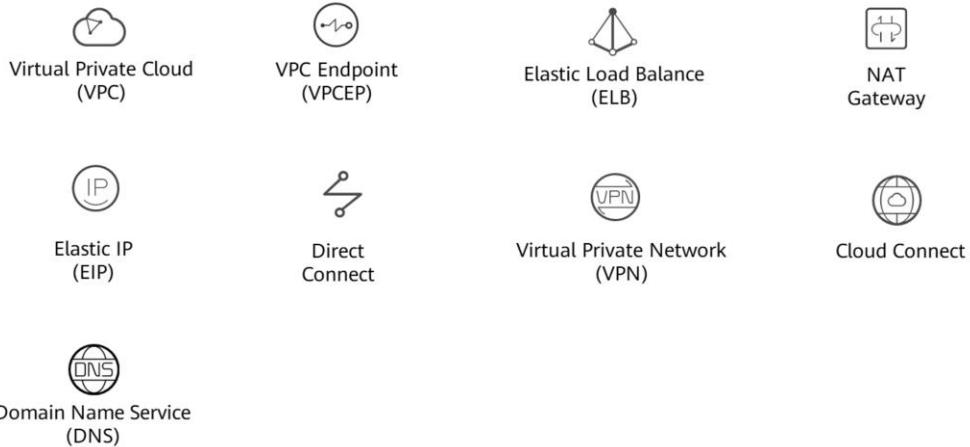
# Foreword

- Network resources are essential to the development of the ICT infrastructure. With network resources, devices and systems can communicate with each other so that enterprises can provide better services to their end users.
- This chapter describes the network services provided by HUAWEI CLOUD.

# Objectives

- On completion of this course, you will be able to:
  - Understand what network services are and what scenarios different services are designed for.
  - Understand how network services work and how you can use them.

## Network Service Overview



4 Huawei Confidential



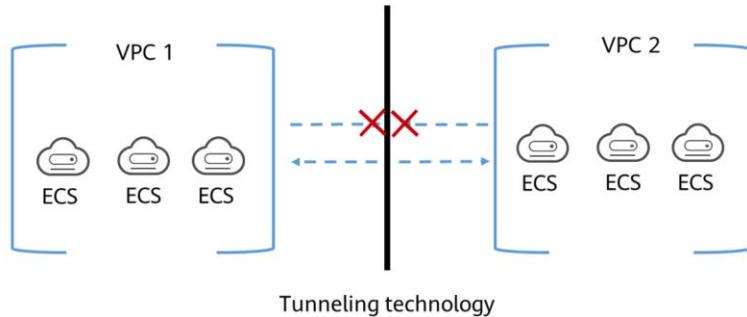
- VPC provides an isolated network environment on HUAWEI CLOUD.
- VPCEP provides secure access to cloud services and private services hosted on HUAWEI CLOUD.
- ELB automatically distributes incoming traffic across multiple backend servers.
- NAT Gateway provides network address translation (NAT) for cloud servers.
- EIP provides independent public IP addresses for accessing the Internet.
- Direct Connect establishes a dedicated channel between an on-premises data center and the cloud.
- VPN establishes an IPsec encrypted channel between an on-premises data center and the cloud.
- CC connects VPCs in multiple regions and allows one or more on-premises data centers to access multiple VPCs.
- DNS provides authoritative DNS services and domain name management services.

# Contents

- 1. Virtual Private Cloud (VPC)**
2. Elastic Load Balance (ELB)
3. Virtual Private Network (VPN)
4. NAT Gateway
5. Other Services

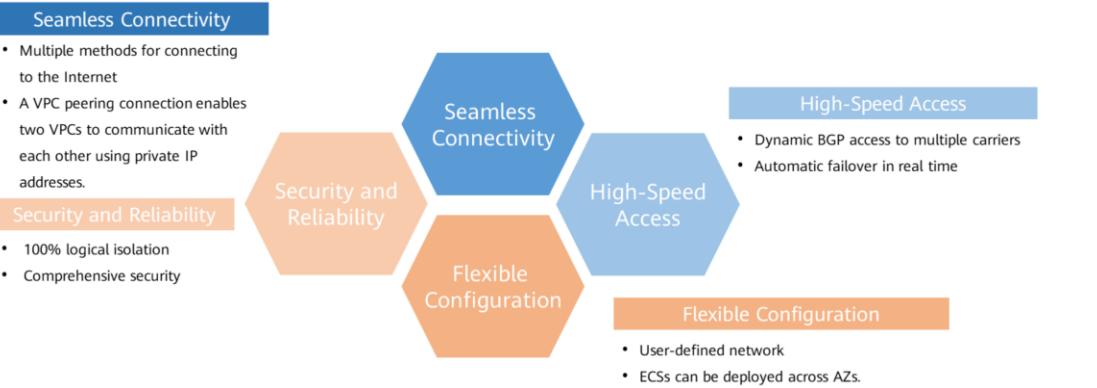
## What Is a VPC?

- A Virtual Private Cloud (VPC) is a logically isolated virtual network. Within your own VPC, you can create subnets, configure route tables, assign EIPs and bandwidths, and configure security groups to manage access control.



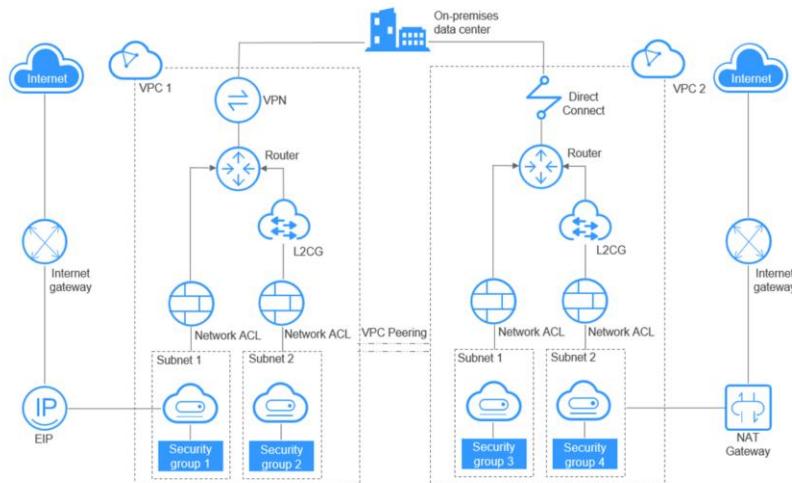
- VPC is the basis of HUAWEI CLOUD networks. VPC provides secure and isolated networks based on tunneling technology. You can customize your own VPCs, including dividing subnets, configuring route tables, specifying IP addresses, and configuring network ACLs and security groups.

# VPC Advantages



- A VPC has many advantages.
  - Flexible configuration: You can customize VPCs, divide subnets as required, and configure DHCP and route tables. ECSs can be deployed across AZs.
  - Security and reliability: VPCs are logically isolated from each other. By default, different VPCs cannot communicate with each other. Network ACLs protect subnets, and security groups protect ECSs.
  - Seamless connectivity: By default, a VPC cannot communicate with the Internet. You can use EIP, ELB, NAT Gateway, VPN, and Direct Connect to enable access to or from the Internet. By default, two VPCs in the same region cannot communicate with each other. You can create a VPC peering connection to enable them to communicate with each other using private IP addresses.
  - High-speed access: Up to 21 dynamic BGP connections are established to multiple carriers. Dynamic BGP provides automatic failover in real time and chooses the optimal path when a network connection fails.

## VPC Architecture



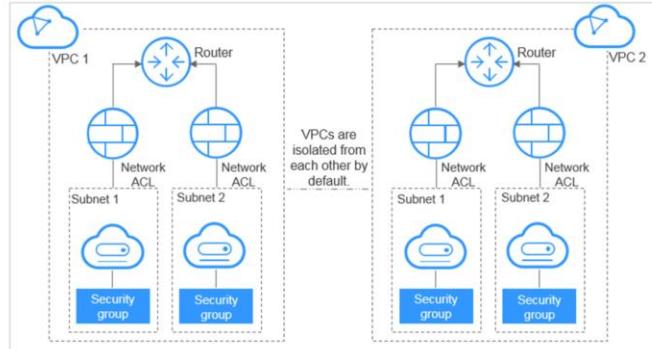
8      Huawei Confidential



- Each VPC consists of a private CIDR block, route tables, and at least one subnet.
  - When you create a VPC, you need to specify the private CIDR block for the VPC. The VPC service supports CIDR blocks 10.0.0.0/8-24, 172.16.0.0/12-24, and 192.168.0.0/16-24.
  - Cloud resources, such as cloud servers and databases, must be deployed in subnets, so you need to divide your VPC into one or more subnets.
  - When you create a VPC, the system automatically generates a default route table for the VPC. The route table ensures that all subnets in the VPC can communicate with each other. If the routes in the default route table cannot meet application requirements (for example, a cloud server without an EIP bound needs to access the Internet), you can create a custom route table.

## Application Scenario - Dedicated Networks on Cloud

- Each VPC represents a private network and is logically isolated from other VPCs. You can deploy your service systems in a private network on the cloud. If you have multiple service systems, for example, a production system and a test system, you can keep them isolated by deploying them in two different VPCs.



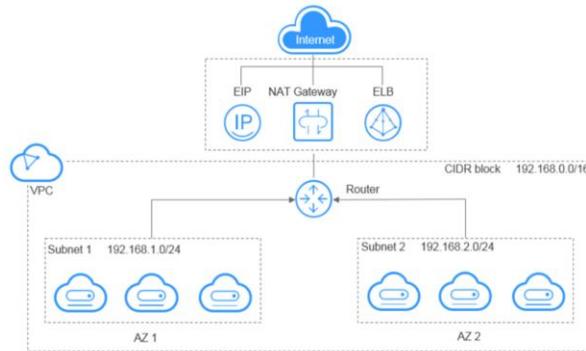
10      Huawei Confidential



- To enable two VPCs in the same region to communicate with each other, you can create a VPC peering connection between them.

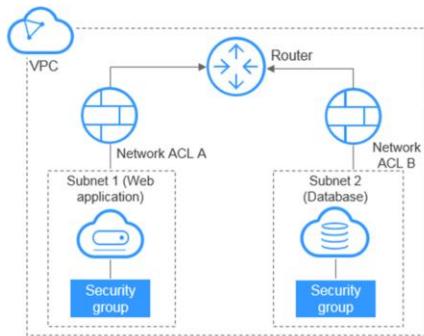
## Application Scenario - Web Application/Website Hosting

- You can host web applications and websites in a VPC and use the VPC as a regular network. With EIPs or NAT gateways, you can connect ECSs running your web applications to the Internet. You can then use load balancers provided by the ELB service to evenly distribute traffic across multiple ECSs.



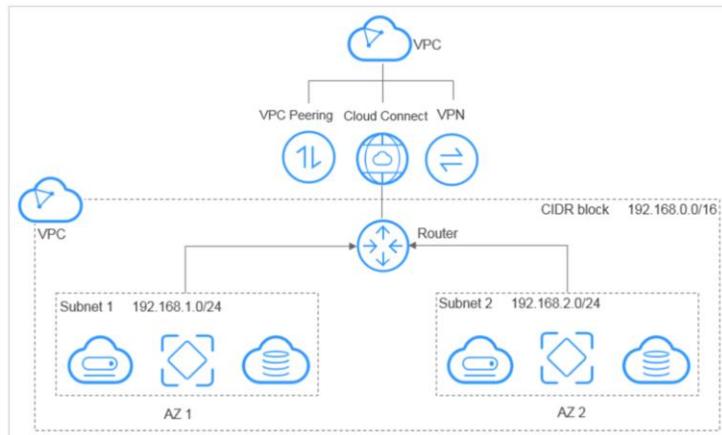
## Application Scenario - Web Application Access Control

- You can create a VPC and multiple security groups to associate web servers and database servers with different security groups and configure different access control rules for security groups. You can launch web servers in a publicly accessible subnet, and also run database servers in subnets that are not publicly accessible. In this way, you can ensure high security.



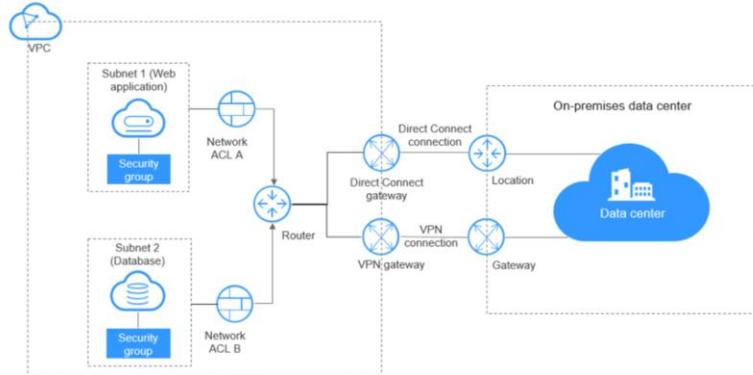
## Application Scenario - VPC Connectivity Options

- You can use the following cloud services to allow two VPCs to communicate with each other.

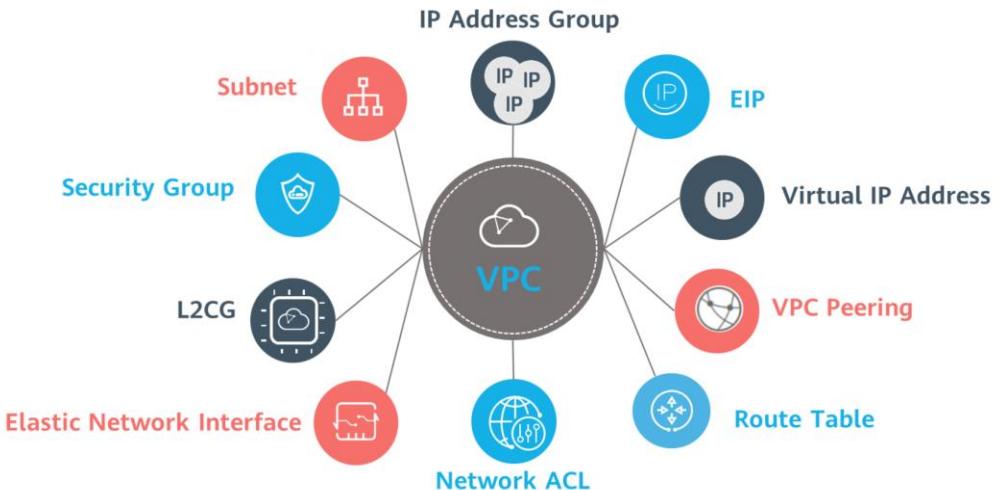


## Application Scenario - Hybrid Cloud Deployment

- If you have an on-premises data center and you do not want to migrate all of your business to the cloud, you can build a hybrid cloud. That way you can keep core data in your own data center.



## VPC Concepts



15      Huawei Confidential



- An elastic network interface is a virtual network card. You can create and configure network interfaces and attach them to your instances (ECSs and BMSs) to create flexible and high availability network configurations.
- An IP address group is a collection of IP addresses that use the same security group rules. You can use an IP address group to manage IP addresses that have the same security requirements or whose security requirements change frequently. An IP address group frees you from repeatedly modifying security group rules and simplifies security group rule management.

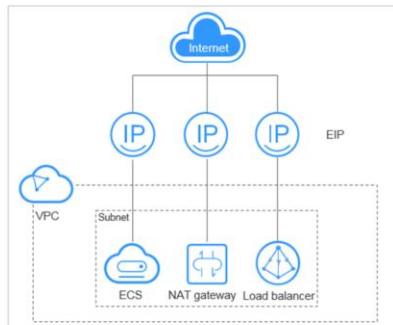
## VPC - Subnet

- A subnet is a unique CIDR block, a range of IP addresses, in your VPC. All resources in a VPC must be deployed on subnets. Once a subnet has been created, its CIDR block cannot be modified.



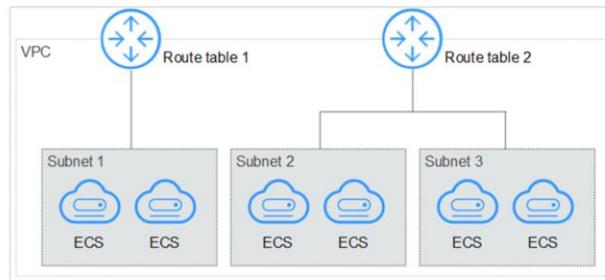
## VPC - EIP

- The Elastic IP (EIP) service enables your cloud resources to communicate with the Internet using static public IP addresses and scalable bandwidths. EIPs can be bound to or unbound from ECSs, BMSs, virtual IP addresses, load balancers, and NAT gateways. Various billing modes are provided to meet diverse service requirements. Each EIP can be used by only one cloud resource at a time.



## VPC - Route Table

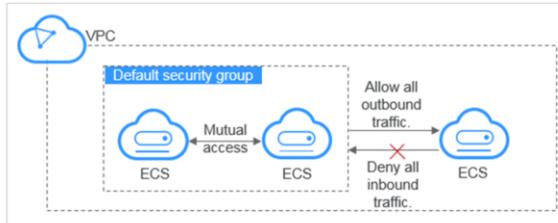
- A route table contains a set of routes that are used to determine where network traffic from your subnets in a VPC is directed. Each subnet in a VPC must be associated with a route table. A route table can be associated with multiple subnets. However, each subnet can only be associated with one route table.



- You can add, query, modify, and delete routes.
- When you create a VPC, the system automatically generates a default route table for the VPC. If you create a subnet in the VPC, the subnet automatically associates with the default route table. You can add, delete, and modify routes in the default route table, but you cannot delete the route table. When you create a VPN connection, the default route table automatically delivers a route that cannot be deleted or modified. If you want to modify or delete the route, you can associate your subnet with a custom route table and replicate the route to the custom route table to modify or delete it.
- You can also create a custom route table and associate subnets that have the same routing requirements with this table. Custom route tables can be deleted if they are no longer required.
- The way you can access the route table module varies by region.
  - If the route table module is not decoupled from the VPC module in your selected region, access the route table module by clicking the **Route Tables** tab on the VPC details page.
  - If the route table module is decoupled from the VPC module in your selected region, access the route table module by clicking **Route Tables** in the left navigation pane of the VPC console.

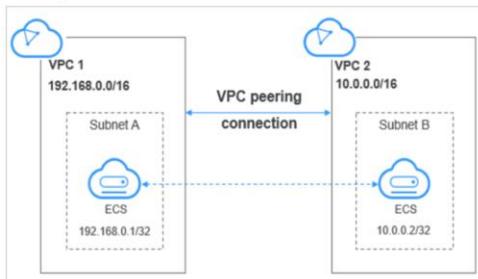
## VPC - Security Group

- A security group is a collection of access control rules for ECSs that have the same security requirements and are mutually trusted within a VPC. After you create a security group, you can create different access rules for the security group, and the rules will apply to any ECS that the security group contains.



## VPC - VPC Peering

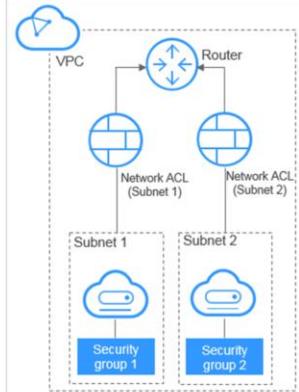
- A VPC peering connection is a network connection between two VPCs in the same region. It enables you to route traffic between them using private IP addresses. You can create a VPC peering connection between your own VPCs, or between your VPC and a VPC of another account within the same region. However, you cannot create a VPC peering connection between VPCs in different regions.



- If you create a VPC peering connection between two VPCs in your account, the system accepts the connection by default. To enable communication between the two VPCs, you need to add routes for the local and peer VPCs.
- If you request a VPC peering connection with a VPC in another account in the same region, the VPC peering connection will be in the **Awaiting acceptance** state. After the owner of the peer account accepts the connection, the connection status changes to **Accepted**. The owners of both the local and peer accounts must configure the routes required by the VPC peering connection to enable communication between the two VPCs.

## VPC - Network ACL

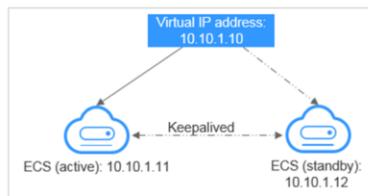
- A network ACL allows you to create rules to control traffic in and out of one or more subnets.



- Similar to security groups, network ACLs control access to subnets, but they add an additional layer of security. Security groups only have allow rules, but network ACLs have both allow and deny rules. You can use network ACLs together with security groups to implement fine-grained and comprehensive access control.

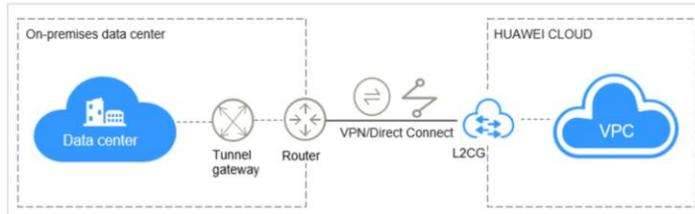
## VPC - Virtual IP Address

- A virtual IP address can be shared among multiple ECSs. An ECS can have both private and virtual IP addresses, and you can access the ECS through either IP address. A virtual IP address has the same network access capability as a private IP address. Virtual IP addresses are used for high availability as they make active/standby ECS switchover possible.



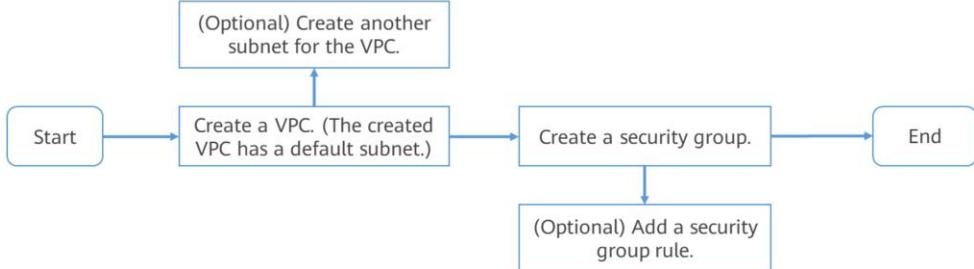
## VPC - L2CG

- An L2CG is a virtual tunnel gateway that works with Direct Connect or VPN to establish network communication between cloud and on-premises networks. The gateway allows you to migrate data center or private cloud services to the cloud without changing subnets and IP addresses.



- A Direct Connect or VPN connection establishes a Layer 3 network tunnel between cloud and on-premises networks, but the subnets on the cloud and on-premises networks cannot overlap. If the cloud and on-premises networks are on the same subnet and need to communicate with each other, you can use a L2CG to enable communication over a Layer 2 network.
- An L2CG is a tunnel gateway of a VPC and corresponds to a tunnel gateway of your data center. An L2CG can work together with a Direct Connect or VPN connection to establish a Layer 2 network between a VPC and your data center.
- A Layer 2 connection connects a VPC subnet to an L2CG and specifies the L2CG to connect to the tunnel gateway in an enterprise data center. This enables the VPC subnet to communicate with the subnet in the enterprise data center at a Layer 2 network.

# VPC Configuration Process



- Before creating your VPCs, determine how many VPCs, the number of subnets, and what IP address ranges you will need. Ensure that the subnets do not overlap with those of the end of VPN or Direct Connect connections.
- For network security, define access control policies based on specific services and minimize the access permissions. For example, a security group allows access from only certain source IP addresses on certain ports.

## VPC Configuration - Subnet

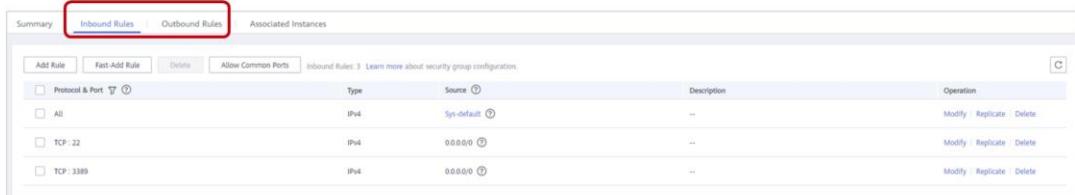
- Each VPC comes with a default subnet. If the default subnet cannot meet your requirements, create one.
- The subnet is configured with DHCP by default. When an ECS in this subnet starts, the ECS automatically obtains an IP address using DHCP.
- An AZ is a physical location where resources use independent power supplies and networks within a given region.

The screenshot shows a 'Default Subnet' configuration page. It includes fields for 'AZ' (set to 'AZ3'), 'Name' (set to 'subnet-406f'), and 'CIDR Block' (set to '192.168.0.0/24'). Below these fields, a note states: 'The CIDR block cannot be modified after the subnet has been created.'

- The CIDR block of a subnet can be within the CIDR block for the VPC. The supported CIDR blocks are 10.0.0.0/8-24, 172.16.0.0/12-24, and 192.168.0.0/16-24.
- An external DNS server address is used by default. If you need to change the DNS server address, ensure that the configured DNS server address is available.
- Dynamic Host Configuration Protocol (DHCP) is a network protocol for local area networks. It means that the server controls a range of IP addresses, and the client can automatically obtain the IP address and subnet mask assigned by the server when logging in to the server.

## VPC Configuration - Security Group

- Your account automatically comes with a default security group. You can add inbound and outbound rules to the default security group or create a new security group.
- Inbound rules control incoming traffic to ECSs in the security group.
- Outbound rules control outgoing traffic from ECSs in the security group.
- Default security group rules



The screenshot shows the 'Inbound Rules' tab selected in a navigation bar. Below it is a table listing four existing rules:

Type	Source	Description	Operation
IPv4	Sys-default	...	Modify Replicate Delete
IPv4	0.0.0.0	...	Modify Replicate Delete
IPv4	0.0.0.0	...	Modify Replicate Delete

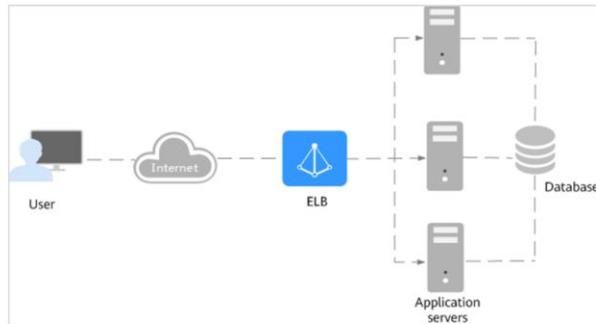
- The default security group cannot be deleted, but you can modify the rules in the default security group.
- If two ECSs are in the same security group but in different VPCs, the ECSs cannot communicate with each other. To enable communications between the ECSs, use a VPC peering connection to connect the two VPCs.
- In a VPC, if you want to copy resources from an ECS in a security group to another ECS in another security group, you can add rules to enable internal network communication between the ECSs and then copy resources. Within a given VPC, ECSs in the same security group can communicate with one another by default. However, ECSs in different security groups cannot communicate with each other by default. To enable these ECSs to communicate with each other, you need to add certain security group rules.

# Contents

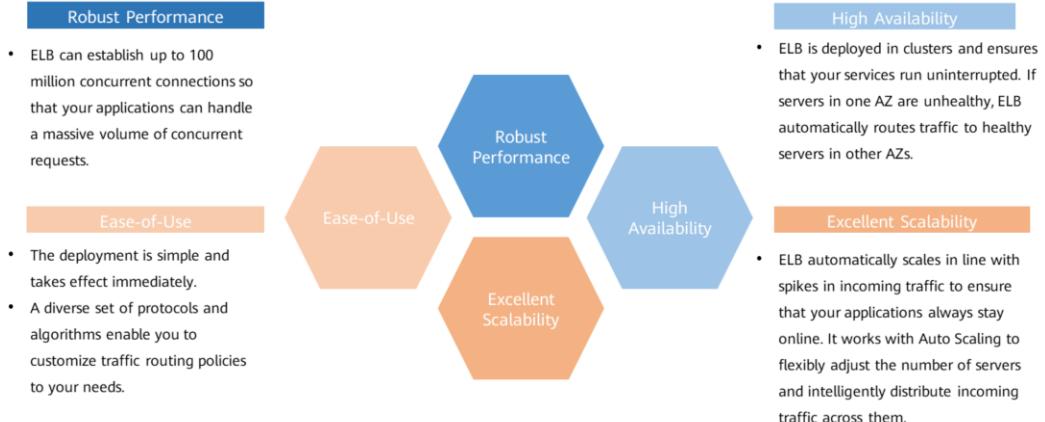
1. Virtual Private Cloud (VPC)
- 2. Elastic Load Balance (ELB)**
3. Virtual Private Network (VPN)
4. NAT Gateway
5. Other Services

## What Is ELB?

- Elastic Load Balance (ELB) automatically distributes incoming traffic across multiple backend servers based on the listening rules you configure. ELB expands the service capabilities of your applications and improves their availability by eliminating single points of failure (SPOFs).



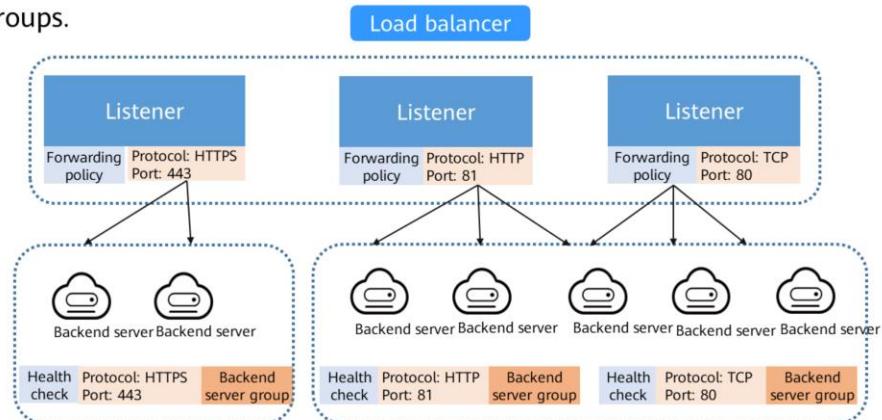
# ELB Advantages



- ELB has the following advantages:
  - Robust performance: ELB can establish up to 100 million concurrent connections so that your applications can handle a massive volume of concurrent requests.
  - High availability: ELB is deployed in clusters and ensures that your services run uninterrupted. If servers in one AZ are unhealthy, ELB automatically routes traffic to healthy servers in other AZs.
  - Excellent scalability: ELB automatically scales in line with spikes in incoming traffic to ensure that your applications always stay online. It works with Auto Scaling to flexibly adjust the number of servers and intelligently distribute incoming traffic across them.
  - Ease-of-use: A diverse set of protocols and algorithms enable you to customize traffic routing policies to your needs while keeping deployments simple.

## ELB Architecture

- ELB consists of three components: load balancers, listeners, and backend server groups.



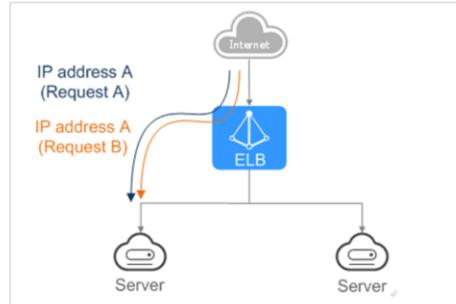
30 Huawei Confidential



- ELB consists of load balancers, listeners, and backend server groups.
  - A load balancer is an instance that distributes incoming traffic across backend servers in one or more availability zones (AZs).
  - A listener uses the protocol and port you specify to check for requests from clients and route the requests to associated backend servers based on the listening rules you define. You can add one or more listeners to a load balancer.
  - A backend server group uses the protocol and port you specify to receive the requests from the load balancer and route them to one or more backend servers. You need to add at least one backend server to a backend server group. You can set a weight for each backend server so that the load balancer can route requests based on their performance. You can also configure health checks for a backend server group to check the health of backend servers in the group. If a backend server is unhealthy, the load balancer stops routing new requests to this server until it recovers.

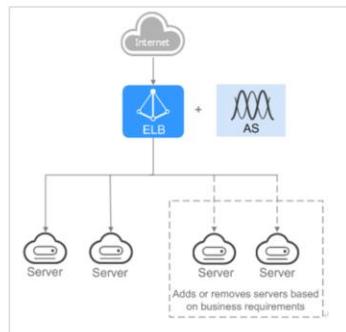
## Application Scenario: Heavy-Traffic Applications

- For an application with heavy traffic, such as a large web portal or mobile app store, ELB evenly distributes incoming traffic to multiple backend servers, balancing the load while ensuring stable performance. Sticky sessions ensure that requests from one client are always forwarded to the same backend server.



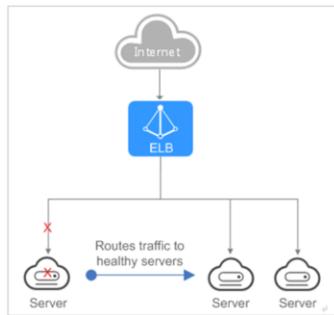
## Application Scenario: Applications with different Traffic

- For an application that has predictable peaks and troughs in traffic volumes, ELB works with AS to add or remove backend servers to keep up with changing demands. One example is flash sales, during which there are predictable traffic spikes that only last a short while. ELB can work with AS to run only the required number of backend servers needed to handle the load of your application.



## Application Scenario: Eliminating SPOFs

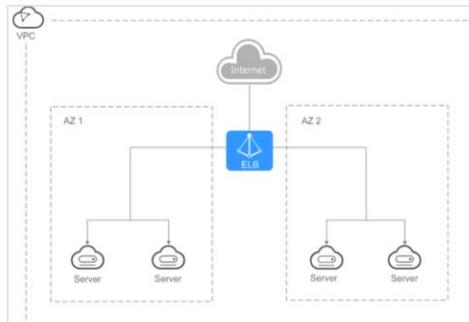
- ELB routinely performs health checks on backend servers. If any backend server is unhealthy, ELB will not route requests to this server until it recovers. This makes ELB a good choice for running applications that require high reliability.



- A single point of failure (SPOF) is a part of a system that, if it fails, will stop the entire system from working. SPOFs are undesirable in any system with a goal of high availability or reliability, such as a business system, software application, or other industrial system.

## Application Scenario: Cross-AZ Load Balancing

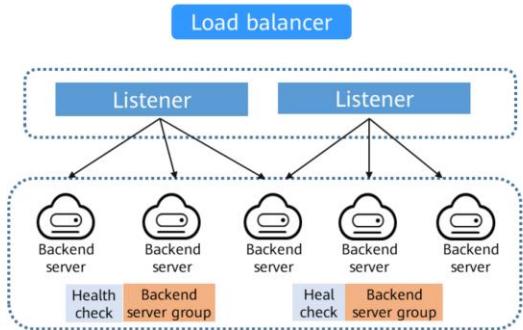
- ELB can distribute traffic across AZs. If an AZ becomes faulty, ELB distributes incoming traffic across backend servers in other AZs. It is useful for applications that require high availability.



- When you choose whether to deploy resources in the same AZ, consider your requirements for disaster recovery and network latency.
  - For high disaster recovery capabilities, deploy resources in different AZs but in the same region.
  - For lower latency, deploy resources in the same AZ.
- If you deploy resources in different AZs, you can use ELB to distribute traffic across AZs for real-time disaster recovery.

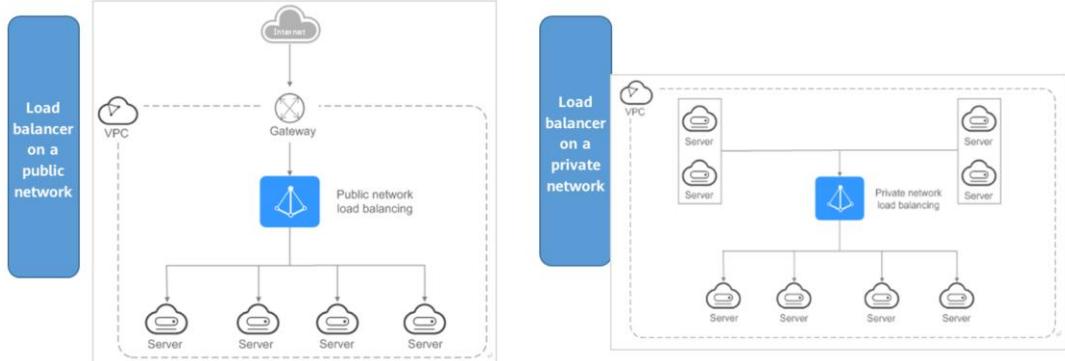
## ELB Concepts

- A load balancer distributes incoming traffic across backend servers.
- A listener listens on requests from clients and routes the requests to backend servers based on the settings that you configure when you add the listener.
- A backend server group is a group of cloud servers that have same features. When you add a listener, you select a load balancing algorithm and create or select a backend server group. Incoming traffic is routed to the corresponding backend server group based on the listener's configuration.
- ELB periodically sends heartbeat messages to associated backend servers to check their health and ensure that traffic is distributed only to healthy backend servers. This can improve the availability of your applications.



## ELB - Load Balancer

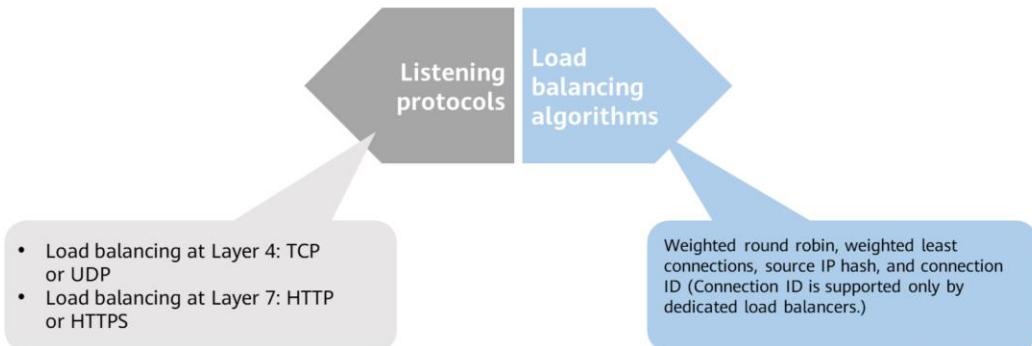
- A load balancer distributes incoming traffic across multiple backend servers. Load balancers can work on both public and private networks.



- Each load balancer on a public network has an EIP bound to it and routes requests from clients to backend servers over the Internet.
- Load balancers on a private network work within a VPC and route requests to backend servers in the same VPC as the clients.

## ELB - Listener

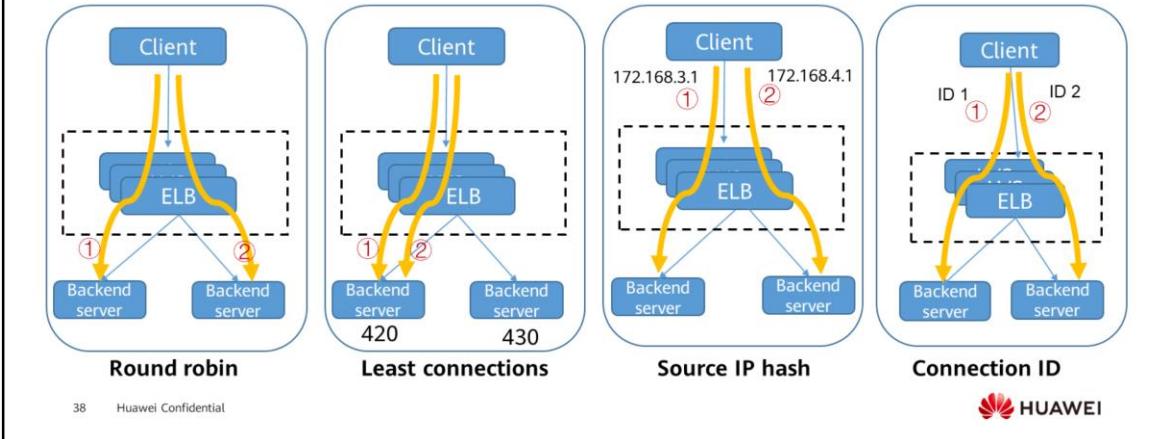
- A listener listens on requests from clients and routes the requests to backend servers based on the settings that you configure when you add the listener.



- A listener specifies the protocol and port used to receive requests from the clients, and the protocol, the port, and the load balancing algorithm to forward the requests to one or more backend servers. A listener also defines the health check configuration, which the load balancer uses to continually check the statuses of backend servers. If a backend server is unhealthy, the load balancer routes traffic to the healthy ones. Traffic routing to this server resumes after it recovers.
- The OSI model consists of the application layer, presentation layer, session layer, transport layer, network layer, data link layer, and physical layer.
  - Protocols at the application layer: HTTP, SNMP, FTP, NFS, Telnet, and SMTP
  - Protocols at the presentation layer: none
  - Protocols at the session layer: none
  - Protocols at the transport layer: TCP and UDP
  - Protocols at the network layer: IP and ICMP
  - Protocols at the data link layer: FDDI, Ethernet, ARPANET, PDN, SLIP, and PPP
  - Protocols at the physical layer: IEEE 802.1A, IEEE 802.2 to IEEE 802.11

## ELB - Backend Server Group

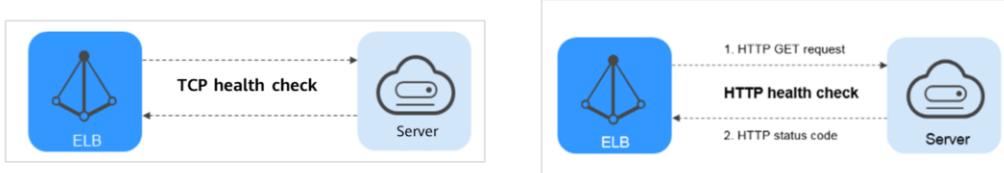
- A backend server group contains at least one backend server to process client requests forwarded by a load balancer. When you add a listener to a load balancer, you specify a backend server group to receive requests from the load balancer using the port and protocol you specify for the backend server group and the load balancing algorithm you select. ELB supports the following load balancing algorithms.



- Each backend server can be given a numeral value from 0 to 100 to indicate the proportion of requests the backend server can receive. The higher the weight, the more requests the backend server receives. You can set a weight for each backend server when you select one of the following algorithms:
  - Weighted round robin: Requests will not be routed to a backend server whose weight is 0, even if the backend server is considered healthy. If none of the servers have a weight of 0, the load balancer routes requests to these servers using the round robin algorithm based on their weights. If two backend servers have the same weights, they receive the same number of requests.
  - Weighted least connections: Requests will not be routed to a backend server whose weight is 0. If none of the servers have a weight of 0, the load balancer calculates each server's overhead using the formula: Overhead = Number of current connections/Server weight. The load balancer routes requests to the backend server with the lowest overhead.
  - Source IP hash: If a backend server's weight is 0, requests will not be routed to this server. If the server weights are not 0, they will not take effect, and requests from the same IP address will be routed to the same backend server.
  - Connection ID: If a backend server's weight is 0, requests will not be routed to this server. If the server weights are not 0, they will not take effect, and requests from the same client and with the same connection ID will be routed to the same backend server.

## ELB - Health Check

- ELB periodically sends heartbeat messages to associated backend servers to check their health and ensure that traffic is distributed only to healthy servers. This can improve the availability of your applications. If a backend server is unhealthy, the load balancer stops routing traffic to it. The load balancer will resume routing requests to the backend server after it recovers.



- How does a health check work?
  - UDP listeners: UDP is used for health checks by default, and UDP probe packets are sent to backend servers to obtain their health results.
  - TCP, HTTP, or HTTPS listeners: HTTP can be used for health checks. ELB sends HTTP GET requests to backend servers to check their health.
- If the health check result of a backend server is **Unhealthy**, you need to check its configuration.
- The security group that contains the backend servers must allow access from 100.125.0.0/16. Otherwise, health checks cannot be performed.
- If UDP is used for health checks, the backend server group's protocol must be UDP.

# ELB Configuration Process

## 1. Creating a Load Balancer

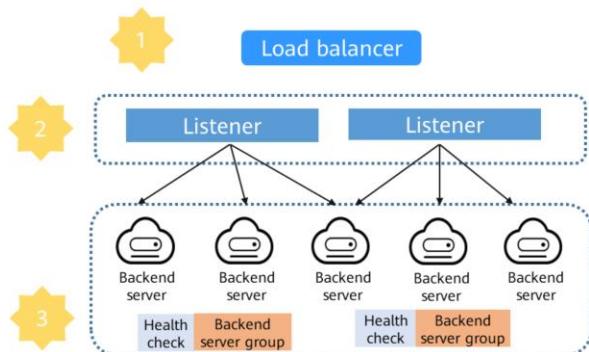
- Click Buy Elastic Load Balancer.
- Select the load balancer type.
- Configure the network.

## 2. Adding a Listener

- Locate the created load balancer.
- Configure the protocol and port.

## 3. Adding a Backend Server Group

- Select a load balancing algorithm.
- Configure a health check.



- A listener specifies the protocol and port used to receive requests from the clients, and the protocol, the port, and the load balancing algorithm to forward the requests to one or more backend servers. A listener also defines the health check configuration, which the load balancer uses to continually check the statuses of backend servers. If a backend server is unhealthy, the load balancer routes traffic to the healthy ones. Traffic routing to this server resumes after it recovers.

## ELB Configuration - Creating a Load Balancer

- Before creating a load balancer, you need to plan its region, network, protocol, and backend servers.

The screenshot shows the configuration page for creating a new Elastic Load Balancer (ELB). The form includes fields for Region (set to AP-Singapore), Network Type (Public network selected), VPC (vpc-501), Subnet (subnet-01), Private IP Address (Automatically-assigned IP), EIP (New EIP selected), and EIP Type (Dynamic BGP). The 'Region' field has a red box around it, indicating it's a required field. The 'EIP' field also has a red box around it.

- Click the icon in the upper left corner to select a region and a project.
- Hover on the upper left to display **Service List**. Under **Networking**, click **Elastic Load Balance**.
- Click **Buy Elastic Load Balancer** and then configure the parameters.
- Click **Next**.
- Confirm the configuration and submit your request.
- View the newly created load balancer in the load balancer list.

## ELB Configuration - Adding a Listener

- After you have created a load balancer, you need to add at least one listener. A listener listens on requests from clients and routes the requests to backend servers based on the settings that you configure when you add the listener.

The screenshot shows the 'Add Listener' wizard with three steps: 1. Configure Listener, 2. Configure Backend Server Group, and 3. Finish. Step 1 is active. It has fields for 'Name' (set to 'listener-vivi') and 'Frontend Protocol/Port' (set to 'HTTP' on port 8881). A note below says: 'Select TCP or UDP for load balancing at Layer 4. Select HTTP or HTTPS for load balancing at Layer 7.' and 'When HTTPS is selected, the backend protocol can only be HTTP.'

- Frontend Protocol/Port:** The load balancer uses the protocol and port to receive requests from clients and forward the requests to backend servers.
- Obtain Client IP Address:**
  - Enable this option if you want to pass source IP addresses of the clients to backend servers.
  - It is enabled for dedicated load balancers by default and cannot be disabled.

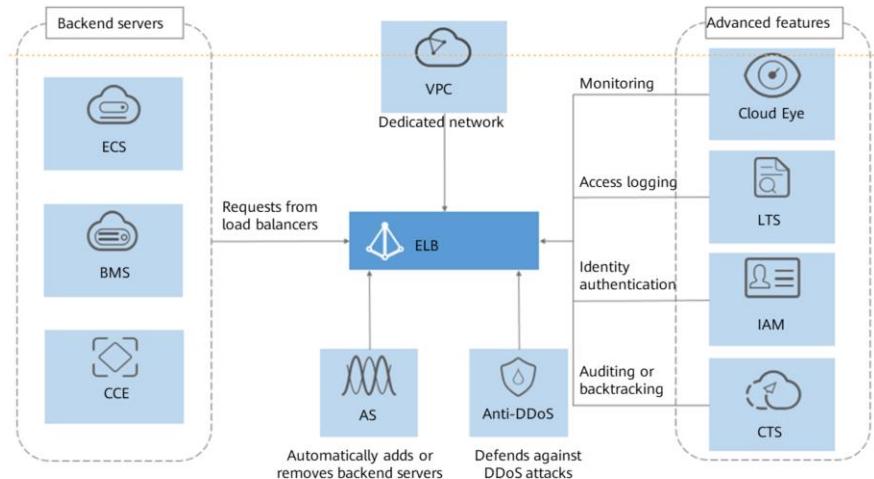
## ELB Configuration - Adding a Backend Server Group

- A backend server group is a collection of cloud servers that have the same features and receive the requests routed by the load balancer.

The screenshot shows the 'Backend Server Groups' tab selected in the navigation bar. A red box highlights the 'Add Backend Server Group' button. The main area displays the configuration for a group named 'server\_group-vivi' under 'Listener-vivi'. The 'Basic Information' section includes fields for Name, Listener, Load Balancing Algorithm (Weighted round robin), and Sticky Session (Disabled). The 'Available servers' section lists two servers: 'ecs-501' and 'ecs-502', both marked as 'Running' and 'Healthy' with a weight of 1 and port 8889. The 'Health Check Result' column shows green icons for both servers.

- The load balancer uses one of the following algorithms to distribute traffic:
  - Weighted round robin: Requests are distributed across backend servers in sequence based on their weights. Backend servers with higher weights receive proportionately more requests, whereas equally-weighted servers receive the same number of requests.
  - Weighted least connections: In addition to the number of active connections established with each backend server, each server is assigned a weight based on its capacity. Requests are routed to the server with the lowest connections-to-weight ratio.
  - Source IP hash: Requests from the same source IP address are routed to the same backend server.
  - Connection ID (only for dedicated load balancers): Requests from the same client and with the same connection ID are routed to the same backend server.

## ELB and Related Services



44 Huawei Confidential



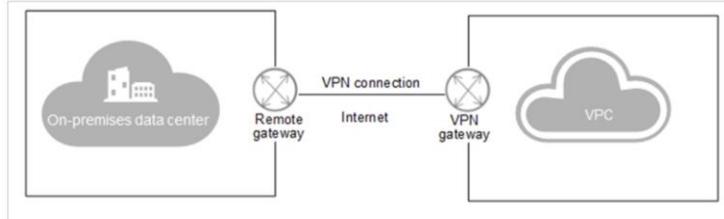
- ECS, BMS, and CCE provide cloud servers or containers to run your applications in the cloud. ELB is required to route traffic to these servers or containers.
- VPC provides IP addresses and bandwidth for load balancers.
- AS can work with ELB to automatically scale the number of backend servers.
- IAM provides authentication for ELB.
- CTS records the operations performed on ELB resources.
- Cloud Eye monitors the status of load balancers and listeners, without the need to install any additional plug-in.
- Anti-DDoS protects load balancers on a public network from DDoS attacks, keeping your applications stable and reliable.
- LTS can store access logs of HTTP or HTTPS requests to your load balancer for query and analysis later if you have enabled access logging.

# Contents

1. Virtual Private Cloud (VPC)
2. Elastic Load Balance (ELB)
- 3. Virtual Private Network (VPN)**
4. NAT Gateway
5. Other Services

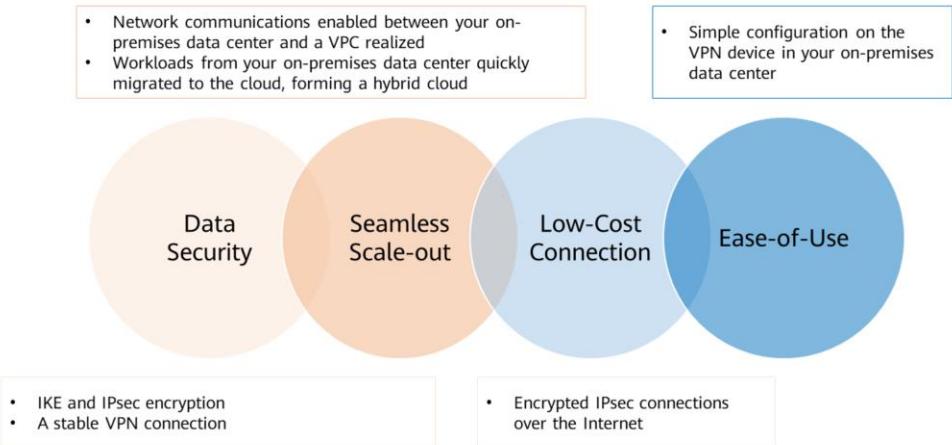
## Virtual Private Network

- Virtual Private Network (VPN) allows you to establish an encrypted, Internet-based communications tunnel between your on-premises data center and a VPC, so you can access resources in the VPC remotely.



- VPN tunnels support three protocols: PPTP, L2TP, and IPsec.

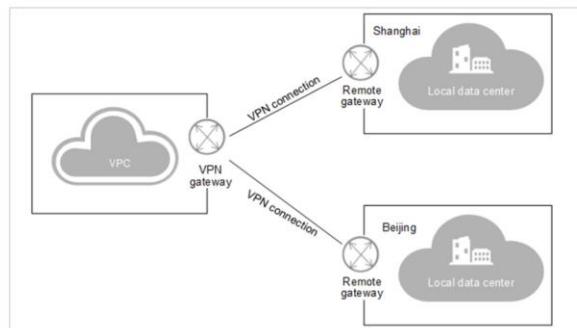
## VPN Advantages



- VPN advantages:
  - Data security: Huawei hardware uses IKE and IPsec to encrypt data to provide carrier-class reliability and ensure a stable VPN connection.
  - Seamless scale-out: With VPN, you can connect your on-premises data center to your VPC and quickly extend services from the data center to the cloud, forming a hybrid cloud.
  - Low-cost connection: Encrypted IPsec connections over the Internet provide a cost-effective alternative to Direct Connect connections.
  - Ease-of-use: You can create an easy-to-use VPN connection by specifying parameters on the VPN console and configuring VPN device in your on-premises data center.

## VPN Networking

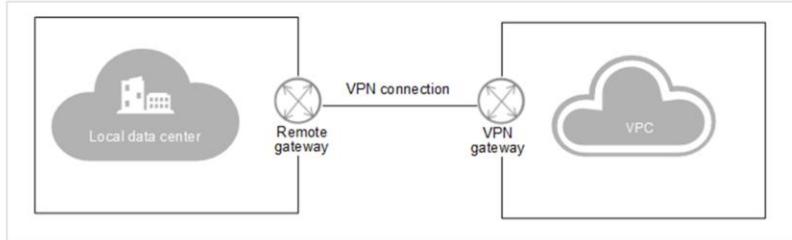
- A VPN consists of a VPN gateway and one or more VPN connections.
- A VPN gateway provides an Internet egress for a VPC and works together with the gateway in your on-premises data center.
- A VPN connection is an encrypted connection that links the VPN gateway to the remote gateway to enable communications between a VPC and your on-premises data center, quickly establishing a secure hybrid cloud.



- VPN components:
  - A VPN gateway is an egress gateway for a VPC. With a VPN gateway, you can create a secure, reliable, and encrypted connection between a VPC and your on-premises data center or between two VPCs in different regions. Each data center must have a gateway, which works as the remote gateway. Each VPC must have a VPN gateway. A VPN gateway needs to be paired with a remote gateway. Each VPN gateway can connect to one or more remote gateways, so you can set up point-to-point or hub-and-spoke VPN connections.
  - A VPN connection is a secure and reliable communications tunnel established between a VPN gateway and a gateway in your on-premises data center. Only IPsec VPNs are supported. VPN connections use IKE and IPsec protocols to cost-effectively and securely encrypt data transmitted over the Internet.

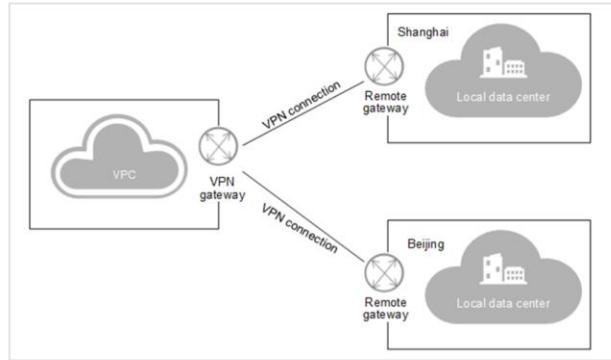
## Site-to-Site VPN Connection

- You can set up a VPN to connect your on-premises data center to a VPC, effectively creating a hybrid cloud.



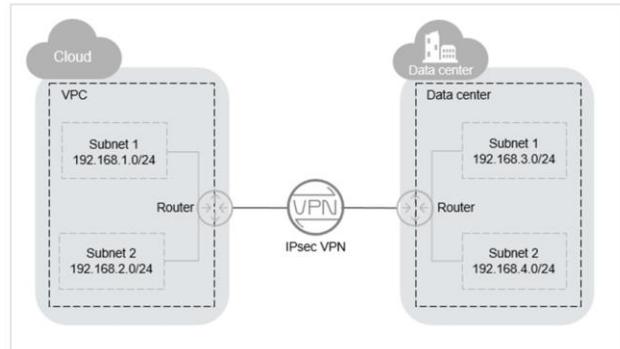
## Hub-and-Spoke VPN Connection

- You can also set up a VPN to connect multiple on-premises data centers to a VPC, also creating a hybrid cloud.



## VPN Concepts - IPsec VPN

- Internet Protocol Security (IPsec) VPN uses a secure network protocol suite that authenticates and encrypts data packets to provide secure encrypted communications between different networks. The VPN service uses an IPsec VPN.



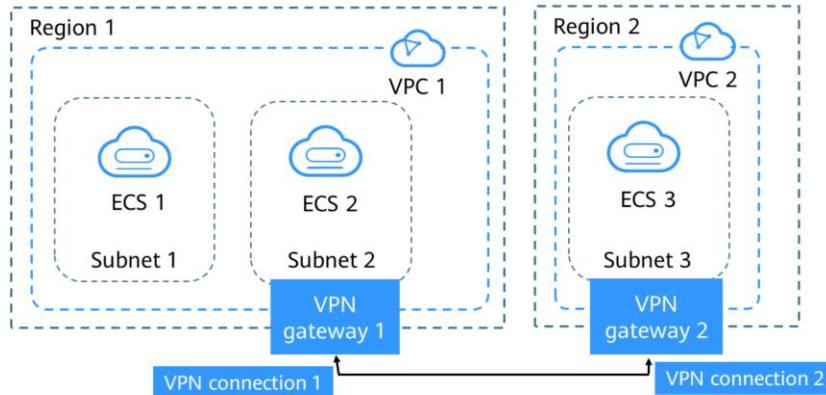
51      Huawei Confidential



- In the figure, the VPC has subnets 192.168.1.0/24 and 192.168.2.0/24. The on-premises data center has subnets 192.168.3.0/24 and 192.168.4.0/24. You can use VPN to enable subnets in the VPC to communicate with those in your on-premises data center.

## VPN Configuration Process

- You can create a VPN gateway and a VPN connection on the management console.



52      Huawei Confidential



- A VPN enables communications between VPCs in different regions.
- In this example, ECS 2 in region 1 needs to communicate with ECS 3 in region 2. A VPN connection linking region 1 and region 2 can make this possible.
- Step 1: Create a VPN gateway in region 1 and configure parameters such as Billing Mode, Region (Region 1), VPC (VPC 1), Billed by, Bandwidth, and encryption policies.
- Step 2: Create a VPN connection in region 1. Select subnet 2 for Local Subnet and subnet 3 for Remote Subnet. Configure the remote gateway. (VPN gateway 2 has not been created. Just enter a random address. You can change it later.)
- Step 3: Create a VPN gateway in region 2 and configure parameters such as Billing Mode, Region (Region 2), VPC (VPC 2), Billed by, Bandwidth, and encryption policies.
- Step 4: Create a VPN connection in region 2. Select subnet 3 for Local Subnet and subnet 2 for Remote Subnet. Configure the remote gateway by entering the IP address of VPN gateway 1.
- Step 5: Change the remote gateway address of VPN connection 1 to the address of VPN gateway 2.
- Step 6: Test the connectivity between ECS 2 and ECS 3 and check the VPN connection status.

## VPN Configuration: VPN Gateway

- To allow your ECSs in a VPC to access your on-premises network, you must first create a VPN gateway.

The screenshot shows a configuration dialog for creating a VPN gateway. The 'Name' field is set to 'Huawei-Vivi'. The 'VPC' dropdown is set to 'vpc-default'. The 'Type' dropdown is set to 'IPsec'. Under 'Billed By', the 'Bandwidth' tab is selected, while 'Traffic' is unselected. A slider for 'Bandwidth (Mbit/s)' is set to 5, with other options like 10, 20, 50, 100, 200, and 300 available. A 'Create VPC' button is located at the top right of the form.

- VPN Gateway**

- You can modify the name and description of a VPN gateway if needed. If the bandwidth of a VPN gateway cannot meet your requirements, you can modify the bandwidth, too. If the number of VPN connections associated with a VPN gateway cannot meet your requirements, you can modify the VPN gateway specifications. You can change the billing mode of a VPN gateway billed by bandwidth from per-use to yearly/monthly.
- If a VPN gateway is no longer required, you can delete it to release network resources as long as it has no VPN connections configured. If it has any connections configured, they have to be deleted before you can delete the gateway.

- VPC:** the name of the VPC that the VPN connects to

- Type:** the VPN type. **IPsec** is selected by default.

- Billed By:** There are two options available, bandwidth, and traffic.

- Bandwidth:** You specify a bandwidth and pay the bill based on the amount of time you use the bandwidth.
- Traffic:** You specify a bandwidth and pay for the total traffic you generate.

- Bandwidth (Mbit/s):**

- The bandwidth (Mbit/s) of the VPN gateway. The bandwidth size is shared by all VPN connections created for the VPN gateway. The total bandwidth size used by all VPN connections created for a VPN gateway cannot exceed the VPN gateway bandwidth size.
- If the network traffic exceeds the VPN gateway bandwidth, the network may get congested and VPN connections may be interrupted. Make sure you configure enough bandwidth.

- You can configure alarm rules on Cloud Eye to monitor the bandwidth.

## VPN Configuration: VPN Connection

- To connect your ECSs in a VPC to your private network, after the VPN gateway is obtained, you also have to create a VPN connection.

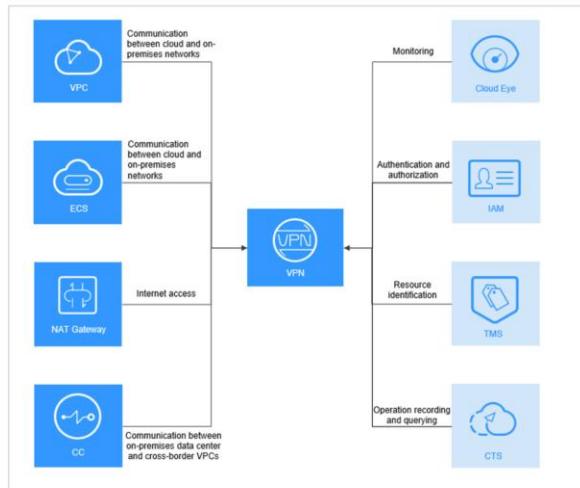
Name: vpn-43fd  
VPN Gateway: Select a VPN gateway to proceed.  
Local Subnet: Select subnet or Specify CIDR block  
Remote Gateway:  
Remote Subnet: Use commas (,) to separate multiple CIDR blocks, for example, 192.168.32.0/24,192.168.54.0/24  
PSK: Enter a pre-shared key.  
Confirm PSK: Enter the pre-shared key again.

54      Huawei Confidential



- VPN Connection:**
  - A VPN connection is an encrypted communications channel established between the VPN gateway in your VPC and that in your on-premises data center. The VPN connection can be modified later.
  - You can delete a VPN connection to release network resources if it is no longer required. When you delete the last VPN connection for a pay-per-use VPN gateway, the associated gateway will be deleted along with it.
- VPN Gateway:** the name of the VPN gateway used by the VPN connection
- Local Subnet:** the VPC subnets that will access your on-premises network through VPN. Possible values are **Select subnet** and **Specify CIDR block**.
- Remote Gateway:** the public IP address of the VPN device translated by the VPN gateway in your on-premises private network. This IP address is used for communications with your VPC.
- Remote Subnet:** the subnets of your on-premises network that will access the VPC through a VPN. The local subnet cannot include the CIDR block of the remote subnet.
- PSK:** Enter 6 to 128 characters. The PSK at both ends of a VPN connection must be the same.

## VPN and Related Services



55      Huawei Confidential



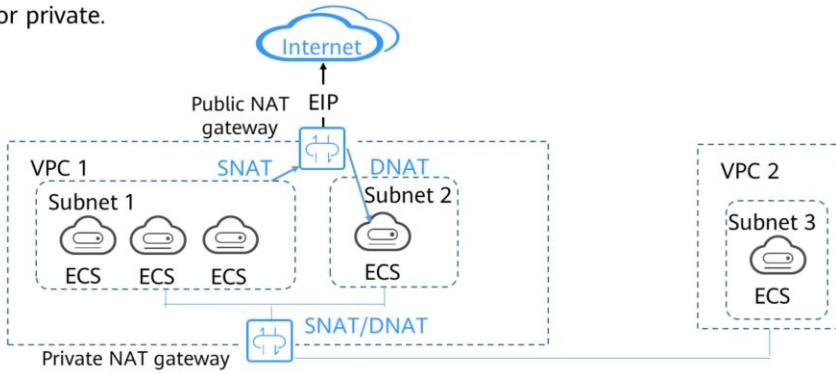
- **VPC** allows you to create a virtual private cloud that your on-premises data center can connect to. It also allows you to create security groups, add inbound and outbound rules, and add ECSs to the security groups, improving ECS access security.
- **Cloud Connect** works together with VPN to enable stable network communications between your on-premises data center and VPCs in different regions.
- **NAT Gateway** allows the servers in your on-premises data center that connected to a VPC using VPN to share EIPs to access the Internet or provide services that are accessible from the Internet.
- **ECS** allows you to create cloud servers, which can then be added security groups to control access.
- **Cloud Eye** monitors VPN resources and allows you to view metrics.
- **IAM** allows you to assign different permissions to different users. It enables fine grained control over your VPN resources.
- **CTS** records operations performed on VPN.

# Contents

1. Virtual Private Cloud (VPC)
2. Elastic Load Balance (ELB)
3. Virtual Private Network (VPN)
- 4. NAT Gateway**
5. Other Services

## NAT Gateway

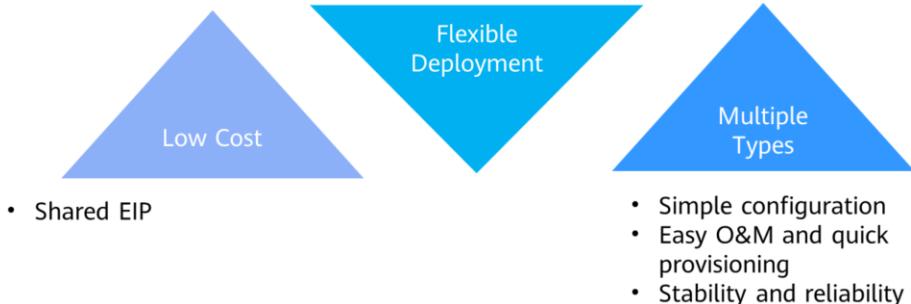
- The NAT Gateway service provides network address translation (NAT) service for servers in a VPC and enables servers to share an EIP to access the Internet. NAT gateways can be either public or private.



- NAT Gateway provides both source NAT (SNAT) and destination NAT (DNAT) for your resources in a VPC and allows servers in your VPC to access or provide services accessible from the Internet.
- A private NAT gateway provides NAT service for servers in a VPC, so that multiple servers can share a private IP address to access or provide services accessible from an on-premises data center or other VPCs.

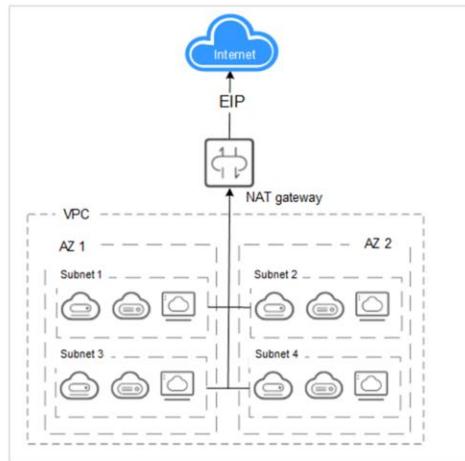
## NAT Gateway Advantages

- Cross-AZ deployment
- The type and EIP of a NAT gateway can be changed at any time.



- NAT Gateway has the following advantages:
  - Flexibility: A NAT gateway can be deployed flexibly across subnets and AZs. A fault in a single AZ does not affect the service continuity of a NAT gateway. The type and EIP of a NAT gateway can be changed at any time.
  - Ease of use: NAT gateway configuration is simple, the O&M is easy, and they can be provisioned quickly. Once provisioned, they are stable and reliable.
  - Cost-effectiveness: When you send data through a private IP address or your applications provide services accessible from the Internet using a NAT gateway, the NAT gateway translates the private IP address to a public IP address. NAT Gateway helps you save money on EIPs and bandwidth.

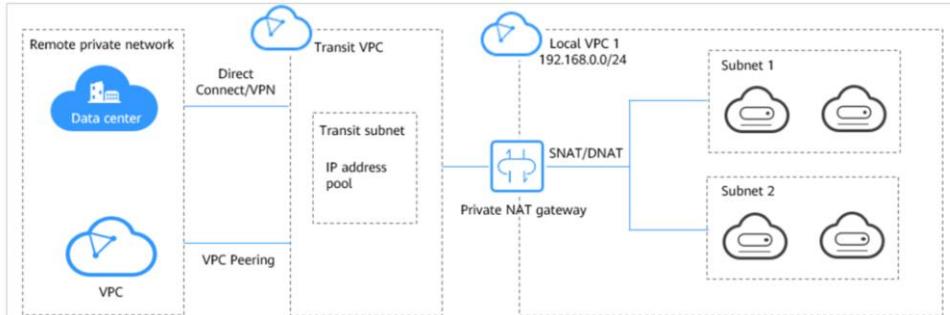
## NAT Gateway Architecture (Public NAT Gateway)



- Public NAT gateways support SNAT and DNAT.
  - SNAT translates private IP addresses into EIPs, allowing servers in a VPC to share an EIP to access the Internet in a secure and efficient way.
  - DNAT enables servers in a VPC to share an EIP to provide services accessible from the Internet through IP address mapping or port mapping.

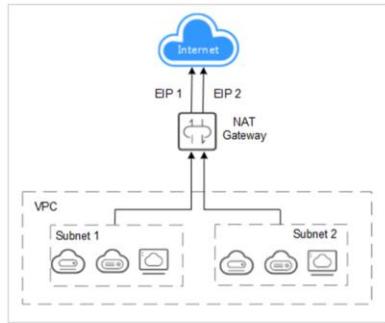
## NAT Gateway Architecture (Private NAT Gateway)

- A private NAT gateway provides NAT service for servers in a VPC, so that multiple servers can share a private IP address to access or provide services accessible from an on-premises data center or other VPCs.



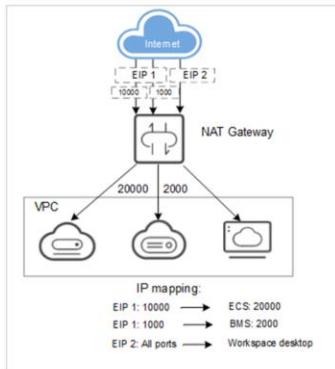
## Public NAT Gateway: Using SNAT

- If your servers in a VPC require Internet access, you can use SNAT to let the servers share one or more EIPs to access the Internet without exposing their IP addresses. NAT Gateway provides different types of NAT gateways for different numbers of connections. You can create multiple SNAT rules to meet different service requirements.



## Public NAT Gateway: Using DNAT

- DNAT lets servers in a VPC to provide services accessible from the Internet.



62      Huawei Confidential



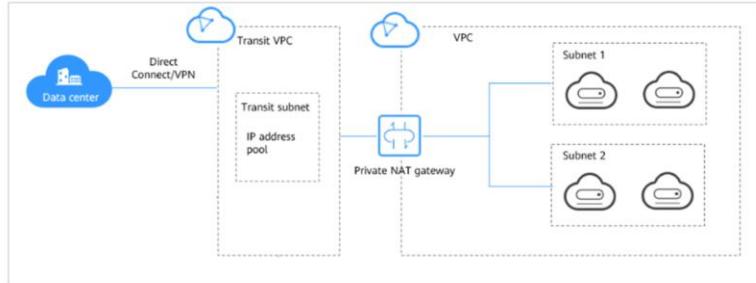
- You can bind an EIP to a DNAT rule. Then NAT gateways can forward requests -from a specific port and over a specific protocol to the EIP by port mapping. -to the EIP to your servers based on IP address mapping.

NAT Gateway allows multiple servers to share an EIP, saving costs on bandwidth.

- A DNAT rule is configured for one server. If there are multiple servers, you can create several DNAT rules to share one or more EIPs.

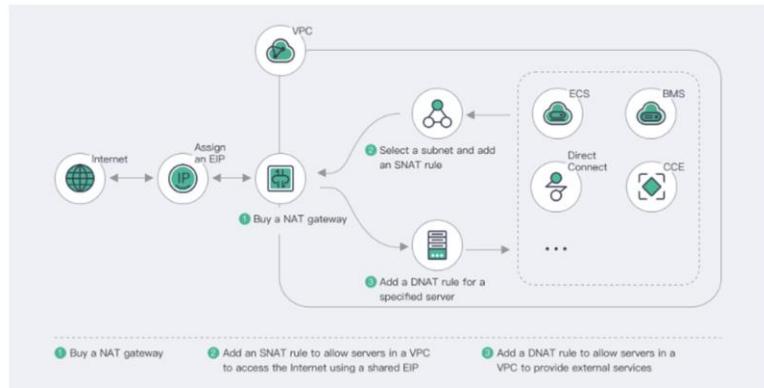
## Private NAT Gateway: Enterprise Network Management

- To ensure security compliance, an enterprise may require that all its branches and departments map their IP addresses to the same IP address for internal communications. To accomplish this, the enterprise can use a private NAT gateway to enable these communications without changing the original network after migrating workloads to the cloud.



# Process for Buying a NAT Gateway

Public NAT gateway:



- Before you use public NAT gateway, buy an EIP.
- SNAT translates private IP addresses into EIPs, allowing servers in a VPC to share an EIP to access the Internet in a secure and efficient way.
- DNAT enables servers in a VPC to share an EIP to provide services accessible from the Internet through IP address mapping or port mapping.
- SNAT and DNAT rules are designed for different functions. If an SNAT rule and a DNAT rule use the same EIP, there may be service conflicts.
- An SNAT rule cannot share an EIP with a DNAT rule with **Port Type** set to **All ports**.

## Buying a NAT Gateway

- When you buy a public NAT gateway, you must specify its VPC, subnet, and type.
- Check whether the default route (0.0.0.0/0) of the VPC is in use by any other gateways. If yes, add another route for the gateway you purchased or add the default route to a new route table that you will associate with the gateway.

The screenshot shows a configuration page for a NAT gateway. The 'Name' field is set to 'nat-Vivi'. The 'VPC' dropdown is set to 'vpc-default' and has a 'View VPC' link. The 'Subnet' dropdown is set to 'subnet-default (192.168.0.0/24)' and also has a 'View VPC' link. Below these, there's a note: 'The selected subnet is for the NAT gateway only. To enable communications over the Internet, after the NAT gateway is created, you need to add rules.' At the bottom, there are four size options: 'Small' (selected), 'Medium', 'Large', and 'Extra-large'. A note below says 'Supports up to 10,000 connections. Learn more'.

- **Subnet:**

- This is the subnet where the public NAT gateway is deployed.
- The subnet must have at least one available IP address.
- The selected subnet cannot be changed after the public NAT gateway is created.

- **Type:**

- The type can be **Small**, **Medium**, **Large**, and **Extra-large**. You can click **Learn more** on the page to view details about each type.

## SNAT Rule Configuration

- If your servers are in a VPC and need to access the Internet, select VPC.
- If your on-premises servers access a VPC over a Direct Connect or VPN connection need to access the Internet, select Direct Connect/Cloud Connect.

The screenshot shows the configuration page for a NAT gateway named 'nat-Vivi'. It highlights the 'Scenario' section, which is set to 'VPC'. Below this, the 'EIP' section is highlighted, showing a table of selected EIPs. The table includes columns for EIP, EIP Type, Bandwidth Name, Bandwidth (Mbps), and Billing Mode. One row is selected, showing '159.138.121.173' as a Dynamic BGP EIP with a bandwidth of 5 Mbps and Pay-per-use billing. A note at the bottom states: 'Selected EIPs: (1): 159.138.121.173. The EIP used for the SNAT rule will be randomly chosen from the ones selected here.'

EIP	EIP Type	Bandwidth Name	Bandwidth (Mbps)	Billing Mode
159.138.121.173	Dynamic BGP	bandwidth-Vivi	5	Pay-per-use

- **Scenario:**

- After the public NAT gateway is created, add SNAT rules to enable your cloud or on-premises servers to access the Internet by sharing an EIP.
- Each SNAT rule is configured for one subnet. If there are multiple subnets in a VPC, you can create several SNAT rules to allow them to share EIPs.

- **Elastic IP:**

- This is the EIP used for accessing the Internet.
- You can select only an EIP that is not bound to any resource, an EIP that is bound to a DNAT rule whose **Port Type** is not set to **All ports**, or an EIP that is bound to an SNAT rule of the current NAT gateway.
- You can select multiple EIPs at once. Up to 20 EIPs can be selected for each SNAT rule. If you have selected multiple EIPs for an SNAT rule, an EIP will be chosen from your selection at random.

## DNAT Rule Configuration

- VPC: A DNAT rule allows servers in a VPC to share an EIP and provide services accessible from the Internet.
- Direct Connect/Cloud Connect: A DNAT rule allows servers in an on-premises data center connected to a VPC through Direct Connect or Cloud Connect to provide services accessible from the Internet.

The screenshot shows the configuration of a DNAT rule. The 'NAT Gateway Name' is set to 'nat-Vivi'. Under 'Scenario', 'VPC' is selected. 'Port Type' is set to 'Specific port'. The 'Protocol' is 'TCP'. The 'EIP' is '159.138.121.173 (5 Mbit/s | Pay-per-use)'. The 'Outside Port' is '22'. The 'Private IP Address' is '192 . 168 . 10 . 1'. The 'Inside Port' is '22'. Bandwidth is '5 Mbit/s' and Billing Mode is 'Pay-per-use'.

- **Scenario:**

- After a public NAT gateway is created, you can add DNAT rules to allow servers in your VPC to provide services accessible from the Internet.
- You can configure a DNAT rule for each port on a server. If multiple servers need to provide services accessible from the Internet, create multiple DNAT rules.

- **Outside Port:**

- This is the port bound to the EIP. This parameter is available if you select **Specific port** for **Port Type**. Ports 1 to 65535 can all be selected.

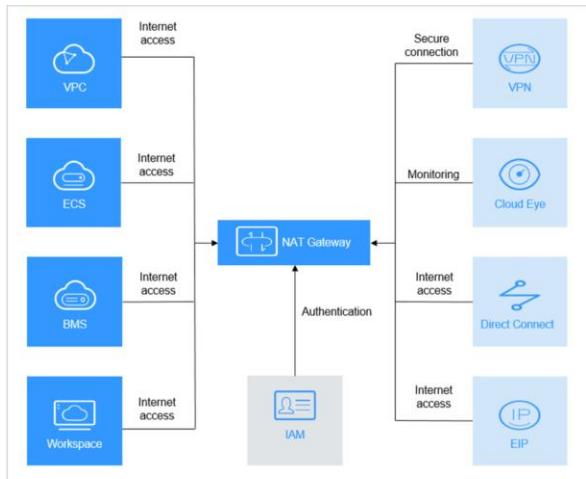
- **Inside Port:**

- This is the port of the server that provides services accessible from the Internet using the DNAT rule. This parameter is available if you select **Specific port** for **Port Type**. The value ranges from 1 to 65535.

- **Port Type:**

- **Specific port:** The NAT gateway forwards requests to your servers only from the outside port and to the inside port configured here, and only if they use the right protocol.
- **All ports:** This is effectively like having a regular EIP bound to your servers. Any requests received by the gateway will be forwarded to your servers, regardless of what port or protocol was used.

## NAT Gateway and Related Services



68      Huawei Confidential



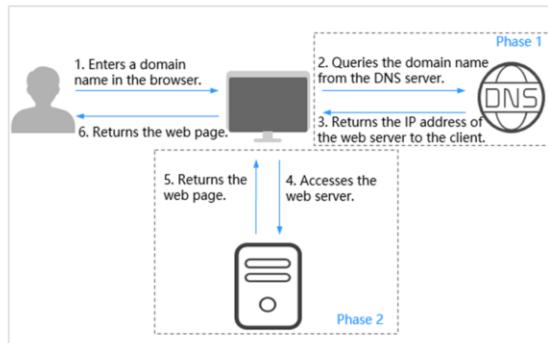
- On-premises servers that need to access the Internet or provide services accessible from the Internet using a NAT gateway can connect to a VPC using Direct Connect.
- On-premises servers that need to access the Internet or provide services accessible from the Internet using a NAT gateway can connect to a VPC through VPN connections.
- A NAT gateway enables cloud servers (ECSs and BMSs) to access the Internet or provide services that are accessible from the Internet.
- ECSs in a VPC can connect to the Internet.
- Servers in a VPC can use an EIP to access the Internet or provide Internet-accessible services through a NAT gateway.
- You can view the monitoring data of NAT gateway resources on the Cloud Eye console.
- If you need to assign different permissions to employees in your enterprise to access your NAT Gateway resources, IAM is a good choice for fine-grained permissions management.

# Contents

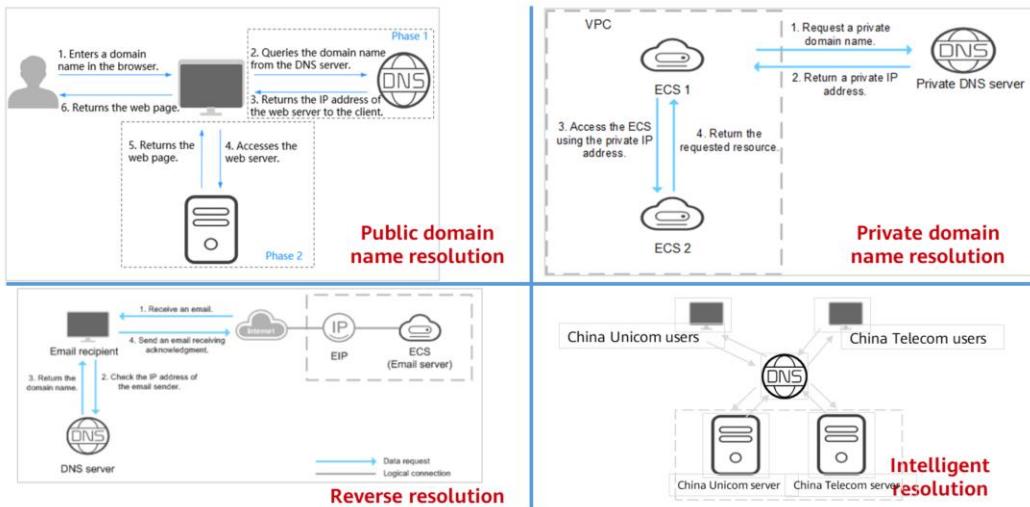
1. Virtual Private Cloud (VPC)
2. Elastic Load Balance (ELB)
3. Virtual Private Network (VPN)
4. NAT Gateway
- 5. Other Services**

## What Is DNS?

- Domain Name Service (DNS) provides highly available and scalable authoritative DNS services that translate domain names into IP addresses required for network connection, reliably directing end users to your applications.



## DNS Resolution Services



71 Huawei Confidential



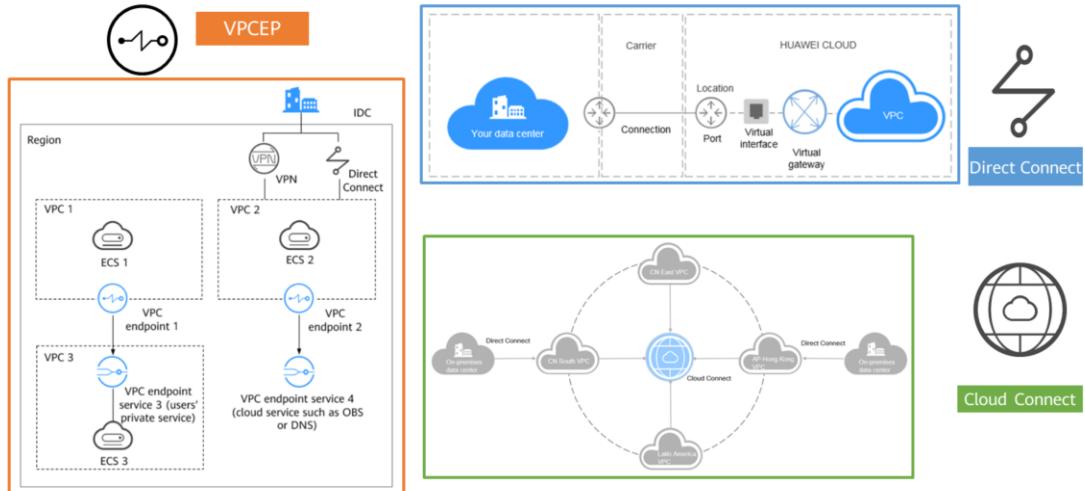
- **Public domain name resolution:** DNS translates domain names like `www.example.com` to public IP addresses like `1.2.3.4`, so that users can access your website or web application over the Internet by entering your domain name in the address box of their browser.
- **Private domain name resolution:** DNS translates domain names like `ecs.com` to private IP addresses like `192.168.1.1` that are used in associated VPCs. With private domain names, your ECSs can communicate with each other within the VPCs without having to connect to the Internet. You can also access cloud services, such as OBS and SMN, over a private network.
- **Reverse resolution:** DNS obtains a domain name based on an IP address. Reverse resolution, or reverse DNS lookup, is typically used to affirm the credibility of email servers. After a recipient server receives an email, it checks whether the IP address and domain name of the sender server are trustworthy and determines whether the email is spam. If the recipient server cannot obtain the domain name mapped to the IP address of the sender server, it concludes that the email was sent by a malicious host and rejects it. It is necessary to configure pointer records (PTR) to point the IP addresses of your email servers to domain names. If no PTR records are configured, the recipient server will treat emails from the email server as spam or malicious and discard them. If you want to build an email server, it is necessary to configure a PTR record to map the email server's IP address to your domain name.
- **Intelligent resolution:** DNS allows you to configure resolution lines. With these resolution lines, you can specify the DNS server that returns different resolution results for the same domain name based on the networks or geographic locations of visitors' IP addresses. For example, if the visitor is a China Unicom user, the DNS server will return an IP address of China Unicom. With this function, you can improve DNS resolution efficiency and speed up cross-network access. You can also create more fine-grained resolution lines based on source IP addresses.

# Domain Name Format and DNS Hierarchy

- **A valid domain name meets the following requirements:**
  - A domain name is segmented using periods (.) into multiple labels.
  - A label can contain supported language-specific characters, letters, digits, and hyphens (-) and cannot start or end with a hyphen.
  - A label cannot exceed 63 characters.
  - The total length of a domain name, including the period at the end, cannot exceed 254 characters.
- **A domain name is divided into the following levels based on its structure:**
  - Root domain: . (a dot)
  - Top-level domain: for example, .com, .net, .org, and .cn
  - Second-level domain: subdomain names of the top-level domain names, such as example.com, example.net, and example.org
  - Third-level domain: subdomain names of the second-level domain names, such as abc.example.com, abc.example.net, and abc.example.org

- DNS allows you to create next-level subdomains within a zone, for example, abc.example.com in example.com and abc.example.com.cn in example.com.cn. However, you cannot expand the domain name further to create subdomains like def.abc.example.com and def.abc.example.com.cn.

## Other Network Services



73      Huawei Confidential

HUAWEI

- VPCEP enables you to access HUAWEI CLOUD services or your own private services securely. It provides flexible networking without having to use EIPs.
- Direct Connect allows you to establish a dedicated network connection that features high speed, low latency, stability, and security between your on-premises data center and the cloud. Direct Connect allows you to maximize legacy IT facilities and leverage cloud services to build a flexible, scalable hybrid cloud computing environment.
- Cloud Connect allows you to connect the VPCs to build a globally connected cloud network with enterprise-grade scalability and communication capabilities.

## Quiz

1. (Single choice) Which of the following is not a component of ELB?
  - A. Backend server group
  - B. Listener
  - C. Load balancer
  - D. NAT Gateway
2. (Single choice) Can resources in a subnet of one VPC communicate with those in a subnet of another VPC in the same region?
  - A. Yes
  - B. No
  - C. Yes, they can communicate with each other by default
  - D. Yes, but VPN is required

- D
- A

## Summary

- This chapter described basic network knowledge and common network cloud services. After completing this course, you will be able to understand the functions of networks as well as how network cloud services work and where you can use these services. For example, a VPC is like the internal network used by an enterprise, and applications can provide Internet-accessible services using EIPs. Mastering these concepts can help you better prepare for cloud migration of legacy systems.

# Recommendations

- Huawei Learning
  - <https://e.huawei.com/en/talent/#/>
- HUAWEI CLOUD technical support
  - <https://support.huaweicloud.com/intl/en-us/help-novice.html>
- HUAWEI CLOUD Academy
  - <https://edu.huaweicloud.com/intl/en-us/>

## Acronyms and Abbreviations

- ACL: access control list
- AS: autonomous system
- BGP: Border Gateway Protocol
- CC: Cloud Connect
- DHCP: Dynamic Host Configuration Protocol
- DNAT: destination network address translation
- DNS: Domain Name System/Domain Name Service
- ECS: Elastic Cloud Server

## Acronyms and Abbreviations

- EIP: Elastic IP
- ELB: Elastic Load Balance
- HTTP: Hypertext Transfer Protocol
- HTTPS: Hypertext Transfer Protocol Secure
- ICT: information and communications technology
- IDC: Internet data center
- IPsec: IP security
- NAT: network address translation

## Acronyms and Abbreviations

- SNAT: source network address translation
- TCP: Transmission Control Protocol
- UDP: User Datagram Protocol
- VPC: Virtual Private Cloud
- VPCEP: VPC Endpoint
- VPN: Virtual Private Network
- Web: World Wide Web (WWW)

# Thank you.

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

Copyright©2021 Huawei Technologies Co., Ltd.  
All Rights Reserved.

The information in this document may contain predictive  
statements including, without limitation, statements regarding  
the future financial and operating results, future product  
portfolio, new technology, etc. There are a number of factors that  
could cause actual results and developments to differ materially  
from those expressed or implied in the predictive statements.  
Therefore, such information is provided for reference purpose  
only and constitutes neither an offer nor an acceptance. Huawei  
may change the information at any time without notice.



## Storage Cloud Services



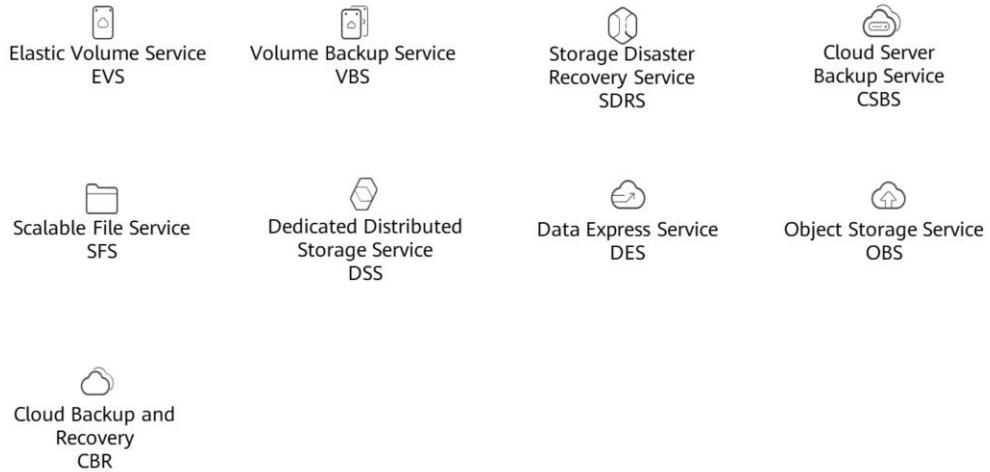
## Foreword

- Data is everywhere. We use USB flash drives and cloud disks to store data, and these devices are called storage devices. That is enough for most of us, but what do you use if you are an enterprise? In today's age of cloud computing, what are the most common storage cloud services?
- In this section, we will cover some common storage services on HUAWEI CLOUD.

# Objectives

- Upon completion of this course, you will:
  - Acquire a basic understanding of cloud storage.
  - Understand the principles behind and uses of common storage services on HUAWEI CLOUD.

# Storage Services Overview



4      Huawei Confidential



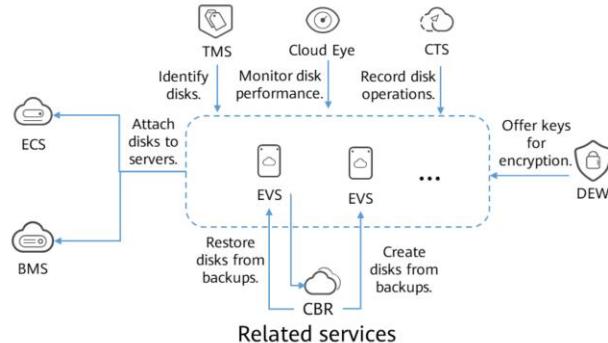
- EVS: persistent block storage for compute services.
- VBS: online backup for EVS disks without the need to stop or restart servers.
- SDRS: protection for cloud servers with zero RPO.
- CSBS: consistent EVS disk backups for ECSs.
- SFS: fully-hosted shared file storage for cloud servers.
- DSS: dedicated, physical storage.
- DES: secure, fast transmission of massive data to HUAWEI CLOUD.
- OBS: stable, secure, and easy-to-use cloud storage
- CBR: backups for cloud servers, disks, and on-premises resources.

# Contents

- 1. Elastic Volume Service**
2. Object Storage Service
3. Scalable File Service

## What Is EVS?

- Elastic Volume Service (EVS) offers scalable block storage for cloud servers. EVS disks offer high reliability and excellent performance. They can be used for distributed file systems, development and testing environments, data warehouse applications, and high-performance computing (HPC).



- EVS disks are like the hard disks on your local computer, except on the cloud. They need to be attached to cloud servers before you can use them. You can initialize EVS disks, create file systems, and then use them for persistent data storage.
- A distributed file system (DFS) is a hierarchical file system, whose physical storage resources may not be directly connected to local nodes, but connected to local nodes through compute networks (compute nodes) or a group of logical partitions or volumes.

## EVS Advantages

Various disk types

- Choose from a range of disk types with different I/O performance specifications.

Real-time monitoring

- With Cloud Eye, you can monitor EVS disk health in real time.

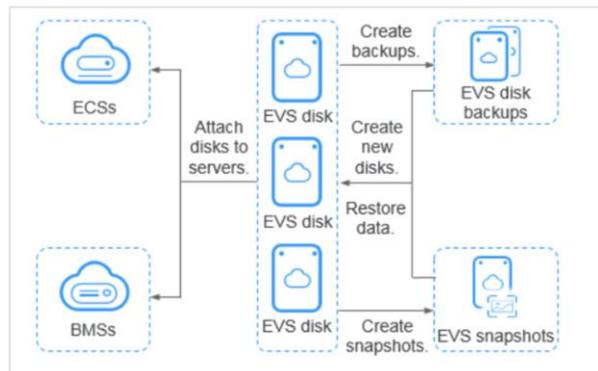
Elastic scalability

- You can expand capacity on-demand and without interrupting services.

High security and reliability

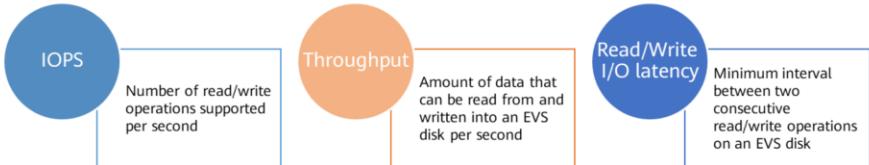
- EVS provides high durability and supports data protection mechanisms including encryption, backup, and snapshot.

## EVS Architecture



- EVS disks are like the hard disks on your local computer, except on the cloud. They need to be attached to cloud servers before you can use them. You can initialize EVS disks, create file systems, and then use them for persistent data storage. Alternatively, you can create backups and snapshots for your EVS disks to improve data reliability.

## EVS Performance Metrics



Parameter	Extreme SSD	Ultra-high I/O	General Purpose SSD	High I/O
Short description	For workloads that demand super-high bandwidth and super-low latency	High-performance disks, excellent for enterprise mission-critical services as well as workloads demanding high throughput and low latency	Cost-effective disks suitable for enterprise office applications	Disks suitable for commonly accessed workloads
Maximum IOPS (for reference)	128,000	50,000	20,000	5,000
Maximum throughput (for reference)	1,000 MB/s	350 MB/s	250 MB/s	150 MB/s
Single-queue access latency (for reference)	200 µs	1 ms	1 ms	1 ms to 3 ms
Typical application scenarios	Databases AI scenarios	Read/write-intensive applications that require ultra-large bandwidth, transcoding services, I/O-intensive scenarios, and latency-sensitive applications	Enterprise office applications, large-scale development and testing, transcoding services, and container system disks	Common office applications, light-load development and testing, and non-system disks

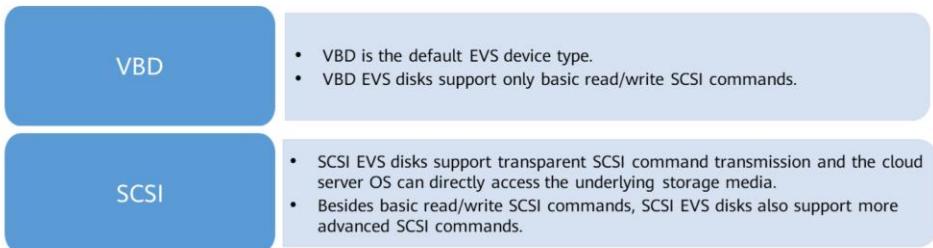
9    Huawei Confidential



- This table lists common EVS disk types. The performance data is for reference only. For the most reliable data, refer to the HUAWEI CLOUD official website.
- EVS disks include as extreme SSD, ultra-high I/O, general purpose SSD, high I/O, or common I/O types, each type offering different performance characteristics. EVS disks differ in performance and price. Choose the disk type most appropriate for your applications.
- Input/Output Operations per Second (IOPS) is the number of read and write operations per second. Disk IOPS = Min. (Maximum IOPS, Baseline IOPS + IOPS per GB x Capacity)

## EVS Device Types

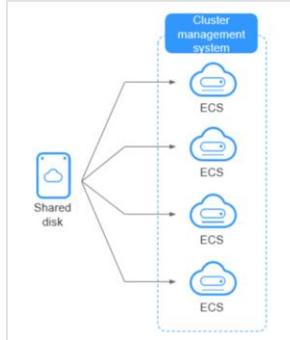
- There are two EVS device types: Virtual Block Device (VBD) and Small Computer System Interface (SCSI).



- SCSI EVS disks: BMSs support only SCSI EVS disks.
- Shared SCSI EVS disks: Shared SCSI EVS disks must be used together with a distributed file system or cluster software. Because most clustered applications, such as Windows MSCS, Veritas VCS, and Veritas CFS, require the usage of SCSI reservations, you are advised to use shared EVS disks with SCSI.

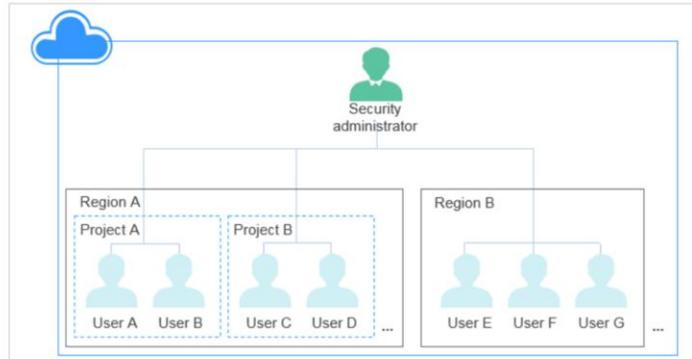
## Shared EVS Disks

- A shared EVS disk can be attached to multiple ECSs or BMSs, and supports concurrent access. Shared EVS disks feature multiple attachments, high-concurrency, high-performance, and high-reliability. They are often used for mission-critical applications that require cluster deployment for high availability (HA).



## EVS Disk Encryption

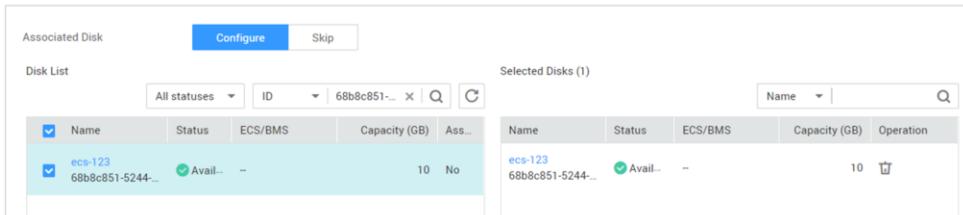
- EVS disks can be encrypted in case your services require extra security.



- The security administrator can grant Key Management Service (KMS) permissions for EVS. These permissions allow you to use disk encryption. The system creates a default master key, which, if you have the permissions, you can use to encrypt EVS disks.

## EVS Disk Backup

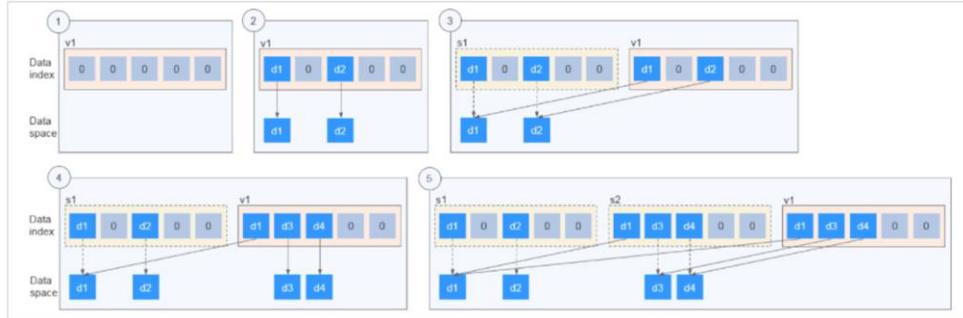
- CBR cloud disk backup allows you to create backups for EVS disks to safeguard the important data on your disks. You can back up EVS disks on the console without stopping the cloud servers. If anything happens to an EVS disk, you can restore the disk data to any point in the past when the backup was created. Disk backups help ensure the integrity and security of your data.



- After a backup policy is applied, the EVS disk data is automatically backed up based on the policy. You can use the backups as the baseline data to create new EVS disks or to restore the backup data to original EVS disks.

## EVS Snapshot

- An EVS snapshot is a complete copy or image of the disk data taken at a specific point in time. They are used for disaster recovery. If anything happens, you can completely restore the disk data to the state from when the snapshot was taken.



- Here we see the process for how snapshots are created over time.
  - EVS disk v1 is created. It has no data.
  - Data d1 and d2 are written to disk v1. Data d1 and d2 are written to new spaces.
  - Now we have snapshot s1 for disk v1. Data d1 and d2 are not saved as another copy elsewhere. Instead, a relationship is established between snapshot s1 and data d1 and d2.
  - Data d3 is written to disk v1 and change data d2 to d4. Data d3 and d4 are written to new spaces, and data d2 is not overwritten. The relationship between snapshot s1 and data d1 and d2 is still valid. Therefore, snapshot s1 can be used to restore data if needed.
  - Snapshot s2 is created for disk v1. A relationship is established between snapshot s2 and data d1, d3, and d4.

## Differences Between EVS Disk Backup and EVS Snapshot

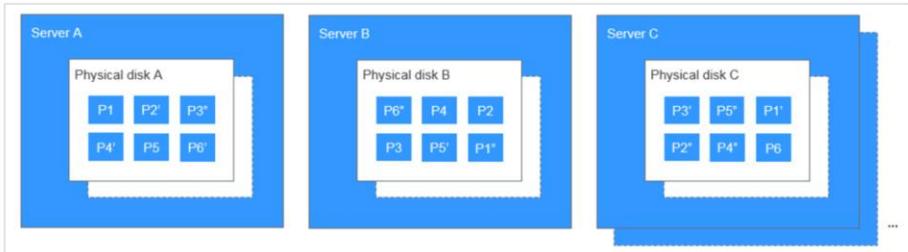
- Both EVS disk backups and snapshots provide redundancy for the EVS disk data. They improve reliability. The following table lists the differences between them.

Item	Data Storage	Data Synchronization	Disaster Recovery Scope	Service Recovery
Backup	Backup data is stored in OBS, instead of EVS disks. Data can be restored even when the EVS disk is damaged.	A backup is a complete copy of a disk at a given point in time. You can also schedule auto backups by configuring backup policies. The backup will not be deleted even if the EVS disk is deleted.	A backup and its source EVS disk have to be in the same AZ. Cloud server backups, in contrast, can be replicated across regions.	You can restore a backup to the original disk or choose to restore it to a new disk. Backups ensure excellent data reliability.
Snapshot	The snapshots are stored with the disk data.	A snapshot is the state of an EVS disk at a specific point in time. When an EVS disk is deleted, its snapshots are also deleted.	The snapshots are in the same AZ as their source EVS disks.	You can use a snapshot to roll back an EVS disk to a previous state, or you can restore the data to a new EVS disk.

- It takes some time to create a backup because data needs to be transferred, but creating or rolling back a snapshot doesn't take as long.

## EVS Three-Copy Redundancy

- The backend storage system of EVS uses three-copy redundancy to guarantee data reliability. Every piece of data is, by default, divided into 1 MB blocks. Three copies of each block are then saved and stored on different nodes in the system based on distribution algorithms. Three-Copy Redundancy has the following features:
  - The storage system automatically distributes three copies of data to different physical disks on different servers. The failure of a single piece of hardware does not affect services.
  - The storage system guarantees strong consistency among data copies.



16      Huawei Confidential



- The storage system ensures strong consistency among copies. For example, the system backs up data block P1 on physical disk A of server A as P1" on physical disk B of server B and P1' on physical disk C of server C. Data blocks P1, P1', and P1" are all copies of the same data block. If physical disk A where P1 resides becomes faulty, P1' and P1" are still available to ensure service continuity.

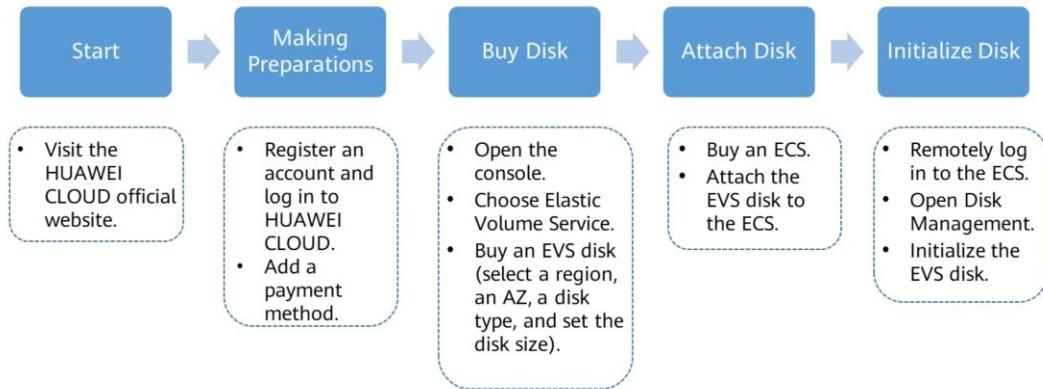
## Data Rebuild

- If a physical server or disk fault is detected, the storage system automatically rebuilds the data. Since data copies are distributed on different storage nodes, data can be rebuilt on different nodes at the same time and each node has only a small amount of data to be rebuilt. This mechanism prevents performance deterioration caused by restoration of a large amount of data on a single node, and therefore minimizes impacts on upper-layer services.



- Each physical disk in the storage system stores multiple data blocks, whose copies are distributed on cluster nodes according to certain rules. If a physical server or disk fault is detected, the storage system automatically rebuilds the data. Since data copies are distributed on different storage nodes, data can be rebuilt on different nodes at the same time and each node has only a small amount of data rebuilt. This system prevents performance deterioration caused by restoration of a large amount of data on a single node, and minimizes impacts on upper-layer services.

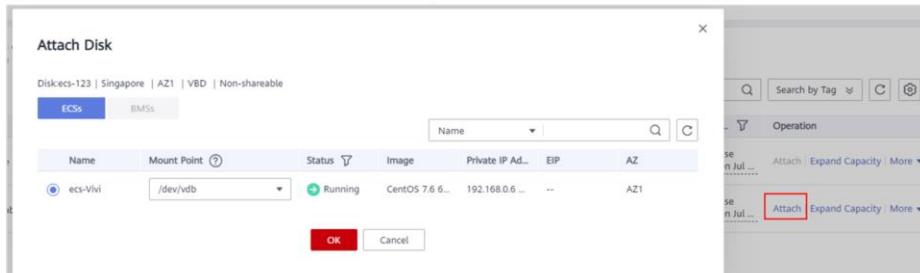
## EVS Configuration Process



- Regions are defined based on geographical location and network latency. Public services, such as Elastic Cloud Server (ECS), Elastic Volume Service (EVS), Object Storage Service (OBS), Virtual Private Cloud (VPC), Elastic IP, and Image Management Service (IMS), are shared within the same cloud region.
- An AZ contains one or more physical data centers. Each AZ has independent cooling, fire extinguishing, moisture-control, and electricity facilities. Within an AZ, computing, network, storage, and other resources are logically divided into multiple clusters. AZs within a region are interconnected using high-speed optical fibers to ensure high availability.

## Attaching an EVS Disk

- An EVS disk cannot be used alone. It can be accessed and used only after being attached to an ECS or BMS and initialized.
  - A non-shared EVS disk can be attached to only one server.
  - A shared EVS disk can be attached to up to 16 servers in the same AZ.



- Separately purchased EVS disks can be used as data disks. In the EVS disk list, the function of such disks is **Data disk**, and their status is **Available**. Data disks need to be attached to servers before you can use them.
- System disks are purchased along with servers and are automatically attached. In the EVS disk list, the function of such disks is **System Disk**, and their status is **In-use**. After a system disk is detached from a server, the disk function changes to **Bootable Disk**, and the disk status changes to **Available**.

## Discussion

- What are the differences between EVS disks and other forms of online storage?
- What are the differences between three-copy redundancy, EVS disk backups, and EVS snapshots?



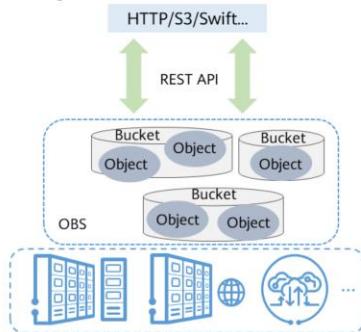
- Three-copy redundancy is provided by the backend storage system of EVS to ensure data reliability. It mainly deals with data losses or inconsistency caused by hardware faults. Whereas, EVS disk backup and EVS snapshot are used to address data loss or inconsistency caused by human error, viruses, or hacker attacks.
- EVS disks cannot be used alone. They must be used together with ECSs or BMSs. Web disks can directly be used to store data. In terms of product category, EVS disks are IaaS products, and online drives are SaaS products.

# Contents

1. Elastic Volume Service
- 2. Object Storage Service**
3. Scalable File Service

## What Is OBS?

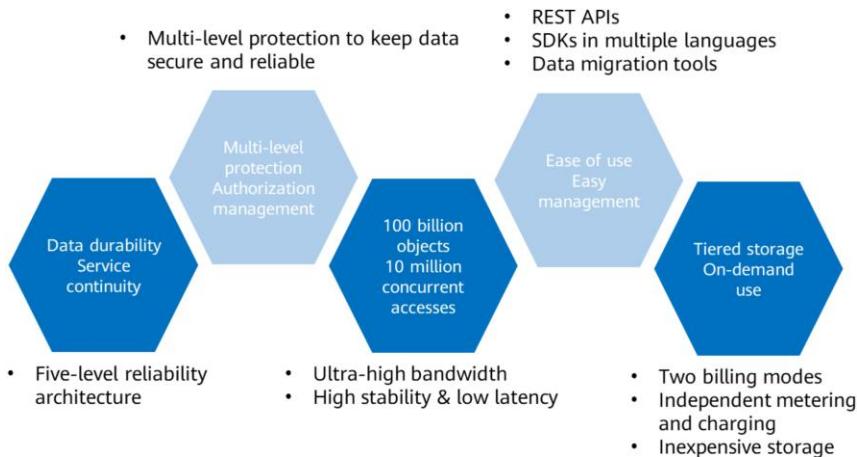
- Object Storage Service (OBS) is a cloud storage service optimized for storing massive amounts of data. It provides unlimited, secure, and highly reliable storage capabilities at a relatively low cost.



- Flat structure, isolated tenant data
- Users can create buckets (like folders) where they can upload objects to or download objects from, and share objects using links.

- OBS provides users with unlimited storage capacity for storing files in any format, catering to the needs of common users, enterprises, and developers. Neither the entire OBS system nor a single bucket limits the storage capacity or the quantity of objects that can be stored.
- As a web service, OBS supports APIs over HTTP and HTTPS. Users can access and manage data stored in OBS anytime, anywhere through OBS Console or using OBS tools. With OBS SDKs and APIs, you can easily manage data stored in OBS and develop upper-layer service applications.

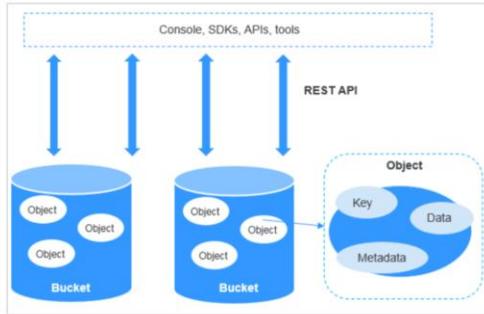
## OBS Advantages



- Here are the advantages of OBS:
  - **Data durability and service continuity:** OBS provides storage for Huawei mobile phone's cloud album to support access from hundreds of millions of users. It delivers a data durability of up to 99.999999999% and service continuity of up to 99.995% by leveraging cross-region replication, cross-AZ disaster recovery, intra-AZ device and data redundancy, slow disk or bad sector detection, and other technologies.
  - **Multi-level protection and authorization management:** OBS keeps your data secure and trusted by using versioning, server-side encryption, URL validation, VPC-based network isolation, log auditing, and fine-grained permission control.
  - **100 billion objects and 10 million concurrent accesses:** With intelligent scheduling and response, OBS optimizes access paths and leverages technologies such as event notification, transfer acceleration, and big data vertical optimization, enabling users to store hundreds of billions of objects in OBS. OBS features ultra-high bandwidth and low latency.
  - **Ease of use and easy management:** OBS offers REST APIs, SDKs in different languages, and data migration tools, making it easier to migrate services to the cloud. It is not necessary to plan for storage capacity beforehand or worry about capacity expansion or reduction, because storage resources are linearly scalable.
  - **Tiered storage and on-demand use:** OBS is billed in pay-per-use and monthly/yearly modes. Data in Standard, Infrequent Access, and Archive storage classes is separately metered and billed, significantly reducing storage costs.

## OBS Architecture

- A bucket is a container for storing objects in OBS. OBS stores all objects in the same logical layer, realizing a flat storage structure instead of a hierarchical one.
- An object is the basic data storage unit in OBS, which is a file and any metadata that describes the file. Data is stored as objects in OBS buckets. An object consists of three parts: key, metadata, and data.



24      Huawei Confidential



- An object is the basic data storage unit in OBS, which is a file and any metadata that describes the file. An object consists of three parts: key, metadata, and data.
  - A key specifies the name of an object. An object key is a UTF-8 string that is between 1 and 1024 characters. Each object in a bucket has a unique key.
  - Metadata describes an object, and is classified into system-defined metadata and user-defined metadata. The metadata is a set of key-value pairs that are assigned to the object stored in OBS. System-defined metadata is automatically assigned by OBS for processing objects. System-defined metadata includes Date, Content-Length, Last-Modified, Content-MD5, and more. User-defined metadata is specified when users upload objects and is used to describe objects.
  - Data refers to the content that an object contains.

## Permanent AK/SK Pair

- OBS supports authentication using a AK/SK pair. It leverages the AK/SK-based encryption to authenticate a request sender. Users can create a permanent AK/SK pair on the My Credentials page.

The screenshot shows the AK Login interface for HUAWEI CLOUD OBS. The 'Service' dropdown is set to 'HUAWEI CLOUD OBS (default)'. The 'Access Key ID' field contains a redacted value. The 'Secret Access Key' field also contains a redacted value. The 'Access Path' field is empty and placeholder text 'Enter an access path (e.g. obs://bucket/folder)' is visible. A checked checkbox labeled 'Remember my access keys.' is located below the 'Access Path' field. A large red rectangular box highlights the 'Log In' button at the bottom of the form.

25      Huawei Confidential



- What is an AK/SK pair?
  - AK refers to an access key ID. SK refers to a secret access key. AK and SK are used together to sign requests cryptographically.
  - The AK and SK authenticate request senders to prevent their request from being modified.

## Temporary AK/SK Pair

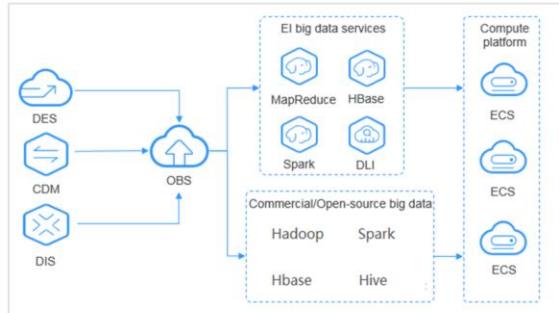
- Temporary AK/SK and security token issued by the system to users are valid for 15 minutes to 24 hours. The temporary AK/SK and security token comply with the least privilege principle and can be used to temporarily access OBS. Error code 403 will be returned if a request does not have a security token.

Status Code	Description
201	The request is successful.
400	Invalid parameters.
401	Authentication failed.
403	Access denied.
500	Internal server error.

- A temporary AK/SK pair works in the same way as an AK/SK pair, but it is temporarily valid and needs to be obtained again after it expires.
- Additionally, a temporary AK/SK pair must be used together with a security token to access all resources of a specified account.

## OBS Application Scenario - Big Data Analytics

- OBS provides inexpensive big data solutions that feature high performance with zero service interruption and eliminate the need for capacity expansion. Such solutions are designed for scenarios involving storage and analysis of massive amounts of data, query of historical data details, analysis of a large number of behavior logs, analysis of public transactions, and statistics collection.



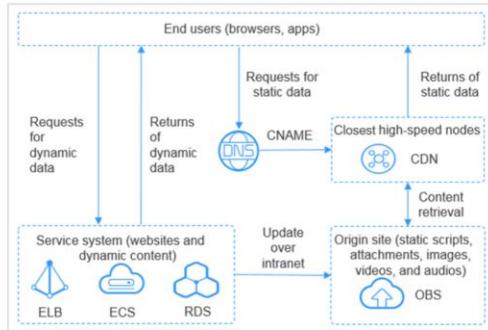
27      Huawei Confidential



- The big data analytics scenarios are classified into the following:
- **Storage and analysis of massive amounts of data:** storage for petabytes of data, batch data analysis, and response for queries of data details in seconds
- **Query of historical data details:** account statement audit, analysis on device energy consumption history, playback of trails, analysis on driving behavior, and refined monitoring
- **Analysis of massive amounts of behavior logs:** analysis and query of learning habits, operation logs, and system operation logs
- **Statistical analysis of public transactions:** crime tracking, associated case queries, traffic congestion analysis, and scenic spot popularity statistics
- In these scenarios, the services recommended for use with OBS include MapReduce Service (MRS), Elastic Cloud Server (ECS), and Data Express Service (DES).

## OBS Application Scenario - Static Website Hosting

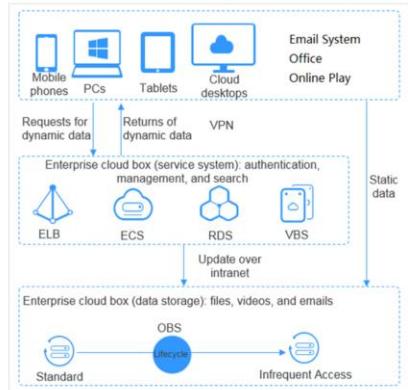
- Dynamic data on end users' browsers and apps directly interacts with the service system on HUAWEI CLOUD. Requests for dynamic data are sent to the service system for processing and then returned to end users. The static data is stored in OBS. The service system processes static data over the intranet, and end users can directly request and read the static data from OBS through the closest high-speed nodes.



- OBS provides a website hosting function that is cost-effective, highly available, and automatically scalable according to traffic volume. With this function and CDN and ECS, you can quickly build a website or an application system with static and dynamic content separated.

## OBS Application Scenario - Enterprise Cloud Box

- OBS offers a highly reliable, inexpensive storage system featuring high concurrency and low latency. Users can scale the storage capacity as their data volume grows.



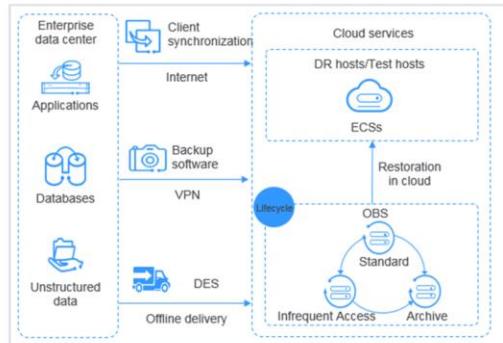
29      Huawei Confidential



- Dynamic data on user's devices such as mobile phones, PCs, and tablets interacts with the enterprise cloud box system on HUAWEI CLOUD. Requests for dynamic data are sent to the service system for processing and then returned to devices. The static data is stored in OBS. Service systems process static data over the intranet and end users can directly request and read static data from OBS. Additionally, OBS allows you to configure lifecycle rules to automatically transition storage classes for objects, reducing storage costs.
- In this scenario, you can use OBS together with ECS, ELB, RDS, and VBS.

## OBS Application Scenario - Backup and Archive

- Users can use the synchronization client, mainstream backup software, or DES to back up your on-premises data and store the data in OBS. Data can be restored from OBS to the DR host or test host on the cloud if necessary.

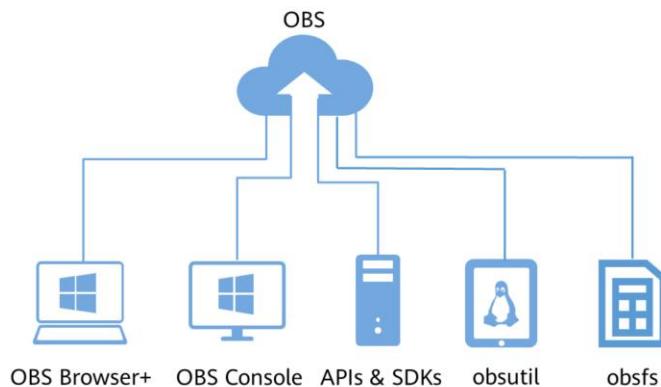


30 Huawei Confidential



- OBS offers a highly reliable, inexpensive storage system featuring high concurrency and low latency. It stores massive amounts of data, meeting the archive requirements for unstructured data of apps and databases.
- In this scenario, you are advised to use OBS together with DES and ECS.

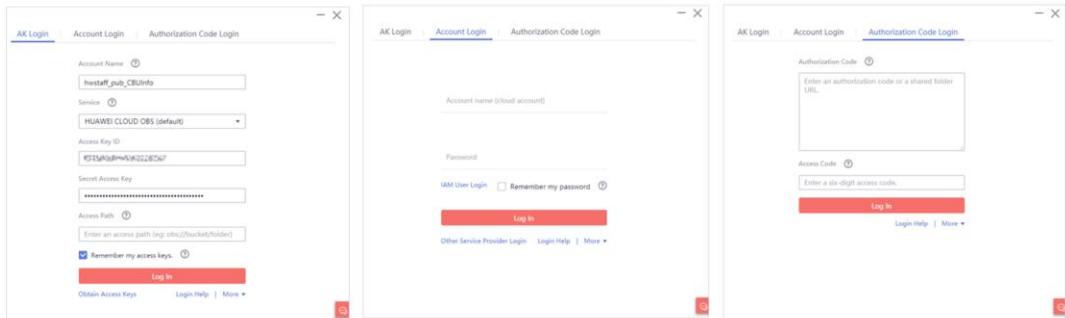
## Accessing OBS



- When users log in to OBS Console using their HUAWEI CLOUD account or as an IAM user, OBS authenticates their account or IAM user credentials. When users access OBS using the tools (OBS Browser+ or obsutil), SDKs, or APIs, OBS requires access keys (AK and SK) for authentication. Therefore, users need to obtain the access keys (AK and SK) before they access OBS using any methods other than OBS Console.
- obsutil is a command line tool for accessing OBS. Users can use this tool to perform basic operations in OBS, such as creating buckets, as well as uploading, downloading, and deleting files or folders. This tool is ideal for batch processing and automated tasks.
- obsfs, built on Filesystem in Userspace (FUSE), is a file system tool provided by OBS for mounting parallel file systems to Linux. It enables you to easily access the infinite storage space on OBS in the same way as you operate a local file system.

## Accessing OBS Using OBS Browser+

- Install OBS Browser+.
  - Download OBS Browser+ from the corresponding download link.
- Log in to OBS Browser+.
  - Access keys (AK/SK), an account, and an authorization code can be used to log in to OBS Browser+.



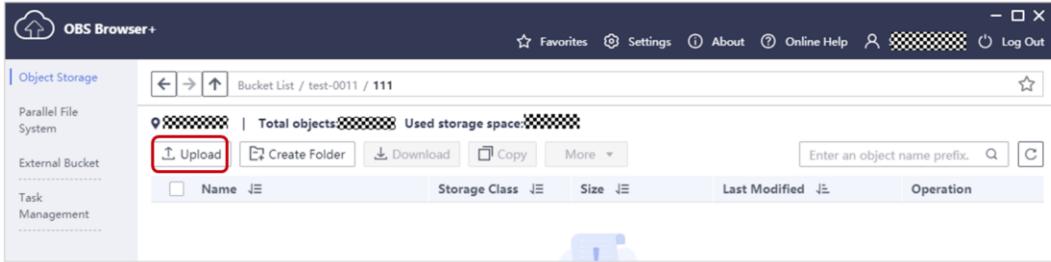
32      Huawei Confidential



- There are some limitations for OBS Browser+:
  - Keeps the login details for a maximum of 100 accounts.
  - Does not support the query or deletion of historically authorized login information.
  - Automatically deletes expired authorization codes.
  - If a proxy is required to access your network environment, configure the network proxy before login.

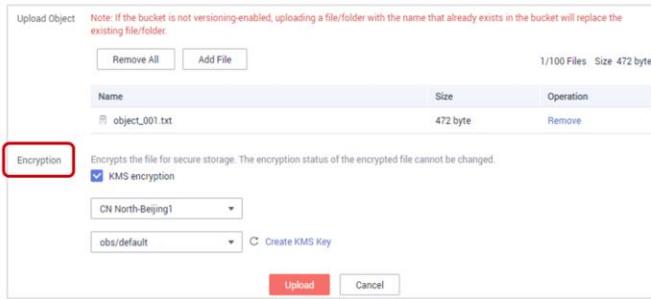
## Accessing OBS Using OBS Browser+

- Upload objects by drag and drop.
- OBS Browser+ provides a powerful drag-and-drop function. Users can drag one or more local files or folders to the object list of a bucket or parallel file system.



## Server-Side Encryption

- With server-side encryption enabled:
  - OBS encrypts your object before saving it on the server.
  - OBS decrypts the object on the server before it is downloaded.



- KMS uses a third-party hardware security module (HSM) to protect keys, enabling users to create and manage encryption keys easily. For security reasons, keys are not displayed in plaintext outside HSMs. With KMS, all operations on keys are controlled and logged, and usage records of all keys can be provided to meet regulatory compliance requirements.
- The objects to be uploaded can be encrypted on the server side using the KMS-provided keys. First, create a key using KMS or use the default key provided by KMS. Then, this selected key will be used to encrypt your objects on the server side before they are uploaded to OBS.
- OBS supports both server-side encryption with KMS-managed keys (SSE-KMS) and server-side encryption with customer-provided keys (SSE-C) by calling APIs. In SSE-C mode, OBS uses the keys and MD5 values provided by customers for server-side encryption.

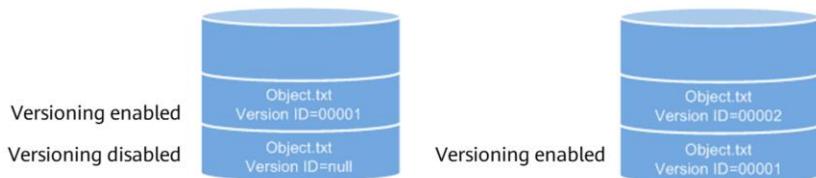
## URL Validation

- Some bad sites may steal links from other sites to enrich their content without any costs. This not only damages the interests of original websites but also increases workloads for the website servers. To avoid link stealing, URL validation is created. OBS also supports whitelist and blacklist settings.
- If Whitelisted Referers is left blank but Blacklisted Referers is not, all websites except those specified in the blacklist are allowed to access the target bucket.
- If Whitelisted Referers is not left blank, only the websites specified in the whitelist are allowed to access the target bucket no matter whether Blacklisted Referers is left blank or not.

- In HTTP, the Referer field is used for a website to detect the web page that accesses a target page. As the Referer field tracks sources, access requests will be blocked or specific pages will be returned if requests are detected not from trusted sources. URL validation checks whether the Referer field in requests matches that configured in the whitelist or blacklist. If they match, the requests are allowed. If they do not match, the requests are blocked or specific pages are returned.

## OBS Versioning

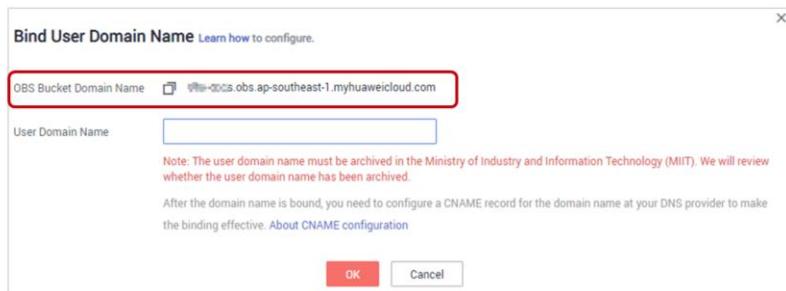
- With versioning enabled, OBS stores multiple versions of an object so that users can quickly search for and restore different object versions or restore data in the event of failures.



- By default, versioning is disabled for buckets. With versioning disabled, if a new object you upload has the same name as the one already stored in the bucket, the new object will overwrite the existing one.
- After you enable versioning for the bucket, the version IDs (null) and contents of the objects that are already in the bucket remain the same. If you upload the new object again, the object versions are shown as the figure on the left. OBS automatically allocates a unique version ID to each new object. Objects with the same name are stored in OBS with different version IDs, as shown in the figure on the right.

## User-Defined Domain Names

- If users want to migrate files from a website to OBS and still want to use their own website link for accessing files stored in OBS, they can bind a user-defined domain name to an OBS bucket.

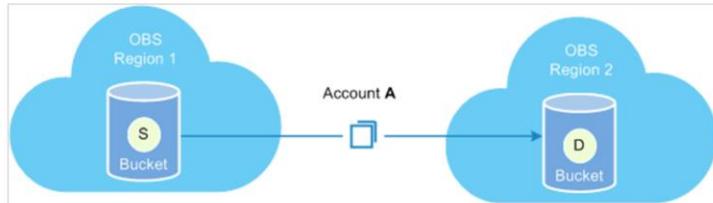


- For example, if the domain name of user A's website is **www.example.com** and the website file is **abc.html**, then the link for accessing **abc.html** on the website is **http://www.example.com/abc.html**. To use **http://www.example.com/abc.html** to access the file stored in OBS, user A can perform the following operations:
  - Create a bucket on OBS, and upload **abc.html** to the bucket.
  - On OBS Console, bind the domain name **www.example.com** to the created bucket.
  - On the DNS server, add a CNAME rule and map **www.example.com** to the bucket's domain name.
  - When requests for accessing **http://www.example.com/abc.html** reach OBS, OBS locates the mapping between **www.example.com** and the bucket domain name, and directs the requests to file **abc.html** in the bucket. That is, a request for **http://www.example.com/abc.html** actually accesses **http://Bucket domain name/abc.html**.
- I will talk about the service types when I explain CDN in detail.
- Here are some constraints for user-defined domain names:
  - Only OBS 3.0 supports the binding of user-defined domain names. The bucket version can be viewed in the **Basic Information** area of the bucket's **Overview** page on OBS Console.
  - A bucket can have a maximum of 5 user-defined domain names bound.
  - User-defined domain names currently allow access requests over only HTTP.

- If you want to use a bound domain name to access OBS over HTTPS, you need to enable CDN to manage HTTPS certificates.

## Cross-Region Replication

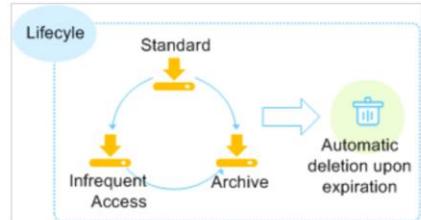
- Cross-region replication provides data disaster recovery across regions, catering to your needs for off-site data backup.
- With replication rules configured, OBS automatically and asynchronously replicates data from a source bucket to a destination bucket in another region.



- With cross-region replication enabled, OBS will replicate the following objects to a destination bucket:
  - Newly uploaded objects (excluding those in the Archive storage class)
  - Updated objects, for example, the object content is updated or the copied ACL is updated
  - Historical objects (To replicate historical objects, **Synchronize Existing Objects** must be enabled.)
- Here list the scenarios where cross-region replication will be used:
  - Cross-region access: Users need to access the same OBS resource in different locations. To minimize the access latency, users can use cross-region replication to create object copies in their nearest region.
  - Data migration: Users need to migrate data stored in OBS from the data center in one region to that in another region.
  - Data backup: To keep data secure and available, users can create explicit backups for all data written to OBS in the data center of another region. This allows users to restore data through the backups if the source data is irreversibly damaged.

## Lifecycle Management

- Users can configure lifecycle rules to periodically delete objects or transition object storage classes.
- Lifecycle rules apply to the following scenarios:
  - Periodically deleting files that are only meant to be retained for specified periods of time
  - Transitioning documents that are seldom accessed to the Infrequent Access or Archive storage class or deleting them



- Lifecycle management rules have two key elements:
  - **Policy:** Users can configure a lifecycle rule and apply it to a subset of objects with a specific prefix, or apply it to the entire bucket (that is, apply it to all the objects in the bucket).
  - **Time:** Users can use a lifecycle rule to specify the number of days after which objects meeting the conditions are automatically transitioned to the Infrequent Access or Archive storage class, or are automatically deleted upon expiration.
    - **Transition to Infrequent Access:** This rule transitions the objects meeting the conditions to the Infrequent Access storage class after the specified number of days since the last object update.
    - **Transition to Archive:** This rule transitions the objects meeting the conditions to the Archive storage class after the specified number of days since the last object update.
    - **Delete upon expiration:** This rule automatically deletes the objects meeting the conditions after the specified number of days since the last object update.

## Permission Management

- OBS leverages the following two methods for access control:

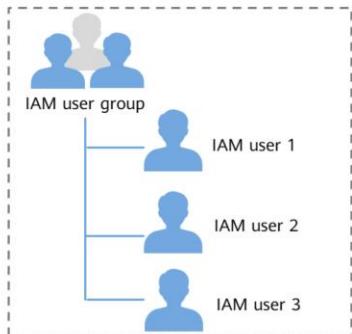


- |  |  |
|--|--|
| <ul style="list-style-type: none"><li>• Apply to cloud resources.</li><li>• Define the actions that are allowed or denied.</li></ul> | <ul style="list-style-type: none"><li>• Apply to a bucket and objects in it.</li><li>• Apply to the objects in a bucket.</li></ul> |
|--|--|

- IAM permissions define the operations that are allowed or denied on your cloud, controlling access to your resources.
- Let's move on to bucket policies and object policies.
  - Bucket policies apply to buckets and objects in them. A bucket owner can use bucket policies to grant IAM users or other accounts the permissions required to operate the bucket and objects in the bucket.
  - Object policies apply to the objects in a bucket.
  - An ACL defines which accounts are granted access and the type of access. You can configure an ACL for a bucket or an object.

## Access Control - IAM Permissions

- Users can create IAM users under a registered cloud service account, and then use IAM permissions to control their access to cloud resources.

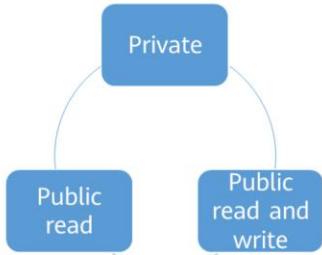


Permission	Description
Tenant Administrator	Allows users to perform any operations on OBS resources.
Tenant Guest	Allows users to query the usage of OBS resources.
OBS Buckets Viewer	Allows users to obtain the bucket list, metadata, and location information.

- IAM permissions define the operations that are allowed or denied on your cloud, controlling access to your resources. The IAM permissions configured for OBS apply to all buckets and objects. To grant an IAM user the permissions required to operate OBS resources, you need to assign one or more OBS permission sets to the user group to which the IAM user belongs.
- IAM permissions are mainly used to authorize IAM users under the same account to,
  - Control access to all cloud resources under the account
  - Control access to all OBS buckets and objects under the account
  - Control access to specified OBS resources under the account

## Access Control - Bucket Policies

- Bucket policies apply to buckets and objects in them. A bucket owner can use bucket policies to grant IAM users or other accounts the permissions required to operate the bucket and objects in the bucket. There are three options for standard bucket policies.



### Standard bucket policies:

- Private:** Only users granted permissions by the ACL can access a bucket.
- Public read:** Anyone can read objects in a bucket.
- Public read and write:** Anyone can read, write, or delete objects in a bucket.

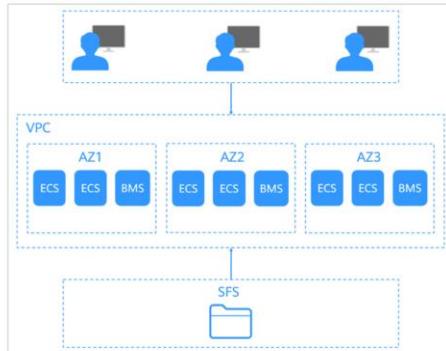
- You may configure bucket policies in the following scenarios:
  - No IAM permissions are configured for access control and you want other accounts to access your OBS resources
  - To grant IAM users different permissions required to access different buckets
  - To grant other accounts the permissions to access your buckets

# Contents

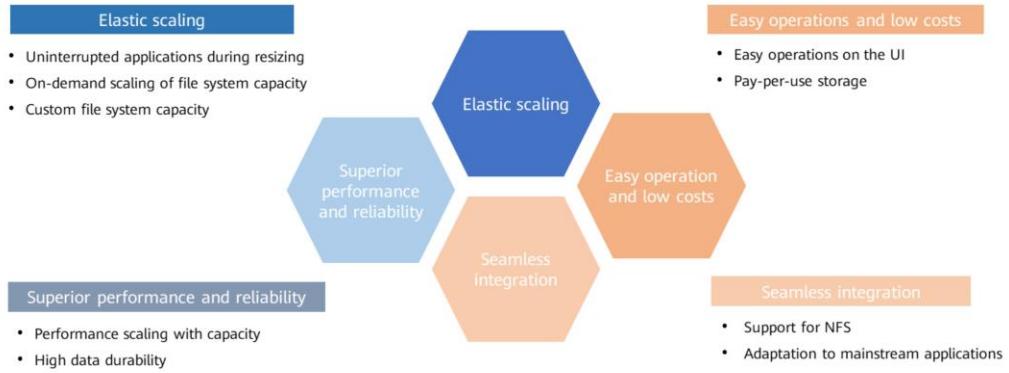
1. Elastic Volume Service
2. Object Storage Service
- 3. Scalable File Service**

## What Is SFS?

- Scalable File Service (SFS) provides reliable, high-performance shared file storage hosted on HUAWEI CLOUD. With SFS, you can enjoy shared file access spanning multiple ECSs, BMSs, and containers.



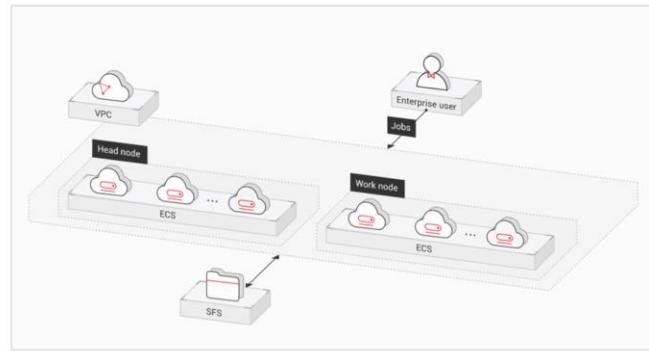
## SFS Advantages



- Compared with traditional file sharing storage, SFS has the following advantages:
  - **Elastic scalability:** Storage can be scaled up or down on demand to dynamically adapt to service changes without interrupting applications. Resizing can be done with a few clicks.
  - **High performance and reliability:** SFS enables file system performance to increase as capacity grows, and delivers a high data durability to support rapid service growth.
  - **Seamless integration:** SFS supports NFS and CIFS protocols. With standard protocols, a broad range of mainstream applications can read and write data in the file system. In addition, SFS is compatible with SMB 2.0, SMB 2.1, and SMB 3.0, so Windows clients can access the shared space.
  - **Simple operation and low cost:** You can create and manage file systems with ease in a GUI. SFS slashes the cost as it is charged on a pay-per-use basis.

## Application Scenario - High Performance Computing

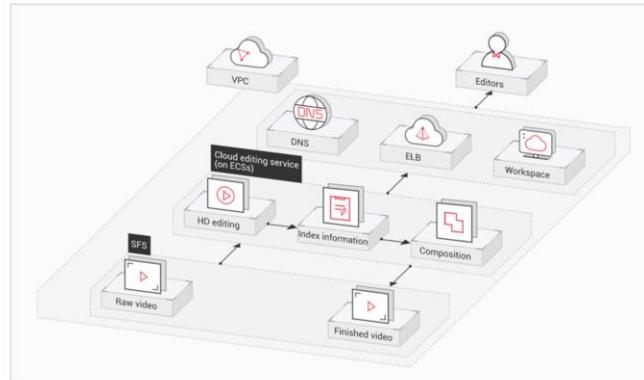
- Shared file storage facilitates industry design, biomedicine, energy exploration, graphic rendering, and heterogeneous computing.



- HPC is short for high performance computing. An HPC system or environment is made up of a single computer system with many CPUs, or a cluster of multiple computer clusters. It can handle a large amount of data and handle computing tasks that would be difficult for regular PCs. HPC has ultra-high capabilities for floating-point computation and can be used for compute-intensive and data-intensive fields, such as industrial design, bioscience, energy exploration, image rendering, and heterogeneous computing.
- Industrial design: In automobile manufacturing, CAE and CAD simulation software are widely used. When the software is operating, compute nodes need to communicate with each other closely, which requires a file system with high bandwidth and low latency.
- Bioscience: The file system should have high bandwidth, lots of storage capacity, and be easy to expand.
  - Bioinformatics: To sequence, stitch, and compare genes.
  - Molecular dynamics: To simulate the changes of proteins at molecular and atomic levels.
  - New drug R&D: To complete high-throughput screening (HTS) to shorten the R&D cycle and reduce the investment required.

## Application Scenario - Media Processing

- Shared file storage facilitates multi-layer HD and 4K video editing, transcoding, composition, and video on demand (VoD).



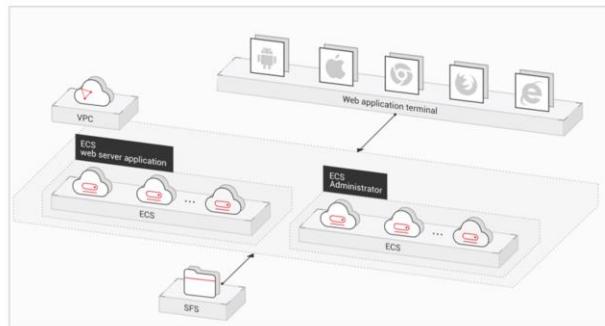
50      Huawei Confidential



- Media processing involves uploading, downloading, cataloging, transcoding, and archiving media materials, as well as storing, invoking, and managing audio and video data. Media processing involves very large video files with high bit rates, so the file systems need to be very large and easily expanded.
- Acquisition, editing, and synthesis of audio and video data require stable, low-latency file systems.
- Concurrent editing requires file systems that can let users share data easily and reliably.
- Video rendering and special effects need processing small files frequently. The file systems must offer high I/O performance.
- SFS is a shared storage service that works like a regular file system. It features high-speed data sharing, dynamic storage tiering, as well as on-demand, smooth, and online resizing. These outstanding features empower SFS to meet the demanding requirements of media processing for robust capacity, throughput, and IOPS, with minimal latency.

## Application Scenario - Content Manage and Web Services

- SFS provides secure storage for content management systems and web applications, facilitating quick online publishing and archiving. Burst traffic at peak hours can also be properly handled, which frees you up from worrying about capacity expansion.



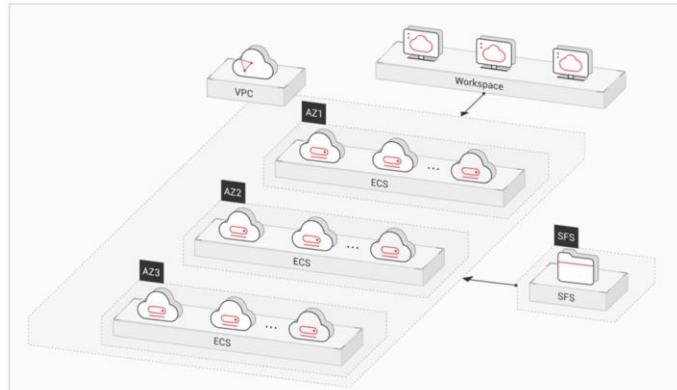
51      Huawei Confidential



- For I/O-intensive website services, SFS Turbo can provide shared website source code directories and storage for multiple web servers, enabling low-latency and high-IOPS concurrent shared access. SFS Turbo file systems are good when dealing with:
- A large number of small files, for instance static websites, where there are many small HTML, JSON, and static image files.
- Read intensive tasks. When there is a lot of existing data but not a lot of write throughput required, SFS Turbo is a good choice.
- Multiple web servers. SFS Turbo is a good way to help ensure high availability for website services.

## Application Scenario - File Sharing

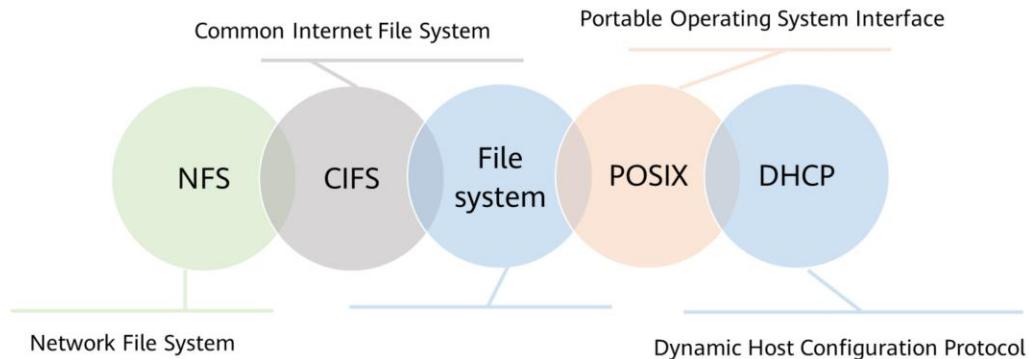
- For companies with a large number of departments and employees, documents and data can be shared and accessed company-wide.



52      Huawei Confidential



## Concepts Related to SFS



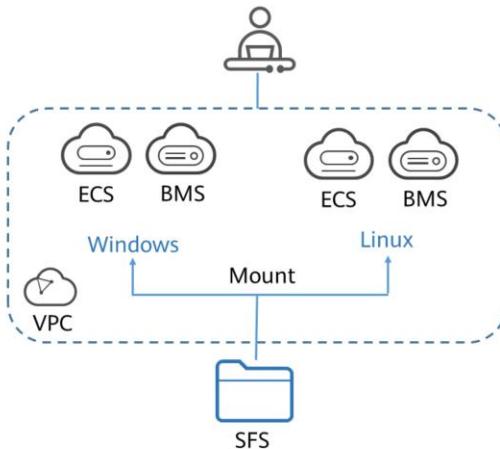
- Concepts related to SFS:
  - Network File System (NFS) is a distributed file system protocol that allows different computers and operating systems to share data over a network.
  - Common Internet File System (CIFS) is a protocol used for network file access. It is a public or open version of the Server Message Block (SMB) protocol, from Microsoft. CIFS allows applications to access files on computers over the Internet and send requests for file services. Using the CIFS protocol, network files can be shared easily between Windows hosts.
  - A file system provides users with shared file storage service through NFS and CIFS. It is used for accessing network files remotely. After a user creates a mount point on the management console, the file system can be mounted to multiple ECSs and is accessible through the standard POSIX.
  - Portable Operating System Interface (POSIX) is a set of interrelated standards specified by Institute of Electrical and Electronics Engineers (IEEE) to define the application programming interface (API) for software compatible with variants of the UNIX operating system. POSIX is intended to achieve software portability at the source code level. That is, a program written for a POSIX compatible operating system may be compiled and executed on any other POSIX operating system.
  - Dynamic Host Configuration Protocol (DHCP) is a LAN network protocol. The server controls an IP address range, and a client can automatically obtain the IP address and subnet mask allocated by the server when logging in to the server. DHCP is not installed as a service component of Windows Server by default. Manual installation and configuration are required.

## SFS Configuration Process

3. Mount the file system to Linux or Windows servers.

2. Ensure that the file system and servers are in the same VPC.

1. Create a file system.



- Create a file system and mount it to multiple ECSs. Then the ECSs can share the file system. There are two types of file systems: SFS Capacity-Oriented and SFS Turbo.

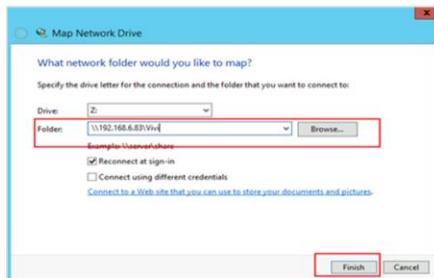
## Using SFS - Mounting an NFS FS to a Linux Server

- After a file system has been created, mount the file system to servers so that the servers can share the file system. In the following example, user root is used to log in to the servers.
- Install support for NFS.
- Run the following command to check whether the file system domain name can be resolved: (SFS Turbo file systems do not require domain name resolution. You can skip this step and directly mount the file system.)
  - nslookup file-system-domain-name
- Run the following command to create a local path for mounting the file system:
  - mkdir /mount-point
- Run the following command to mount the file system to the server. Currently, the file system can be mounted to Linux servers using NFS v3 only.
  - mount -t nfs -o vers=3 timeo=600 file-system-domain-name mount-point
  - After the file system is mounted, check that you can access the file system on the server.

- In the Linux command line:
  - nslookup is used to resolve domain names.
  - mkdir creates a directory. We will use it to create a mount point for the filesystem.
  - mount mounts the filesystem. The -t argument is to specify the type of the file system, in this example, nfs. The -o argument is to set the protocol version and configure a timeout interval, in this example, v3 and 600s.

## Using SFS - Mounting a CIFS FS to a Windows Server

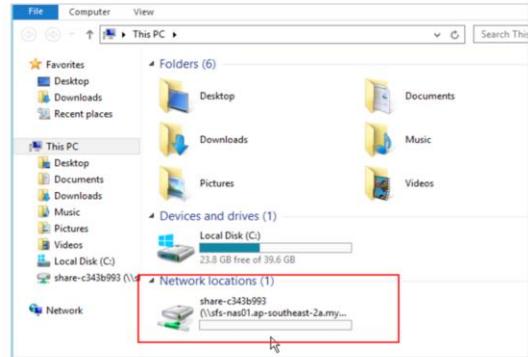
- In the following example, a CIFS file system is mounted to a Windows ECS:
  - Log in to an ECS running Windows Server 2012.
    - Click Start, right-click Computer, and choose Map network drive.
    - In the dialog box that is displayed, enter the file system mount address, specifically, \\file-system-domain-name\path.



- Prerequisites:
  - Check the type of the ECS operating system. Different operating systems require different commands for NFS client installation.
  - A file system has been created, and its mount address has been obtained.
  - At least one ECS that belongs to the same VPC as the file system exists.
  - The IP address of the DNS server used for resolving the file system domain name has been configured on the ECS. SFS Turbo file systems do not require domain name resolution.

## Using SFS - Unmounting a File System

- If a file system is no longer used and needs to be deleted, unmount the file system and then delete it.
- Linux:
  - Log in to the ECS.
  - Run `umount /mount-point`
- Windows:
  - Log in to the ECS.
  - Right-click the file system to be unmounted and choose Disconnect.
  - The file system has been unmounted when it disappears from the network location.
  - File system unmounted



- Prerequisites:
  - Check the type of the ECS operating system. Different operating systems require different commands for NFS client installation.
  - A file system has been created, and its mount address has been obtained.
  - At least one ECS that belongs to the same VPC as the file system exists.
  - The IP address of the DNS server used for resolving the file system domain name has been configured on the ECS. SFS Turbo file systems do not require domain name resolution.

## Using SFS - Configuring VPCs

- Multiple VPCs can be configured for an SFS file system so that servers belonging to different VPCs can share the same file system.

The screenshot shows a configuration page for an SFS file system named 'sfs-Vivi-mp'. It includes tabs for 'Basic Info' and 'Mount Point Info'. Under 'Basic Info', details are listed: Name (sfs-Vivi-mp), Protocol Type (NFS), Available Capacity (GB) (1.00), Region (AP-Bangkok), Created (Jul 28, 2021 15:42:10 GMT+08:00), ID (8dc27f33-d917-4630-8dc7-7dd0110e7484), Status (Available), Maximum Capacity (GB) (1.00), AZ (AZ1), and AZ1. Below this, there are 'Authorizations' and 'Tags' tabs. The 'Authorizations' tab displays a note: 'Only EC2s on VPCs can access file systems. If there are no available VPCs, apply for VPCs first.' A red-bordered button labeled 'Add Authorized VPC' is present, along with a message: 'You can add 19 more authorized VPCs and 299 more authorized addresses/segments.' A table lists an authorized VPC entry: Name (vpc-mp). The 'Tags' tab is also visible. At the bottom right, there is an 'Operation' section with a '1 Add Delete' link.

- Servers belonging to different VPCs can share the same file system as long as those VPCs are authorized in the SFS console.

## Differences Among SFS, OBS and EVS

Dimension	SFS	OBS	EVS
Definition	SFS provides on-demand high-performance file storage, which can be shared by multiple servers. Using SFS is like mounting a remote directory from a Windows or Linux server.	OBS provides massive, secure, reliable, and cost-effective data storage for users to store data of any type and size.	EVS provides scalable block storage that features high reliability and high performance to meet various service requirements. An EVS disk is similar to a hard disk on a PC.
Data storage logic	Stores files. Data is sorted and displayed in files and folders.	Stores objects. Files stored directly automatically generate system metadata, which can also be customized by users.	Stores binary data. Files cannot be stored directly. To store files, you need to format the disk first.
Access method	SFS systems can be accessed only after being mounted to servers through NFS or CIFS. A network address must be specified or mapped to a local directory for access.	OBS buckets can be accessed through the Internet or Direct Connect. The bucket address must be specified for access, and transmission protocols HTTP and HTTPS are used.	EVS disks can be used and accessed from applications only after being attached to ECSS or BMSS and initialized.
Application scenario	HPC, media processing, file sharing, content management, and web services	Big data analysis, static website hosting, online video on demand (VoD), and gene sequencing	HPC, enterprise clustered applications, enterprise application systems, and development and testing

## Quiz

1. (True or false) Before attaching an EVS disk to an ECS, you must stop the ECS.  
True  
False
2. (Multiple-choice) Which of the following is not an OBS function?
  - A. Cross-region replication
  - B. Versioning
  - C. URL validation
  - D. Attached to cloud servers for use

- False. You do not need to stop the ECS when attaching an EVS disk.
- D. EVS disks are attached to ECSs for use.

## Summary

- Where there is data, there is a need for data storage. After studying the content presented here, we should have a new understanding of storage types and we should understand HUAWEI CLOUD storage services a little better. As more and more enterprises migrate to the cloud, we are more able to better meet their storage requirements if we understand the positioning, principles, and usages of various storage services, for example, which storage services are suitable for video cloud and which are the best for databases.

# Recommendations

- Huawei iLearning
  - <https://e.huawei.com/en/talent/>
- HUAWEI CLOUD Help Center
  - <https://support.huaweicloud.com/intl/en-us/help-novice.html>
- HUAWEI CLOUD Academy
  - <https://edu.huaweicloud.com/intl/en-us/>

## Acronyms and Abbreviations

- AK/SK: Access Key ID/Secret Access Key
- API: Application Programming Interface
- AZ: Availability Zone
- BMS: Bare Metal Server
- CAD/CAE: Computer Aided Design/Computer Aided Engineering
- CIFS: Common Internet File System
- DES: Data Express Service
- DHCP: Dynamic Host Configuration Protocol
- ECS: Elastic Cloud Server
- EVS: Elastic Volume Service
- HA: High Available

## Acronyms and Abbreviations

- HPC: High Performance Computing
- HTTP: Hypertext Transfer Protocol
- HTTPS: Hypertext Transfer Protocol over Secure Sockets Layer
- IAM: Identity and Access Management
- IOPS: Input/Output Operations per Second
- NAS: Network Attached Storage
- NFS: Network File System
- OBS: Object Storage Service
- POSIX: Portable Operating System Interface
- SCSI: Small Computer System Interface
- SDK: Software Development Kit

## Acronyms and Abbreviations

- SFS: Scalable File Service
- SSD: Solid-State Drive
- VBD: Virtual Block Device
- VPC: Virtual Private Cloud

# Thank you.

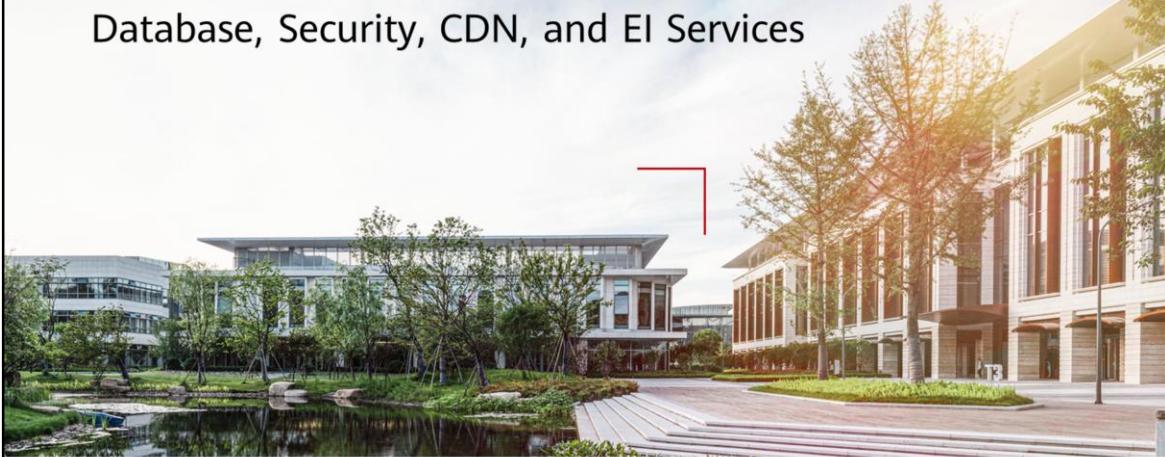
把数字世界带入每个人、每个家庭、  
每个组织，构建万物互联的智能世界。  
Bring digital to every person, home, and  
organization for a fully connected,  
intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.  
All Rights Reserved.

The information in this document may contain predictive  
statements including, without limitation, statements regarding  
the future financial and operating results, future product  
portfolio, new technology, etc. There are a number of factors that  
could cause actual results and developments to differ materially  
from those expressed or implied in the predictive statements.  
Therefore, such information is provided for reference purpose  
only and constitutes neither an offer nor an acceptance. Huawei  
may change the information at any time without notice.



## Database, Security, CDN, and EI Services



# Foreword

- In addition to compute, storage, and networking services, enterprises need database services, security services, Content Delivery Network (CDN), and EI services. These services can be billed on a pay-per-use basis and are easy to maintain, helping enterprises reduce investment and facilitate O&M.
- This chapter introduces database services, security services, CDN, and EI services.

# Objectives

- Upon completion of this course, you will:
  - Understand the basic concepts.
  - Understand the service positioning, principles, and functions.

# Contents

## **1. Database Services**

- Database Basics
  - Database Portfolio
  - RDS for MySQL
  - RDS for PostgreSQL
  - Document Database Service (DDS)

## 2. Security Services

## 3. Content Delivery Network (CDN)

## 4. EI Services

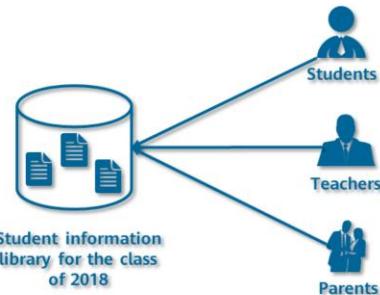
# Databases and Instances



- A database is a collection of files that contain data organized using a given model.



- An instance contains a set of background processes and memory structures. It is the data management software that connects users and the operating system (OS).



- We all know that data can be stored in multiple media formats, such as in memory or saved to disks. In fact, a database is also a medium for storing data.
- To be more specific, a database stores electronic documents, and you can add, intercept, update, and delete data in these documents.
- All operations on the data of databases, such as defining data, querying data, maintaining data, and managing database operations, are performed using database instances. Your applications only interact with the databases only through the instances.
- The smallest management unit of RDS is a database instance. A database instance is an isolated database environment running in the cloud. You can use RDS to create and manage database instances running various DB engines.

## Database Types

A relational database organizes data using a relational model. Data is stored in **rows and columns**. A user retrieves data from a database through a query, which is a type of command that qualifies certain areas of the database. A relational model can be simply understood as a two-dimensional table model, and a relational database is a way of organizing data consisting of two-dimensional tables and their relationships.

Relational database

A non-relational database refers to a non-relational data storage system not compliant with **ACID** properties.

Non-relational database

- ACID stands for atomicity, consistency, isolation, and durability.
  - Atomicity: Atomicity is the guarantee that series of database operations in an atomic transaction will either all occur or none will occur. If an error occurs during transaction execution, the transaction will be rolled back to the state from before it was committed.
  - Consistency: A consistent transaction will not violate integrity constraints placed on the data by the database rules. That is, executing a transaction cannot destroy the integrity or consistency of database data.
  - Isolation: Isolation means that concurrent transactions are executed sequentially. It guarantees the individuality of each transaction and prevents them from being affected by other transactions.
  - Durability: Once a transaction is committed, it will remain in the system even in the event of a system failure.

# Contents

## 1. Database Services

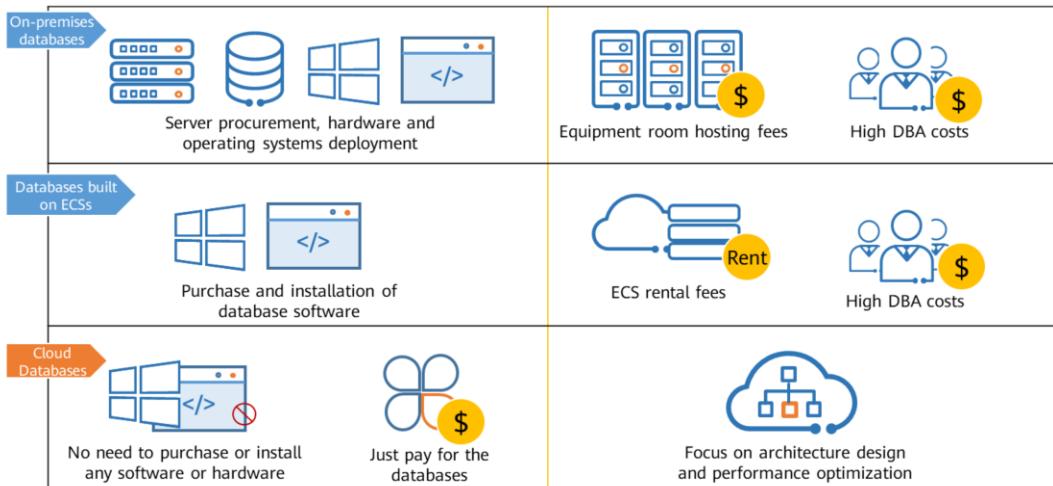
- Database Basics
- Database Portfolio
  - RDS for MySQL
  - RDS for PostgreSQL
  - Document Database Service (DDS)

## 2. Security Services

## 3. Content Delivery Network (CDN)

## 4. EI Services

## Differences Between Cloud and Other Database Solutions



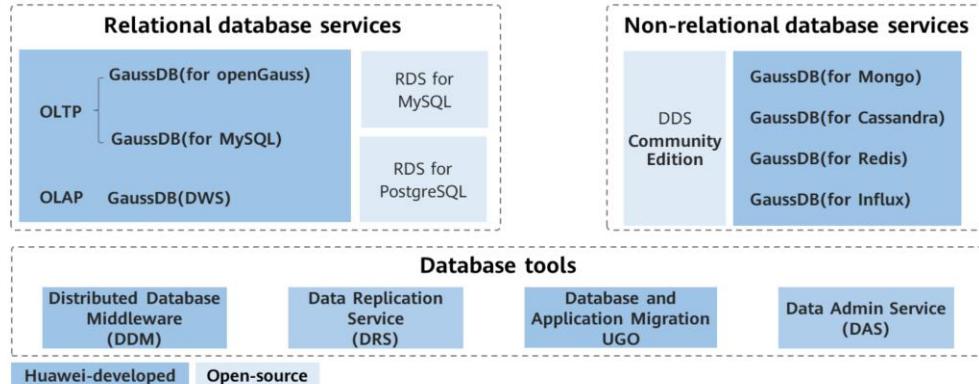
8      Huawei Confidential



- If customers want to build their own databases, they need to purchase hardware such as database servers and switches. If the hardware is damaged or replaced, the cost of repairing or replacing it is typically at least 30% of the project budget. It costs at least 3000 CNY per year to host 1U of cabinet space. If there are two 1U servers and a 1U intranet switch required for databases, the total hosting fee would be 9,000 CNY ( $3,000 \times 3$ ) in a year. The monthly salary of a junior DBA engineer is at least 5,000 CNY per month. If building the databases occupies 30% of the engineer's workload, the yearly labor cost is 18,000 CNY ( $5,000 \times 12 \times 30\%$ ). The sunk cost of this project is considerable. Open-source databases cannot be optimized. To ensure database reliability, customers have to prepare backup resources, which means more money. Public network traffic and domain name transfer are not free either.
- If customers want to deploy databases on ECSs, they need to purchase primary/standby ECS instances. Physical devices are provided by the service provider. Customers do not need to pay for the equipment room. They only need to hire DBA engineers to operate and maintain the database services. Elastic resources are provided. But open-source databases cannot be optimized, and backup represents a separate cost, along with traffic over a public network.
- Using cloud databases, customers only need to pay for the DB instances. The service provider provides the physical devices and maintains databases at its own cost. Resources are elastic and there is no charge for any public network traffic. Even the domain name generated for the DB instance is free, and regular updates help keep your instances updated to the latest MySQL version.

# HUAWEI CLOUD Database Portfolio

- GaussDB is an open-source database designed for small and medium enterprises to achieve the ultimate in cost-effectiveness. GaussDB is a Huawei-developed database that meets the high reliability and performance requirements of governments and enterprises.



9      Huawei Confidential



- Introduction to HUAWEI CLOUD database services:
  - PostgreSQL is an object-relational database management system (ORDBMS) derived from the POSTGRES package based on the 4.2 version written at the University of California, Berkeley. Many leading POSTGRES concepts were not around until fairly late in the development of business databases.
  - NoSQL refers to non-relational databases. Traditional relational databases are unable to keep up with the ultra-large-scale processing and massive concurrent SNS website requests involved with Internet Web 2.0 websites. NoSQL databases are designed to address the challenges of handling the multiple data types involved in large-scale data collections, especially where big data applications are concerned. NoSQL databases come in a variety of types based on different data models. The main types are key-value pair, wide column, document, and graph.
  - Distributed Database Middleware (DDM) works with the RDS service to remove a single node's dependency on hardware, facilitate capacity expansion to address data growth challenges, and ensure fast response to query requests. DDM eliminates the bottlenecks in capacity and performance and ensures that concurrent access is possible for a massive amount of data.
  - Data Replication Service (DRS) is a stable, secure, and efficient cloud service for online database migration and real-time database synchronization. DRS simplifies data transmission between databases and reduces data transfer costs.
  - Data Admin Service (DAS) enables you to manage DB instances on a web-based console, simplifying database management and improving efficiency and security.

# Contents

## **1. Database Services**

- Database Basics
- Database Portfolio
- **RDS for MySQL**
- RDS for PostgreSQL
- Document Database Service (DDS)

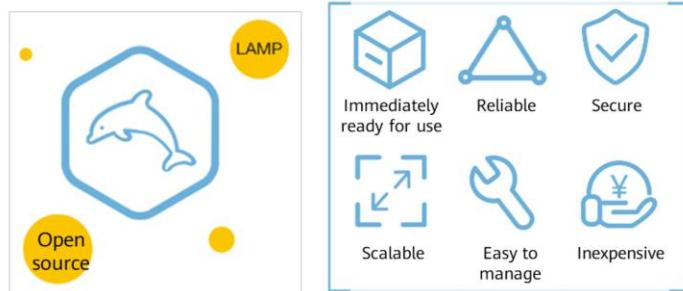
## 2. Security Services

## 3. Content Delivery Network (CDN)

## 4. EI Services

## What Is RDS for MySQL?

- MySQL is one of the world's most popular open-source relational databases. It works with the Linux, Apache, and PHP (LAMP) stack to provide efficient web solutions. RDS for MySQL is reliable, scalable, inexpensive, easy to manage, and immediately ready for use, freeing you to focus on developing your services.



- RDS for MySQL includes a comprehensive performance monitoring system, multi-level security protection measures, and a professional database management platform that allow you to easily set up and scale up databases. On the RDS for MySQL console, you can perform necessary tasks and no programming is required. The console simplifies operations and reduces routine O&M workloads, so you can stay focused on application and service development.
  - It uses a stable architecture and supports a range of web applications. It is also a cost-effective solution for small and medium enterprises.
  - A web-based console is available for you to monitor a comprehensive range of performance metrics.
  - You can flexibly scale resources based on your service requirements but you still pay only for what you use.

## Advantages of RDS for MySQL



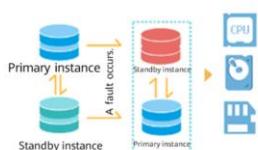
- Huawei enhanced MySQL kernel (HWSQL) provides 3 times higher performance in high-concurrency scenarios.



- A web-based management console provides an easy way to create, scale, monitor, and operate DB instances.



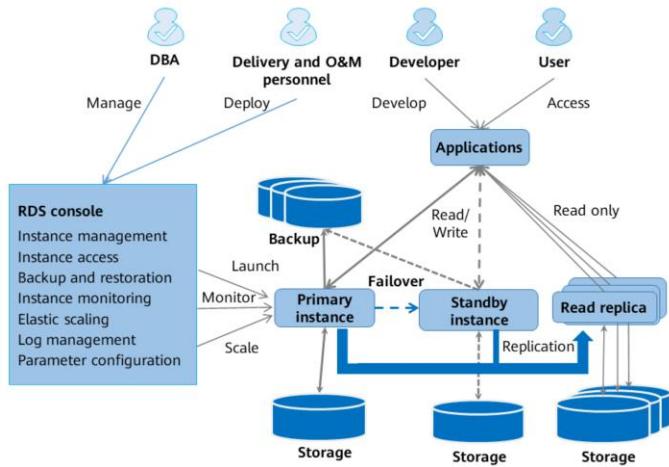
- RDS for MySQL is certified by China's Ministry of Public Security, uses security groups and VPCs to control access to databases, and supports post-incident audit.



- Enhanced semi-synchronous replication prevents data loss. Automatic failover takes only a few seconds, ensuring a low recovery time objective (RTO).

- Performance: RDS offers stable and high-performance database services using servers that have been proven robust by customer success in a wide range of applications. With RDS for MySQL, users can detect slow SQL statements and optimize database performance based on the provided suggestions. Users can access DB instances from the ECSs deployed in the same region as RDS over the intranet. This shortens the response time of applications and reduces the cost incurred by public network traffic.
- Security: RDS uses VPC and network security groups to isolate and secure your DB instances. RDS controls access by using IAM users and security groups; transmission is encrypted using TLS and SSL; and stored data is protected by static encryption and tablespace encryption. When a user deletes a DB instance, all data stored in the instance is deleted. If a DB instance is reachable over a public network, it may be vulnerable to distributed denial-of-service (DDoS) attacks, so RDS is protected by multiple layers of firewalls that can effectively defend against attacks like DDoS and SQL injections.
- Efficiency: RDS is scalable and easy to upgrade. It is billed on a pay-per-use basis and there is no need to purchase any hardware or software. System hosting and O&M are all handled by HUAWEI CLOUD, and since you only pay for the resources actually used, your resource utilization is effectively 100%.
- Reliability: RDS uses a hot standby architecture, so if there is an issue with an instance, failover takes a few seconds. It automatically backs up data every day and uploads the backup data to Object Storage Service (OBS). You can restore data from backups or to any point in time during the backup retention period. Deleted DB instances are moved in the recycle bin. You can rebuild them from the recycle bin within the retention period.

# Architecture of RDS for MySQL



13      Huawei Confidential



- RDS for MySQL provides:
  - Elastic scaling
    - **Horizontal scaling:** Add or delete read replicas (maximum number of read replicas: 5).
    - **Vertical scaling:** Change instance class and add storage (up to 10 TB).
  - Backup and restoration
    - **Backup:** Automated, manual, full, and incremental backups are supported. Backups can be added, deleted, queried, or replicated.
    - **Restoration:** Data can be restored to any point in time within the backup retention period, or to a new or an original DB instance. The backup retention period is up to 732 days.
  - Log management: Slow query logs and error logs can be queried and downloaded.
  - Parameter configuration: Database administrators (DBAs) can tune database performance by optimizing DB engine parameters based on metrics and logs. DB engine parameters can be added, deleted, modified, queried, reset, compared, and replicated.

## Application Scenarios of RDS for MySQL



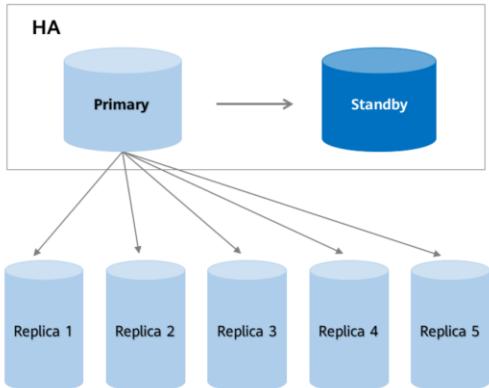
14      Huawei Confidential

HUAWEI

- RDS for MySQL is mainly used in the following scenarios:
  - Users of public cloud platforms other than HUAWEI CLOUD generally use RDS for MySQL.
  - Start-ups choose RDS for MySQL in the early stages because they need ways to support fast growth on a limited budget.
  - MySQL is used widely by Internet, e-commerce, and game enterprises. When migrating databases to the cloud, these types of enterprises choose RDS for MySQL.
  - IoT applications tend to be very large scale and they need to be extremely reliable. RDS for MySQL is the first choice for IoT enterprises because it allows for a large number of concurrent connections and does not require customers to reconstruct their applications.

## RDS for MySQL Features - Cross-AZ HA

### Cross-AZ HA



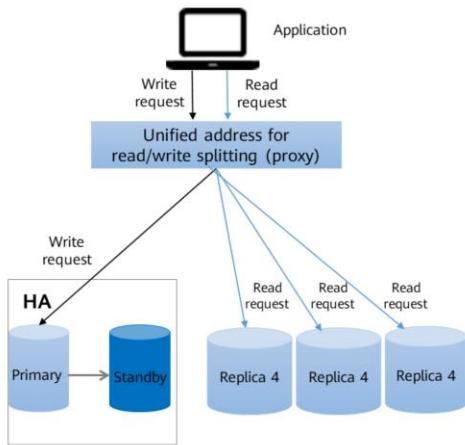
15      Huawei Confidential

### Functions

- Cross-AZ HA supports switchover in seconds.
- Up to 5 read replicas can be created for offloading read traffic.
- Standby DB instances are invisible to users. Users can access DB instances through virtual IP addresses.
- Read replicas cannot exist alone and must come with single or primary/standby DB instances.



## RDS for MySQL Features - Read/Write Splitting



16      Huawei Confidential

### Functions

- A single read/write splitting address is provided, transparent to applications.
- Read-only permissions can be configured for each node.
- Instance health check is performed. If a DB instance breaks down or the latency exceeds what is supported, read requests are no longer allocated to the instance.

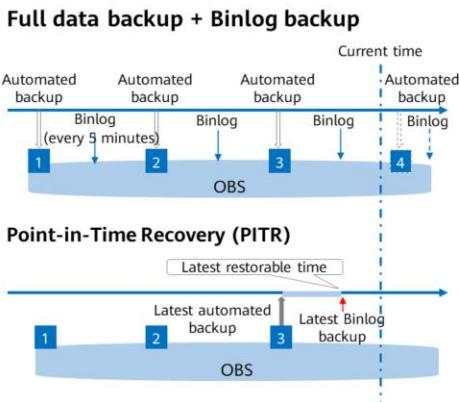
### Advantages

- A single read/write splitting address is provided, and read/write splitting does not require application reconstruction.
- The read weight assigned to a read replica is configurable.



- Read/write splitting enables read and write requests to be automatically routed through a read/write splitting address. You can enable read/write splitting after read replicas are created. Write requests are automatically routed to the primary DB instance and read requests are routed to read replicas by user-defined weights.

## RDS for MySQL Feature - Point-In-Time Recovery (PITR)



### Functions

- Instance-level restoration in seconds is supported.
- Automated backups can be configured to be saved for up to 732 days (approximately 2 years).
- You can restore data to any point in time at least 5 minutes ago and restore the data to a new DB instance or to the original DB instance.

### Advantages

- The backup retention period is up to 732 days.
- RDS provides free backup space approximately equal to your purchased storage space.

# Contents

## **1. Database Services**

- Database Basics
- Database Portfolio
- RDS for MySQL
- **RDS for PostgreSQL**
- Document Database Service (DDS)

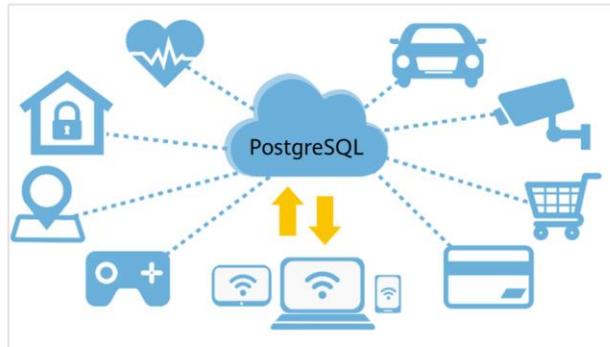
## 2. Security Services

## 3. Content Delivery Network (CDN)

## 4. EI Services

## What Is RDS for PostgreSQL?

- RDS for PostgreSQL is a typical open-source relational database that excels in data reliability and integrity. It supports Internet e-commerce, geographic location application systems, financial insurance systems, complex data object processing, and other applications.



19      Huawei Confidential

 HUAWEI

- PostgreSQL is based on Postgres, which was developed at the University of California, Berkeley. After more than 30 years of development, PostgreSQL has become the most powerful **open-source** database in the world. It has earned a reputation for reliability, stability, and data consistency, and has become the preferred open-source relational database for many enterprises.
- PostgreSQL is an open source object-relational database management system focused on extensibility and standards compliance. It is known as the most advanced open source database. RDS for PostgreSQL is designed for enterprise-oriented OLTP scenarios and supports NoSQL (JSON, XML, or hstore) and GIS data types. It has earned a reputation for reliability and data integrity, and is suitable for websites, location-based applications, and complex data object processing.
- RDS for PostgreSQL supports the Postgres plugin, which provides excellent spatial performance.
- RDS for PostgreSQL is a cost-effective solution for a range of different scenarios. You can flexibly scale resources based on service requirements and you only pay for what you use.

## Advantages of RDS for PostgreSQL

### Ease-of-use

- Services can be provisioned in minutes, and multiple specifications are available.

### Reliability

- The primary and standby instances can fail over in the event of a fault.

### Efficient management

- A range of metrics are monitored and can be viewed on the console.

### Scalability

- Resources are used on demand and can be scaled flexibly.

### High performance

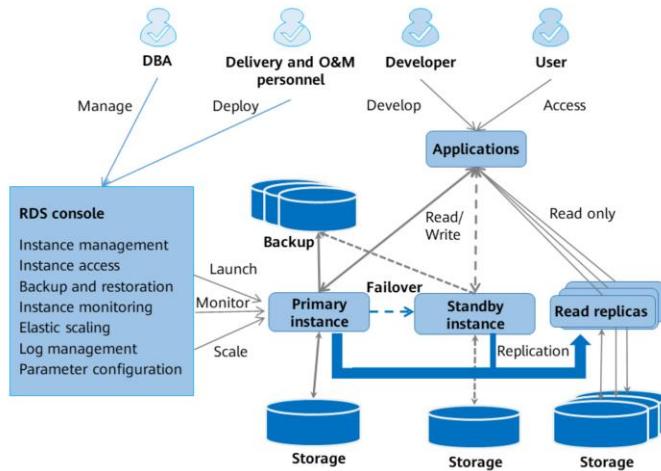
- Read replicas can be created for read/write splitting.

### Easy migration

- Data Replication Service (DRS) provides online and offline migration and is compatible with third-party databases.

- RDS for PostgreSQL has many advantages. Currently, it is mainly used to migrate Oracle databases.

## Architecture of RDS for PostgreSQL

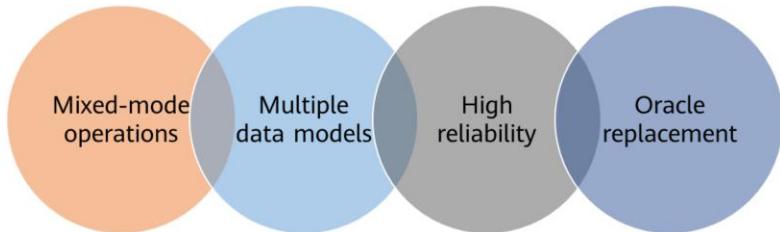


21 Huawei Confidential



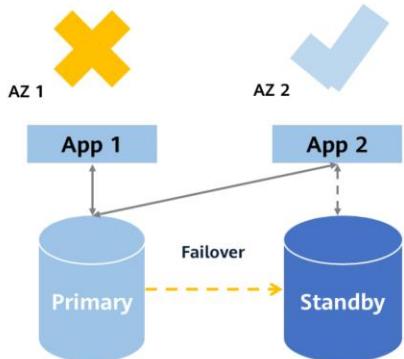
- RDS for PostgreSQL has the following features:
  - Database type: PostgreSQL 9.5, 9.6, 10.0, 11, 12, and Enhanced Edition are provided.
  - Security: Multiple security measures such as VPCs, subnets, security groups, and SSL are provided to protect databases and user privacy.
  - HA: Data is automatically synchronized from a primary DB instance to a standby DB instance. If the primary DB instance fails, services are quickly and automatically switched over to the standby DB instance.
  - Monitoring: Key performance metrics of RDS DB instances are monitored. These metrics include the CPU usage, memory usage, storage space usage, I/O activity, database connections, QPS, TPS, buffer pools, and read/write activities.
  - Elastic scaling
    - Horizontal scaling: Read replicas (up to five for each instance) can be created or deleted.
    - Vertical scaling: DB instance classes can be modified.
    - Instances can be scaled out with a few clicks and without interrupting services.
  - Log management: Slow query logs and error logs can be queried.
  - Parameter configuration: Database administrators (DBAs) can tune database performance by optimizing DB engine parameters based on metrics and logs.

## Applications of RDS for PostgreSQL



- Mixed-mode operations combining OLTP and OLAP are supported.
- Multiple data models are applicable to spatiotemporal, geographic, heterogeneous, image, text retrieval, time series, stream computing, and multi-dimensional scenarios.
- Huawei provides you with a reliable database service and keeps your data consistent.
- To replace Oracle databases, there are two solutions available:
  - Use RDS for PostgreSQL Enhanced Edition.
  - Use RDS for PostgreSQL Community Edition and Oracle plug-ins.

## RDS for PostgreSQL Features - High Availability

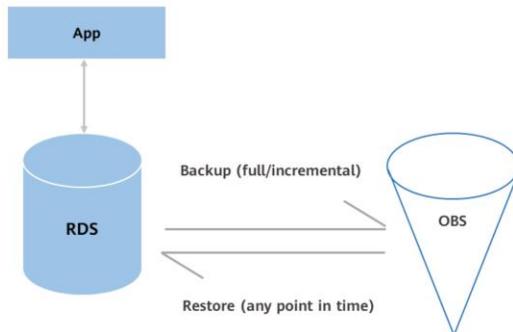


### Benefits of the HA cluster architecture:

- You can choose a failover policy to prioritize reliability or availability.
- DB instances can be deployed in one AZ or across AZs and can automatically fail over within a cluster.
- You can manually switch a primary instance to standby to simulate a fault.
- A read replica can automatically associate itself with a new primary node.
- A switchover can be completed in seconds.
- The standby database does not handle traffic. It only ensures RTO.
- A Huawei-developed HA Monitor module is used.
- Virtual IP addresses can be switched completely invisibly to the applications.
- Multiple primary/standby switchovers can be performed.
- Automatic fault detection is provided.

- RTO stands for Recovery Time Objective. It is the length of time from when an IT system breaks down and services stop to when the system recovers.
- HA Monitor is an HA monitoring module.

## RDS for PostgreSQL Features - Point-In-Time Recovery (PITR)



- Backup cycle: 7 to 732 days
- Pay-per-use: Free EVS storage space equal to the requested storage and virtually limitlessly expandable
- Reliability: Up to 11 nines of data reliability
- Security encryption: KMS encryption and multiple protections

**Data archived in OBS can be restored to any point in time.**

- RDS stands for Relational Database Service.
- EVS stands for Elastic Volume Service.
- OBS stands for Object Storage Service.

# Contents

## **1. Database Services**

- Database Basics
- Database Portfolio
- RDS for MySQL
- RDS for PostgreSQL
- **Document Database Service (DDS)**

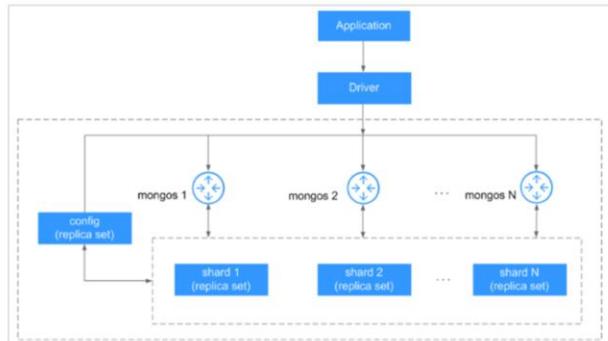
## 2. Security Services

## 3. Content Delivery Network (CDN)

## 4. EI Services

## What Is DDS?

- Document Database Service (DDS) is a high-performance, highly availability MongoDB-compatible database service that is scalable and secure. It provides one-click deployment, elastic capacity expansion, disaster recovery, backup, restoration, monitoring, and alarm reporting.



26      Huawei Confidential



- Each DDS cluster is an independent document database. A sharded cluster consists of a config node, and multiple mongos and shard nodes.
- Data read and write requests are forwarded by the mongos nodes, which read configuration settings from config, and then allocate the read and write requests to the shards, making it easy to cope with high concurrency scenarios. In addition, each config node, along with the shards in its cluster, is replicated in triplicate to ensure high availability.

## DDS Advantages

100% MongoDB compatibility

- You can migrate on-premises MongoDB databases to the cloud without reconstructing your services.

Efficient O&M

- You can monitor DB instances from a convenient UI and expand storage in just a few clicks.

Reliable, available, and secure

- You can create and save automated or manual backups of your DB instance to ensure data security.

3 types of architectures

- You can use clusters, replica sets, and single nodes as required.

- Fully compatible
  - DDS is a document-oriented NoSQL database that is fully compatible with MongoDB.
- Reliable, available, and secure
  - The security system consists of VPCs, subnets, security groups, Anti-DDoS, and SSL, which collectively can defend against a wide range of attacks and keep your data secure. DDS supports audit logs, which can be stored for up to two years. DDS supports fine-grained permission control. The cluster and replica set support high availability. If the primary node is faulty, the secondary node quickly takes over services. The switchover process is unnoticeable to the applications.
  - DDS supports both automated and manual backup. The maximum retention period for an automated backup is 732 days. A manual backup can be retained until you delete it.
  - You can restore a DB instance from a backup file. Replica sets support point-in-time recovery at the instance, database, and table-level.
- Efficient O&M
  - DDS console is a visualized instance management platform. You can restart, back up, or restore an instance in just a few clicks.
  - DDS monitors key performance metrics of DB instances and DB engines in real time, including the CPU usage, memory usage, storage space usage, command execution frequency, delete statement execution frequency, insert statement execution frequency, and the total number of active connections.

## Basic Concepts

- A DDS cluster consists of three types of nodes: mongos, config, and shard, each of which has different functions.

**mongos**

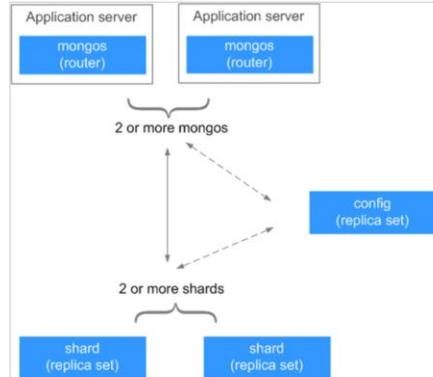
- Each mongos node routes read and write requests, providing a unified interface for accessing DB instances.

**config**

- A config node is deployed as a replica set and stores instance configuration data.

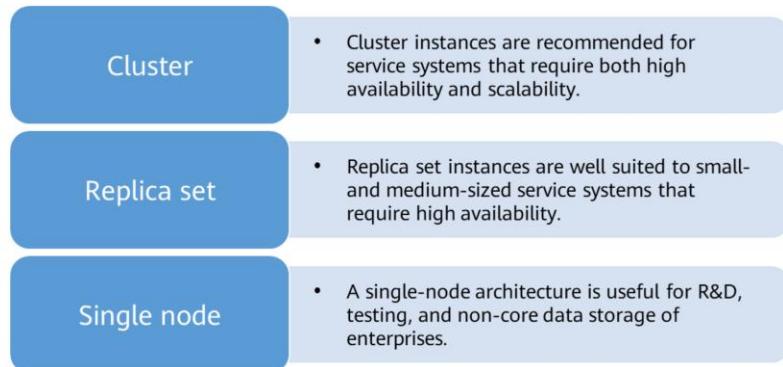
**shard**

- Shard nodes store user data.



## Overview Architecture

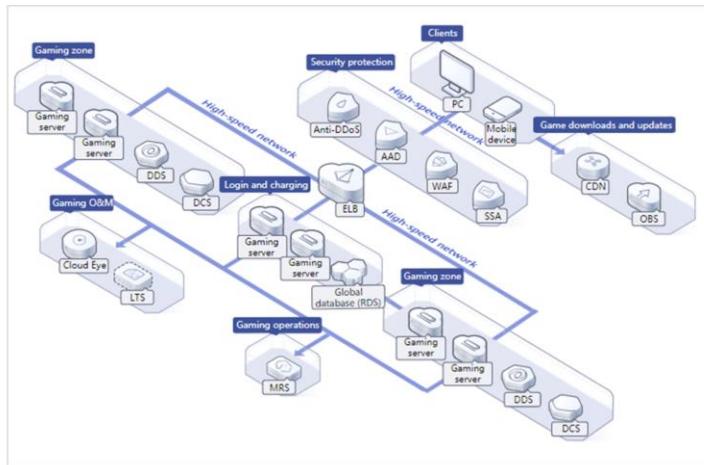
- DDS supports the following deployment modes:



- DDS supports three deployment modes: cluster, replica set, and single node.
  - DDS provides sharded cluster instances comprised of a config node paired with multiple shards and mongos nodes.
  - A replica set consists of three nodes: primary, secondary, and hidden. The three-node architecture is set up automatically, and the three nodes automatically synchronize data with each other to ensure data reliability.
  - The single node architecture is a supplementary deployment mode that is useful for R&D, testing, and non-core data storage.

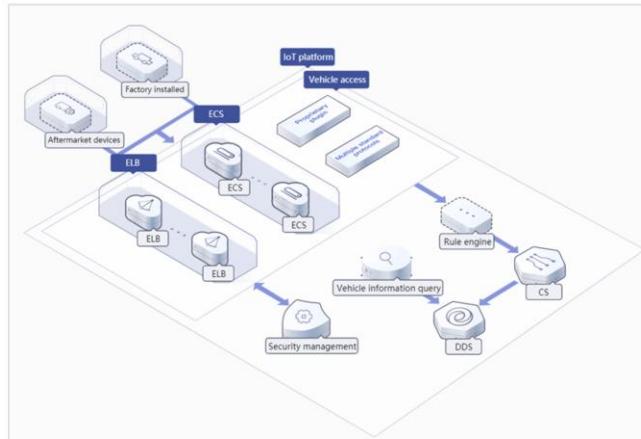
## Applications - Gaming

- DDS offers fast, reliable access to increasingly complex player profiles, including details such as character scores, items acquired and other details. For MMO games, the highly-available architecture of DDS clusters and replica sets can provide a smooth gaming experience even during peak hours.



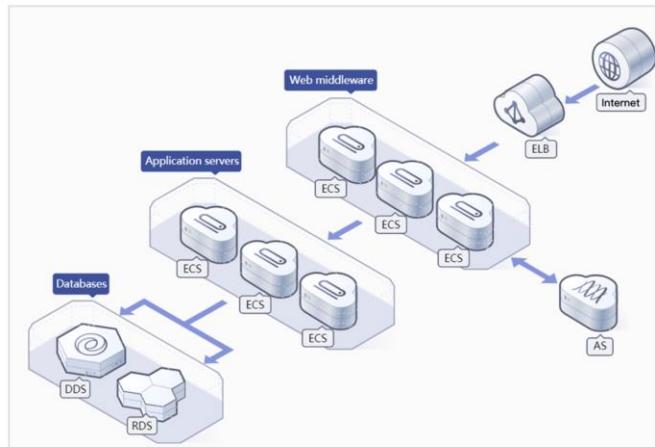
## Applications - IoT

- IoT applications feature high-concurrency writes, diverse data types, and sudden spikes in data volumes. With high performance and asynchronous data writes, DDS is able to process data as fast as in-memory databases when and where it is needed. In addition, the quantities and specifications of mongos and shard nodes in DDS cluster instances can be dynamically increased to meet growing demands, making DDS ideal for IoT applications.



## Applications - Internet

- DDS replica sets use a three-node architecture to deliver reliability and enable disaster recovery. The three data nodes form an anti-affinity group and are deployed on different physical servers to automatically synchronize data. The primary and secondary nodes provide services. Each node has an independent private network address and works with the driver to distribute read load.



# Differences Between Cloud and Other Database Solutions

- Benefits: Cloud database O&M is more efficient, freeing up your database team to focus on database architecture design.

## [On-premises Databases]

- Server procurement and hardware and operating systems deployment
- High hosting fees
- OS and database O&M

Database architecture design
Database tuning
Elastic scaling
High availability
Backup and restoration
Version upgrades and patch installation
Database software installation
OS version upgrade and patch installation
OS installation
Server deployment and maintenance
Rack stacking
Equipment room, power supply, air conditioning, and network infrastructure

## [Databases on an ECS]

- Database hardware procurement and installation
- Costs of renting cloud servers
- Database O&M

Database architecture design
Database tuning
Elastic scaling
High availability
Backup and restoration
Version upgrades and patch installation
Database software installation
OS version upgrade and patch installation
OS installation
Server deployment and maintenance
Rack stacking
Equipment room, power supply, air conditioning, and network infrastructure

## [Cloud Databases]

- No hardware or software investment
- Focused on database architecture design
- Focused on database application optimization

Database architecture design
Database tuning
Elastic scaling
High availability
Backup and restoration
Version upgrades and patch installation
Database software installation
OS version upgrade and patch installation
OS installation
Server deployment and maintenance
Rack stacking
Equipment room, power supply, air conditioning, and network infrastructure

 HUAWEI

- Cloud databases help you reduce the total cost of ownership (TCO) and O&M workload, freeing you to stay focused on developing key services.

# Contents

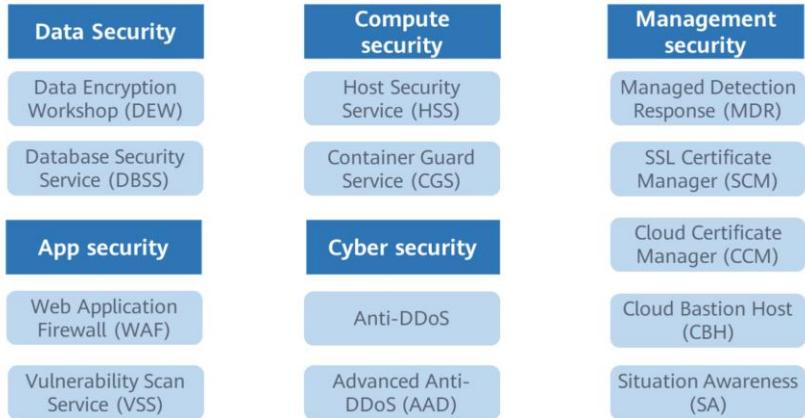
1. Database Services
2. **Security Services**
  - Customer Requirements on Cloud Security
    - HSS
    - WAF
    - DEW
    - IAM
3. Content Delivery Network (CDN)
4. EI Services

# Customer Requirements on Cloud Security

CSA Top Threats		Key Security Requirements for Enterprise Cloudification		
<ul style="list-style-type: none"><li>• Data Leakage</li><li>• Insufficient identity, credential, and access management</li><li>• Insecure ports and APIs</li><li>• System vulnerabilities</li><li>• Account hijacking</li><li>• Malicious insiders</li></ul>	<ul style="list-style-type: none"><li>• Advanced persistent threat (APT)</li><li>• Data loss</li><li>• Insufficient due diligence</li><li>• Abuse and nefarious use of cloud services</li><li>• Denial of service (DoS)</li><li>• Shared technology vulnerabilities</li></ul>	<b>Continuous services</b> <ul style="list-style-type: none"><li>• Defend against network attackers and hackers.</li><li>• Comply with laws and regulations.</li></ul>	<b>Controllable O&amp;M</b> <ul style="list-style-type: none"><li>• Configure security policies.</li><li>• Detect and eliminate risks.</li><li>• Audit and trace operations.</li></ul>	<b>Data confidentiality</b> <ul style="list-style-type: none"><li>• Prevent data breach. Data is accessible only to authorized staff.</li></ul>

# HUAWEI CLOUD Security Services

- Build a series of top-quality security services for ensuring data security.



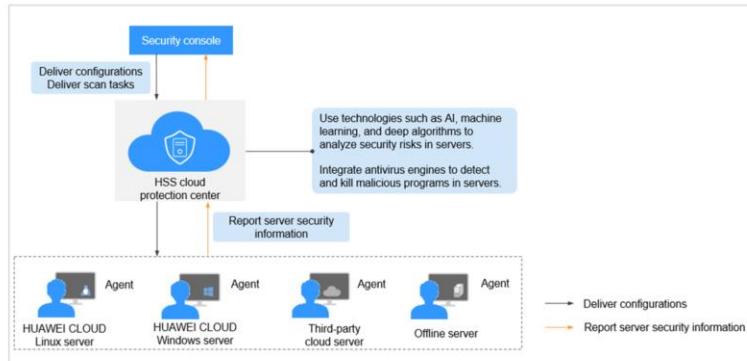
- Security services are developed to address different aspects of information security. Today we'll talk about five types of security services.

# Contents

1. Database Services
2. **Security Services**
  - Customer Requirements on Cloud Security
  - HSS
  - WAF
  - DEW
  - IAM
3. Content Delivery Network (CDN)
4. EI Services

## What Is HSS?

- Host Security Service (HSS) helps you identify and manage the assets on your servers, eliminate risks, and defend against intrusions and web page tampering. There are also advanced protection and security operations functions available to help you easily detect and handle threats.



## HSS Features

Centralized management

- You can easily manage, scan, and protect your servers from a single console.

Precision defense

- HSS blocks attacks with pinpoint accuracy by using advanced detection technologies and diverse libraries.

Lightweight agent

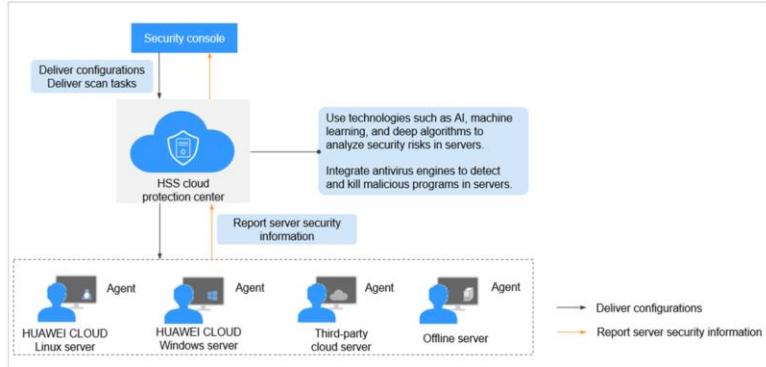
- The lightweight agent occupies only very limited resources, having no impact on system performance.

Comprehensive protection

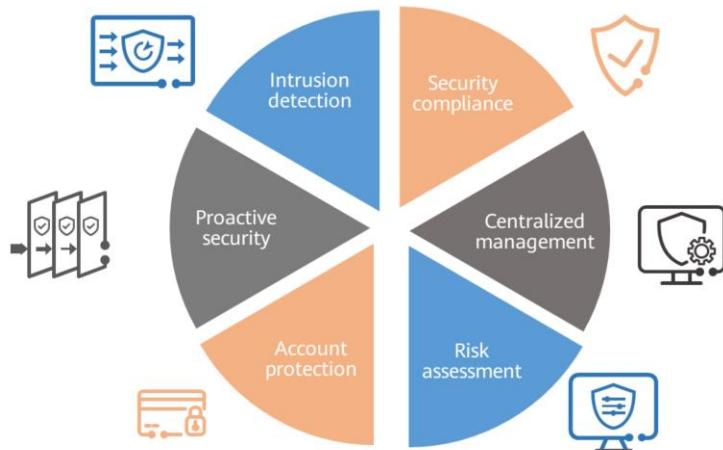
- Prevention before, protection during, and scanning and inspection after any attack.

## How HSS Works

- Install the HSS agent on your servers, and you will be able to monitor the server security status and identify risks in a region from the HSS console.



## HSS Applications



41      Huawei Confidential

 HUAWEI

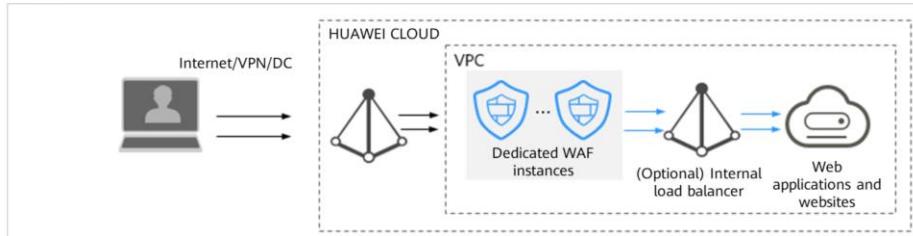
- HSS applications
  - Compliance: HSS protects accounts and systems on cloud servers, helping enterprises meet compliance standards.
  - Centralized management: You can manage servers, security configurations, and security events all from the HSS console. You can reduce security risks from a single convenient portal and keep management costs down.
  - Risk assessment: HSS scans your servers for risks, including unsafe accounts, ports, software vulnerabilities, and weak passwords, and prompts you to eliminate any security risks identified and harden the system in a timely manner.
  - Account protection: Accounts are protected before, during, and after a security event. You can also use 2FA to block brute-force attacks on accounts, enhancing the security of your cloud servers.
  - Proactive defense: You can count and scan your server assets, check and fix vulnerabilities and unsafe settings, and proactively protect your network, applications, and files from attacks.
  - Intrusion detection: You can scan all possible attack vectors to detect and fight APTs and other threats in real time, protecting your system from their impacts.

# Contents

1. Database Services
2. **Security Services**
  - Customer Requirements on Cloud Security
  - HSS
  - WAF
  - DEW
  - IAM
3. Content Delivery Network (CDN)
4. EI Services

## What Is WAF?

- Web Application Firewall (WAF) keeps your website safe and stable. It comprehensively examines website service traffic to accurately identify malicious requests and block attacks, ensuring best-of-class system security and stability for your applications and data.



- These days, more and more enterprises are migrating critical services online. Although the migration brings economic benefits and improves flexibility, it also introduces new security threats and compliance requirements. In recent years, attacks on web applications have been increasing, and new attack techniques have been emerging one after another. Traditional firewalls are not up to the task. So, here comes WAF, a powerful tool to defend against attacks at the application layer.

# WAF Features

## Comprehensive Protection

- WAF uses an extensive built-in attack signature library to detect and block dozens of common online attacks.

## Top-notch Reliability

- WAF ensures zero service interruptions with distributed deployment, 24/7 monitoring, and remote disaster recovery.

## Industry-leading Technologies

- WAF uses an industry-leading engine to accurately identify a wide range of threats, greatly improving the threat discovery rate.

## Flexible Configuration

- WAF provides multiple built-in configuration fields, enabling users to customize rules for focused protection.

- Basic web protection:
  - Backed by an extensive preset reputation database, WAF can defend against the Open Web Application Security Project (OWASP) top 10 threats, vulnerability exploits, web shells, and other threats.
  - WAF detects and blocks varied attacks, such as SQL injection, XSS attacks, remote overflow vulnerabilities, file inclusion, Bash vulnerabilities, remote command execution, directory traversal, sensitive file access, and command/code injection.
  - WAF provides web shell detection, protecting web applications from web shells.
  - Precise identification:
    - WAF uses a wide range of techniques to identify attacks. For example, WAF uses a dual-engine architecture, combining built-in semantic analysis engine and regex engine. WAF enables users to configure blacklist and whitelist rules, so WAF has a low false positive rate.
    - WAF can automatically decode common codes no matter how many times they are encoded. WAF can decode a wide range of code types, including url\_encode, Unicode, XML, C-OCT, hexadecimal, HTML escape, and base64 code, case confusion, JavaScript, shell, and PHP concatenation confusion.
    - WAF deep inspection identifies and blocks evasion attacks, including those that use homomorphic character obfuscation, command injection with deformed wildcard characters, UTF7, data URI scheme, and other techniques.
    - WAF header detection inspects all header fields in received requests.

## How WAF Works

- After a website is connected to WAF, all website access requests are forwarded to WAF first. Then, WAF inspects the traffic, filters out malicious traffic, and routes only normal traffic to the origin server, keeping the origin server secure, stable, and available.



- To enable WAF, after purchasing a WAF instance, go to the WAF console and connect the website to be protected to the WAF instance. After that, all website access requests go to WAF first. Then, WAF inspects the traffic, filters out attacks, and routes only normal traffic to the origin server, keeping the origin server secure, stable, and available.
- The process of forwarding website traffic to the origin server through WAF is called back-to-source. WAF inspects traffic originating from the client and uses WAF back-to-source IP addresses to forward normal traffic to the origin server. To the origin server, source IP addresses of all requests are the WAF back-to-source IP addresses. In this way, the IP address of the origin server is hidden from the client.

## WAF Application Scenarios



- WAF application scenarios:
  - Basic protection: WAF helps users defend against common web attacks, such as command injection and sensitive file access.
  - Promotions on e-Commerce platforms: A large number of malicious requests may go to service interfaces during online promotions. WAF enables customizable rate limiting rules to defend against CC attacks. This protects website services from breakdowns caused by too many concurrent requests while ensuring responses to legitimate requests.
  - Defense against zero-Day vulnerabilities: If website services fail to recover quickly from the impact of zero-day vulnerabilities in third-party web frameworks or plug-ins, WAF will update the preset protection rules immediately to ensure service security and stability. WAF functions as an extra protection on a third-party network. Compared with directly fixing vulnerabilities on the third-party architecture, using WAF rules to block risks is a quicker way.
  - Data leak prevention: WAF prevents malicious actors from using methods such as SQL injection and web shells to bypass application security and gain remote access to web databases and other sensitive information. Users can customize WAF data masking rules for:
    - Accurate identification of attacks: WAF uses semantic analysis & regex to examine traffic from different dimensions, precisely detecting malicious traffic.
    - Detection of attacks that are encoded several times: WAF detects a wide range of distortion attack patterns with 7 decoding methods to prevent bypass attempts.
  - Web tamper prevention: WAF ensures that attackers cannot leave web shells on protected web servers or tamper with web page content, preventing damage to customer credibility. Users can customize web tamper prevention rules to detect Trojans and prevent web pages from

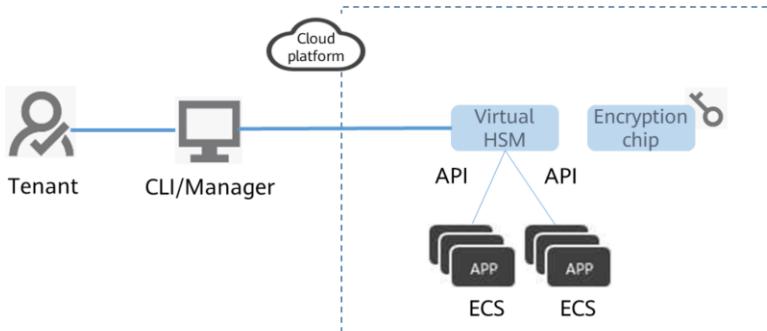
being tampered with.

# Contents

1. Database Services
2. **Security Services**
  - Customer Requirements on Cloud Security
  - HSS
  - WAF
  - DEW
  - IAM
3. Content Delivery Network (CDN)
4. EI Services

## What Is DEW?

- Data Encryption Workshop (DEW) is a cloud data encryption service. It provides Key Management Service (KMS), Key Pair Service (KPS), and Dedicated Hardware Security Module (Dedicated HSM).

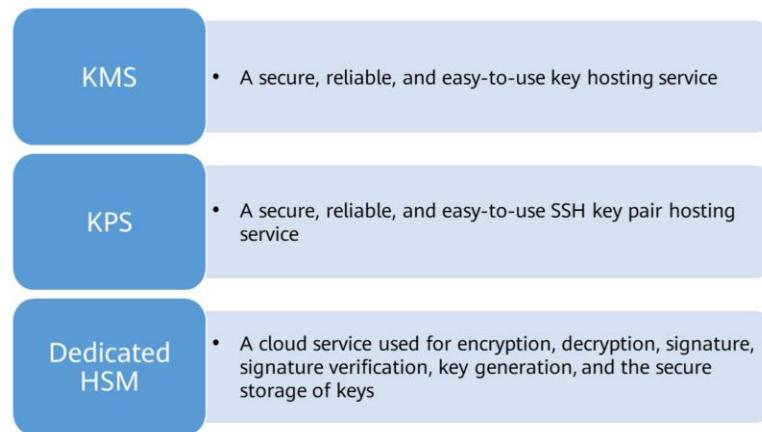


48      Huawei Confidential



- Data is a core enterprise asset, and data breaches can result in immeasurable losses. DEW can encrypt customer data and protect it from data leaks.
- DEW uses HSMs to protect your keys, and can be integrated with other HUAWEI CLOUD services to address data security, key security, and key management issues. You can also develop your own encryption applications based on DEW.

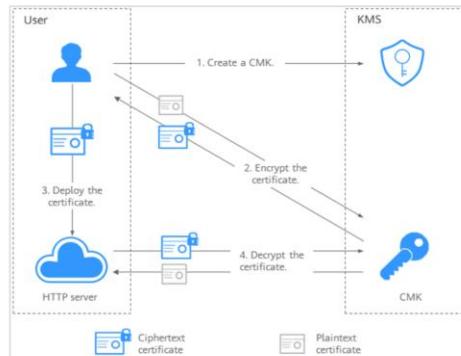
## DEW Services



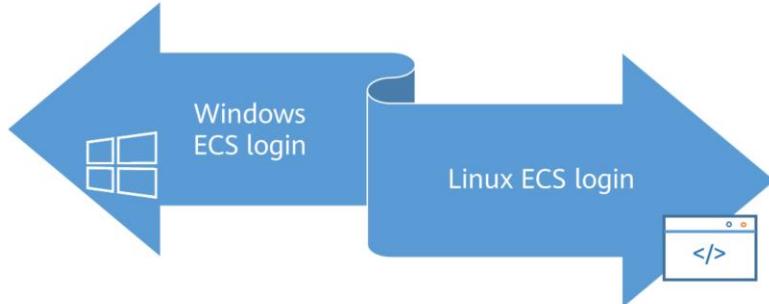
- DEW consists of three services. Let's take a look at them.

## KMS Application: Small Data Encryption and Decryption

- Scenario: You can use online tools on the KMS console or call KMS APIs to directly encrypt or decrypt small amounts of data with a CMK, for instance, passwords, certificates, or phone numbers.

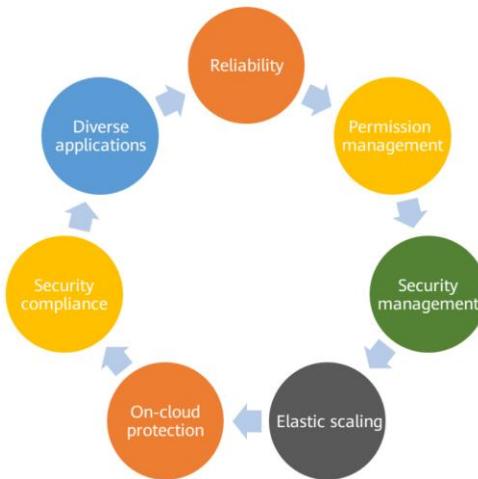


## KPS Applications



- Password authentication is required to log in to a Windows ECS. First of all, you must obtain the administrator password (the password an administrator or another account configured in Cloudbase-Init) generated during the initial installation of the ECS from the private key file downloaded when you create the ECS. This secure password is randomly generated.

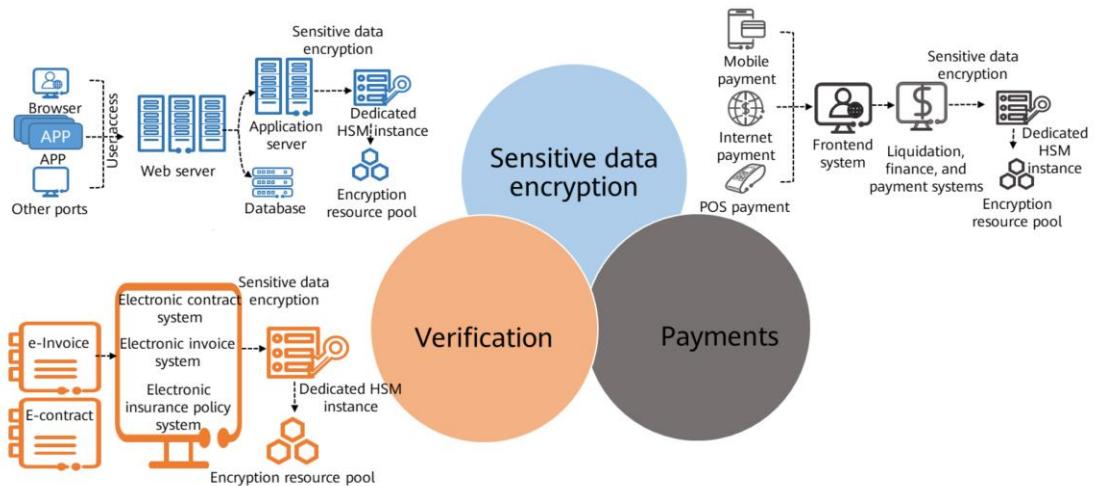
## Dedicated HSM Features



- Product advantages:
  - On-cloud protection: Dedicated HSM can transfer offline encryption capabilities to the cloud, reducing your O&M costs.
  - Elastic scaling: You can flexibly increase or decrease the number of HSM instances deployed based on your service needs.
  - Security management: Dedicated HSM separates device management from the management of content (sensitive information). As a user of the device, you can control the generation, storage, and access of keys. Dedicated HSM only monitors and manages devices and related network facilities. Dedicated HSM O&M personnel have no access to customer keys.
  - Permissions management: Sensitive commands are classified hierarchically for permissions management, so you can prevent unauthorized execution. Several authentication types are supported, such as username/password and digital certificates.
  - Reliability: Dedicated HSM provides China State Cryptography Administration (CSCA) certified and FIPS 140-2 validated level 3 HSMs to protect your keys, guaranteeing high-performance encryption services to meet even the most stringent security requirements. Dedicated HSM chips are used exclusively by each instance. Even if some hardware chips are damaged, the service are not affected.
  - Security compliance: Dedicated HSM provides State Cryptography Administration (SCA) validated HSM instances, helping you protect your data on Elastic Cloud Servers (ECSs) and meet compliance requirements.
  - Diverse applications: Dedicated HSM offers finance HSM, server HSM, and

signature server HSM instances to flexibly adapt to a diverse range of service scenarios.

## Dedicated HSM Application Scenario



53      Huawei Confidential



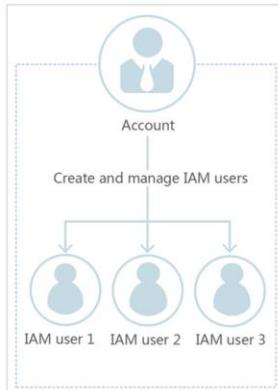
- Applications:
  - Sensitive data encryption: government public services, Internet enterprises, and system applications that contain immense volumes of sensitive information
  - Payment: payment and prepayment applications, such as transportation cards and e-commerce platforms
  - Verification: Dedicated HSM can ensure the confidentiality and integrity of electronic contracts, invoices, insurance policies, and medical records during transmission and storage.

# Contents

1. Database Services
2. **Security Services**
  - Customer Requirements on Cloud Security
  - HSS
  - WAF
  - DEW
  - IAM
3. Content Delivery Network (CDN)
4. EI Services

## What Is IAM?

- Identity and Access Management (IAM) helps you manage your users and control their access to HUAWEI CLOUD services and resources.



- A typical enterprise has multiple IT administrators, each responsible for managing different resources. It's more secure to not give every administrator super administrator permissions. Thanks to HUAWEI CLOUD IAM, an enterprise administrator can create multiple users with separate permissions.
- Credentials authenticate a user on the HUAWEI CLOUD console or APIs. Credentials include a password and access keys. The enterprise administrator manages both their own credentials and the credentials of IAM users they create.

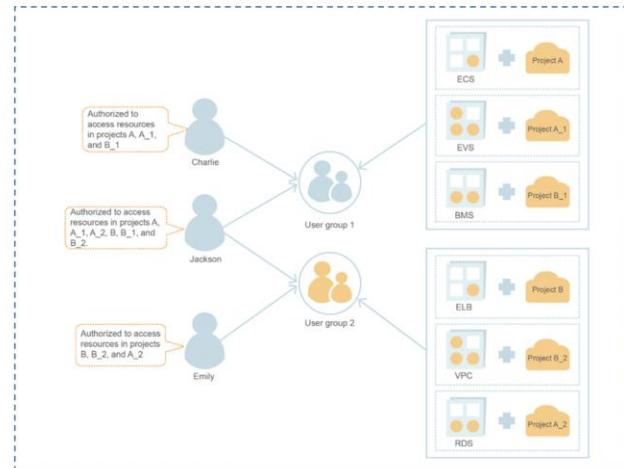
## Why Choose IAM?

Federated access with existing enterprise accounts

Finer access control of HUAWEI CLOUD resources

Delegated access to resources across accounts

## Finer Access Control of HUAWEI CLOUD Resources

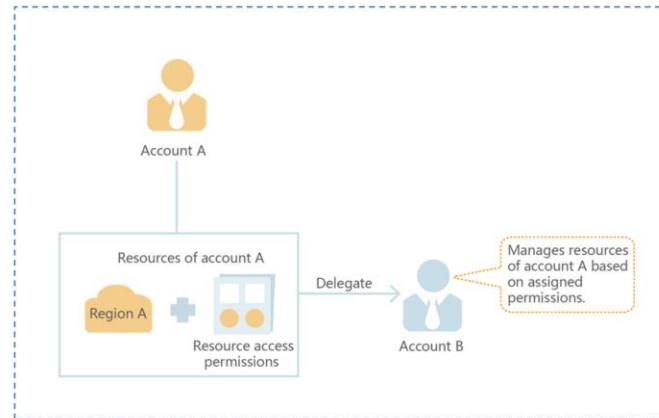


57      Huawei Confidential



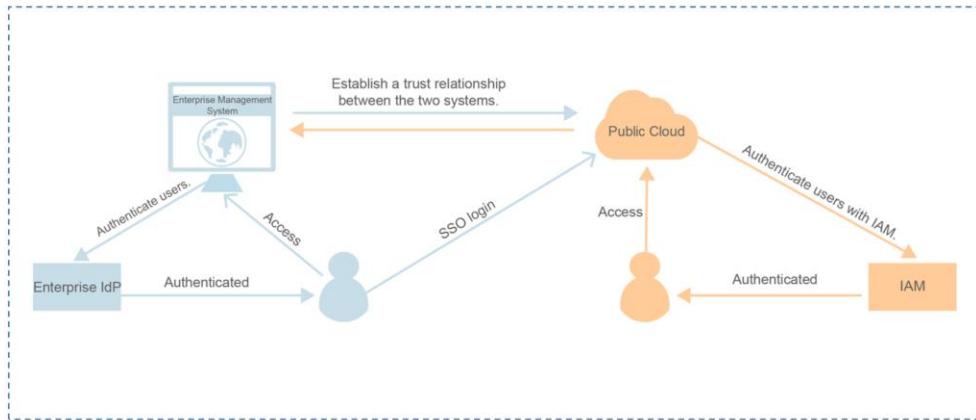
- Now you've purchased HUAWEI CLOUD resources such as Elastic Cloud Server (ECS), Elastic Volume Service (EVS), and Bare Metal Server (BMS), create IAM users for different teams or applications in your enterprise. Each user gets the permissions they need to do their job. IAM users each have a username and password to log in to HUAWEI CLOUD and access resources under your account.

## Delegated Access to Resources Across Accounts



- Say you've created an agency for a professional O&M company. The company can then use its own account to manage certain resources in your account. IAM allows you to modify or cancel delegated permissions at any time. In this figure, account A is the delegating party, and account B is the delegated party.

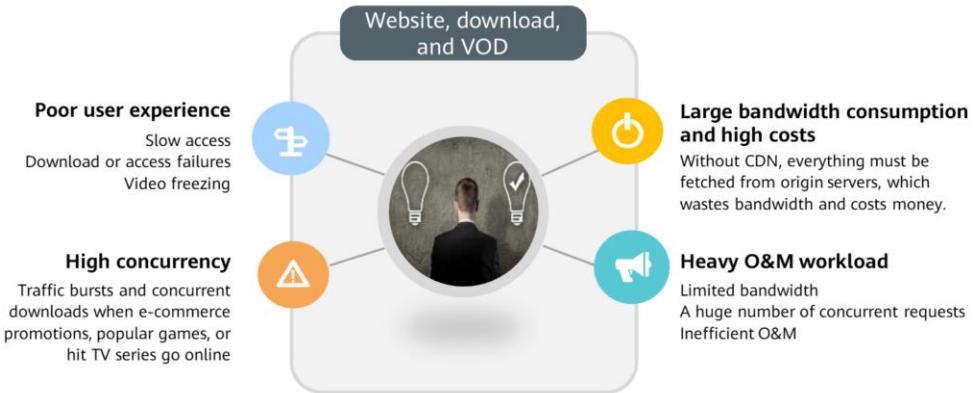
## Federated Access with Existing Enterprise Accounts



# Contents

1. Database Services
2. Security Services
- 3. Content Delivery Network (CDN)**
4. EI Services

## Pain Points



## What Is CDN?

- Content Delivery Network (CDN) is an intelligent virtual network built on top of existing Internet infrastructure. Origin content is cached on CDN nodes around the world so users can quickly obtain desired content from nearby nodes.



62      Huawei Confidential



- CDN speeds up website response and improves website availability, breaking through the bottlenecks caused by low bandwidth, heavy user access traffic, and uneven distribution of nodes.

## Node Distribution in the Chinese Mainland

- HUAWEI CLOUD CDN operates 2,000+ nodes in the Chinese mainland. These nodes are connected to the networks of top carriers in China such as China Telecom, China Unicom, China Mobile, and China Education and Research Network (CERNET), as well as many small and medium-sized carriers. At least 100 Tbit/s of bandwidth is reserved for response to traffic bursts, and bandwidth expansion is not limited. CDN precisely schedules user requests to the most appropriate edge nodes, providing efficient and reliable acceleration.



Nodes in the Chinese mainland

## Node Distribution Outside the Chinese Mainland

- 500+ nodes across over 70 countries and regions, international private lines, and Tbit/s-level redundant bandwidth.



## Advantages of CDN

### Global Presence

HUAWEI CLOUD CDN has over 2,000 nodes in the Chinese mainland and over 500 nodes outside the Chinese mainland. The network bandwidth is higher than 100 Tbit/s.

### Intelligent Scheduling

- Accurate and evolving global IP geolocation database
- Dynamic adjustment of nodes to deliver cache to users based on real-time analysis

### Security

- Secure and reliable content delivery services
- Advanced network security capabilities throughout the network, such as data transmission over HTTPS and hotlink protection

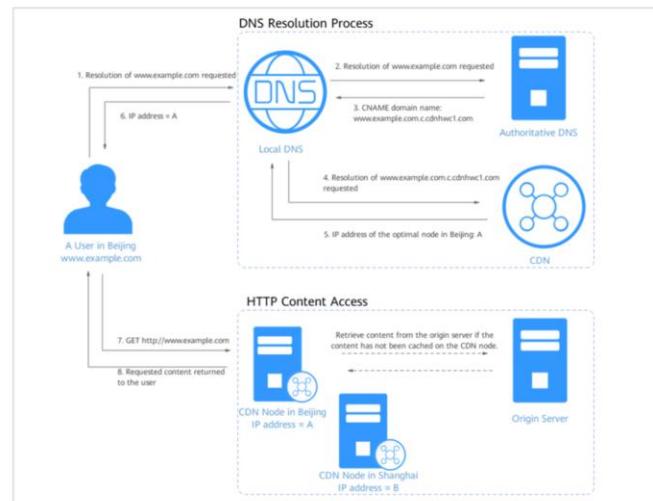
### Ease of Use

You can manage your domain names and logs, customize configurations (such as cache policies), and analyze domain data on the easy-to-use CDN console.

### Reliability

One-stop acceleration, including website, download, video, and whole site acceleration, meeting a wide range of requirements

## How Does CDN Work?

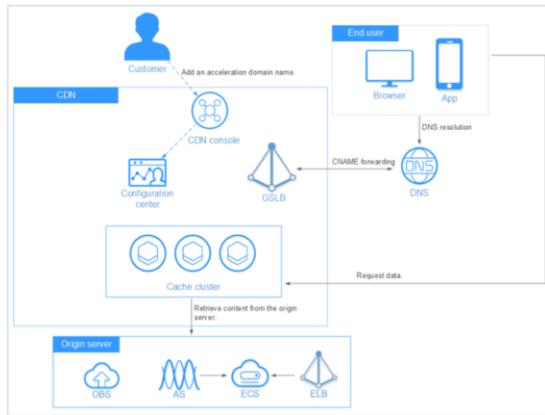


66      Huawei Confidential



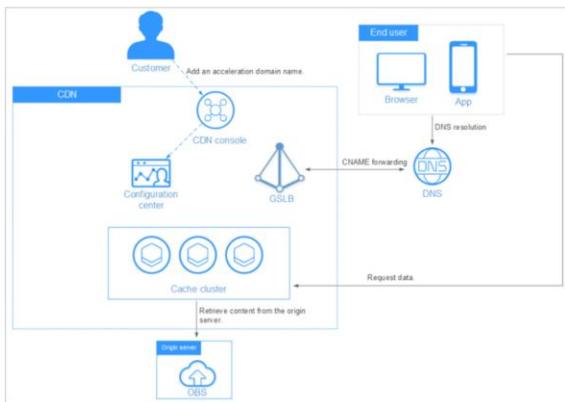
- The process is as follows:
  - A user enters the domain name of a website (for example, `www.example.com`) in the browser. A DNS request is sent to the local DNS server.
  - The local DNS server checks whether it caches the IP address of `www.example.com`. If yes, the local DNS server directly returns the cached IP address to the user. If no, the local DNS server sends a resolution request to the authoritative DNS server.
  - The authoritative DNS server resolves the domain name and finds that the domain name points to `www.example.com.ccdnhwcl.com` (CNAME record of the domain name).
  - The request is directed to CDN. CDN performs intelligent domain resolution and provides the user with the IP address of the optimal CDN node, for example, the Beijing CDN node, which responds the fastest.
  - The user's browser obtains the IP address of the Beijing CDN node.
  - The user's browser sends the access request to the Beijing CDN node.
    - If the Beijing CDN node has cached the requested content, it directly sends the content to the user and ends the request.
    - If the Beijing CDN node has not cached the content, it retrieves the content from the origin server. The retrieved content is cached on the Beijing CDN node based on custom cache policies. Then, the Beijing CDN node sends the content to the user and ends the request.

## Application Scenarios - Website Acceleration



- Website Acceleration
  - CDN is perfect for web portals, e-commerce platforms, news apps, and user generated content (UGC)-focused apps. It provides excellent acceleration for static content under an acceleration domain name. In addition, it supports custom cache policies. You can set the maximum cache age as needed.
- Advantages
  - Quick configuration: Domain names can be configured in just six simple steps.
  - Secure acceleration: HTTPS and referer validation ensure high security.
  - Flexible configuration: Content can be cached permanently or temporarily, or not cached.
- CDN can be used together with OBS, ECS, and DNS to build an E2E solution.

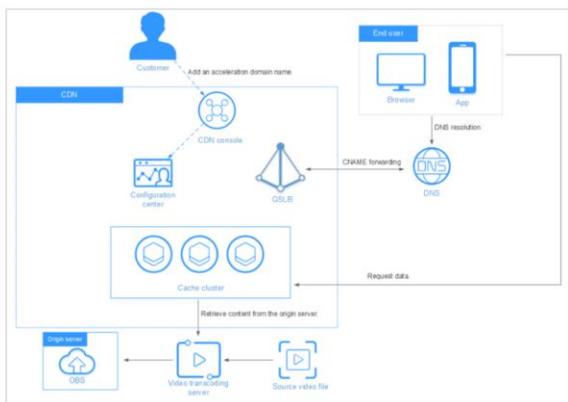
## Application Scenarios - Download Acceleration



- Download Acceleration
- CDN is useful for download clients, game clients, app stores, websites that provide download services based on HTTP or HTTPS, and apps that require updates in real time, such as mobile games.
- Advantages
  - Real-time analysis: Log monitoring and statistical analysis are performed in real time.
  - Reliability: HTTPS acceleration and referer validation ensure high security.
  - Cost-effectiveness: CDN interworks with OBS to further enhance performance and reduce costs.
- CDN can be used together with OBS and DNS to build an E2E solution.

- Conventional download services need to process a large number of download requests and requests for downloading large files. If origin servers have to handle all these requests directly, the servers will be overloaded. With CDN download acceleration, content to be downloaded is distributed to edge nodes, easing the pressure on origin servers and ensuring high-speed downloads.

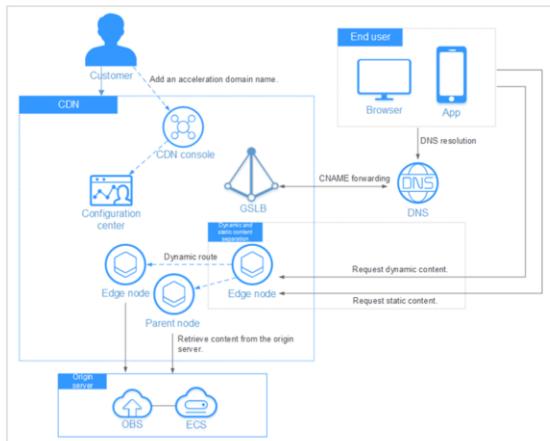
## Application Scenarios - VOD Acceleration



- **VOD Acceleration**
- CDN is a must if you intend to provide on-demand audiovisual services or live streaming services over the HTTP Live Streaming (HLS) protocol. Such services include online education, video sharing, and music or video on demand.
- **Advantages**
  - Real-time monitoring: Data such as traffic and bandwidth generated is displayed in CDN in real time.
  - Security: Referrer validation protects copyrighted images from being used.
  - Flexible configuration: Content can be cached permanently or temporarily, or not cached.
- CDN can be used together with OBS and DNS to build an E2E solution.

- Conventional audiovisual on-demand services cause heavy load on servers and consume enormous amounts of bandwidth. Low speed negatively affects user experience. CDN provides fast, reliable, and secure acceleration for audiovisual on-demand services by delivering the content to all edge nodes, so that users can obtain that content from the nearest node anywhere and anytime.
- CDN supports the video formats MP4, HLS, and FLV and audio formats MP3, ACC, OGG, and FLAC.

## Application Scenarios - Whole Site Acceleration



- Whole Site Acceleration
- CDN is a good option for websites that consist of both dynamic and static content, and for websites that involve a large number of ASP, JSP, or PHP requests.
- Advantages
  - Separation of dynamic and static content: Dynamic and static content is accelerated separately.
  - Secure acceleration: HTTPS and referer validation ensure high security.
  - Sequential retrieval: If the number of content retrieval requests to an origin server increases sharply, you can set a threshold. Once the threshold is exceeded, the retrieval requests are queued for response based on the time the requests are sent.
- CDN can be used together with OBS, ECS, and DNS to build an E2E solution.

- CDN's whole site acceleration accelerates both dynamic and static content. Static content is obtained from nearby nodes, whereas dynamic content is retrieved from the origin server through the fastest possible route. Congested routes are bypassed to load dynamic pages more quickly.

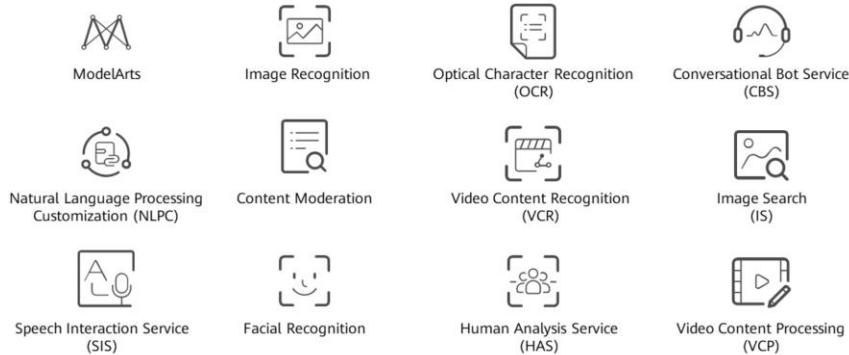
# Contents

1. Database Services
2. Security Services
3. Content Delivery Network (CDN)
- 4. EI Services**



## Huawei EI Service Panorama - Artificial Intelligence

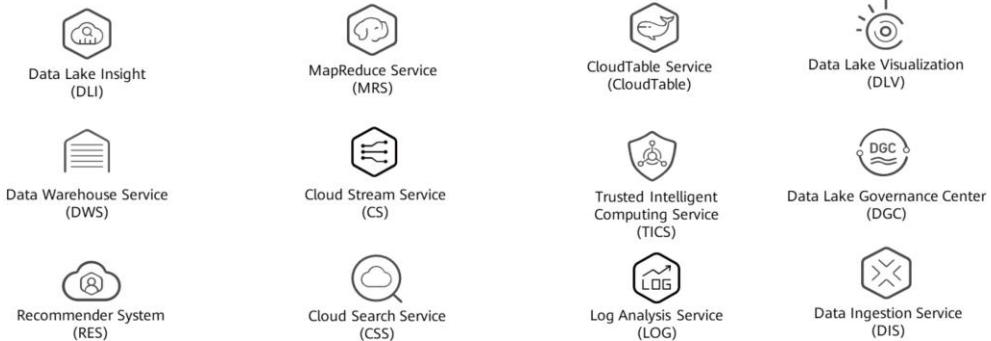
- HUAWEI CLOUD provides comprehensive AI and big data cloud services to facilitate the intelligent upgrades of governments and enterprises and build ubiquitous and pervasive AI.



- HUAWEI CLOUD EI includes AI services and big data services. This slide introduces the former.

## HUAWEI CLOUD EI Service Panorama - Big Data

- HUAWEI CLOUD provides comprehensive AI and big data cloud services to facilitate the intelligent upgrades of governments and enterprises and build ubiquitous and pervasive AI.



- HUAWEI CLOUD EI includes AI services and big data services. This slide introduces the latter.

## One-Stop AI Development Platform ModelArts

- ModelArts is a one-stop AI development platform. For machine learning and deep learning, it supports data preprocessing, semi-automated data labeling, distributed training, automated model building, and on-demand deployment of device-edge-cloud models. ModelArts helps AI developers build and deploy models quickly and manage the lifecycle of AI workflows.



**ModelArts 3.0**

Intelligent sensing, cognition, and decision-making



**ModelArts Pro**

World's first enterprise-grade AI application development suite

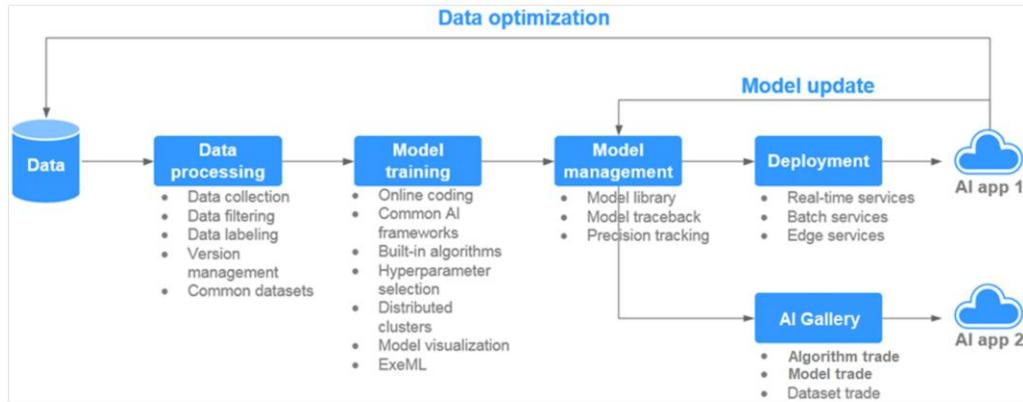


**Knowledge Compute**

New path integrating industry expertise with AI

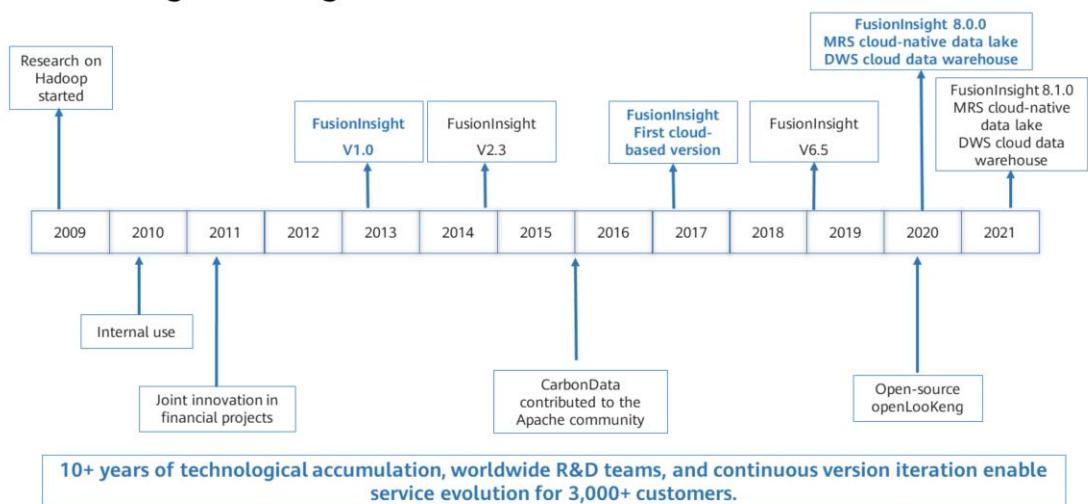
- With data preprocessing, semi-automated data labeling, distributed training, automated model building, and model deployment on devices, edge devices, and HUAWEI CLOUD, ModelArts helps AI developers quickly build and deploy models and easily manage the AI development lifecycle. ModelArts runs through all phases of AI development, including data processing, algorithm development, model training, and model deployment. The underlying technologies support a wide range of heterogeneous compute resources, allowing developers to flexibly select and use resources. In addition, ModelArts supports popular open-source AI development frameworks such as TensorFlow and MXNet. Developers can also use self-developed algorithm frameworks to match their usage habits.
- ModelArts simplifies AI development. It provides convenient and easy-to-use processes for AI developers with different levels of experience. For example, service developers can use ExeML to quickly develop AI applications without considering models or code; AI beginners can use built-in algorithms to develop AI applications without considering model development; AI engineers can use multiple development environments and operational processes and modes to extend code and quickly build models and applications.

## Functions of ModelArts

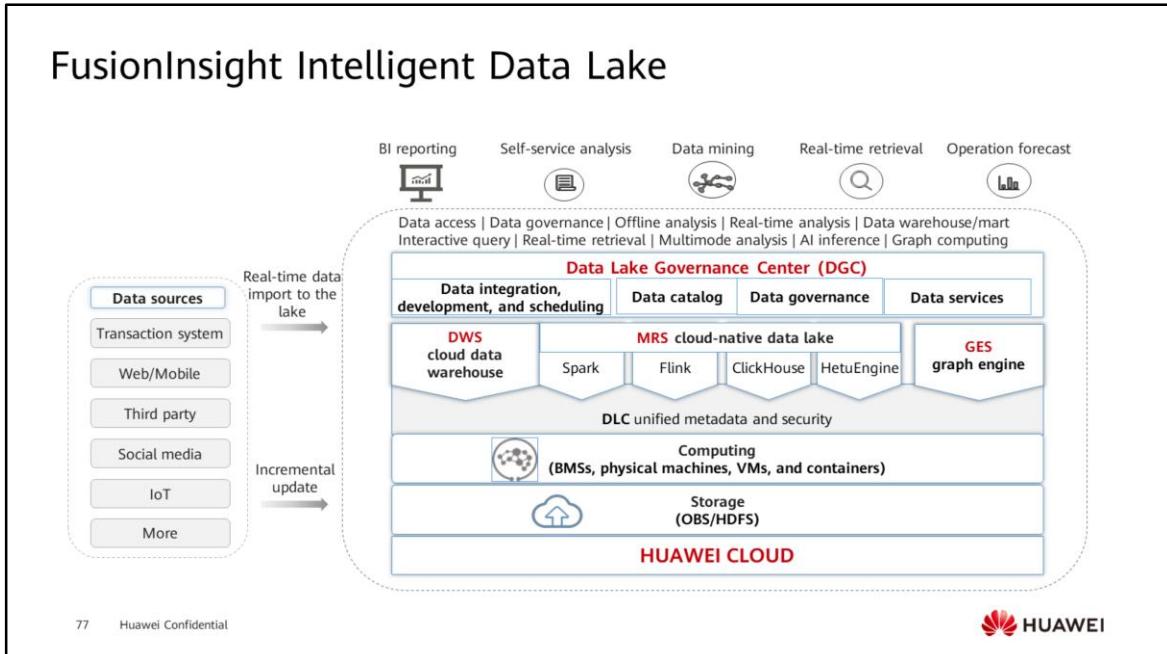


- ModelArts supports the entire AI application development process, including data processing, and model training, management, and deployment. It also provides AI Gallery for sharing models.
- ModelArts supports various AI application scenarios, such as image classification, object detection, video analysis, speech recognition, product recommendation, and exception detection.

## FusionInsight Intelligent Data Lake - Milestones



# FusionInsight Intelligent Data Lake

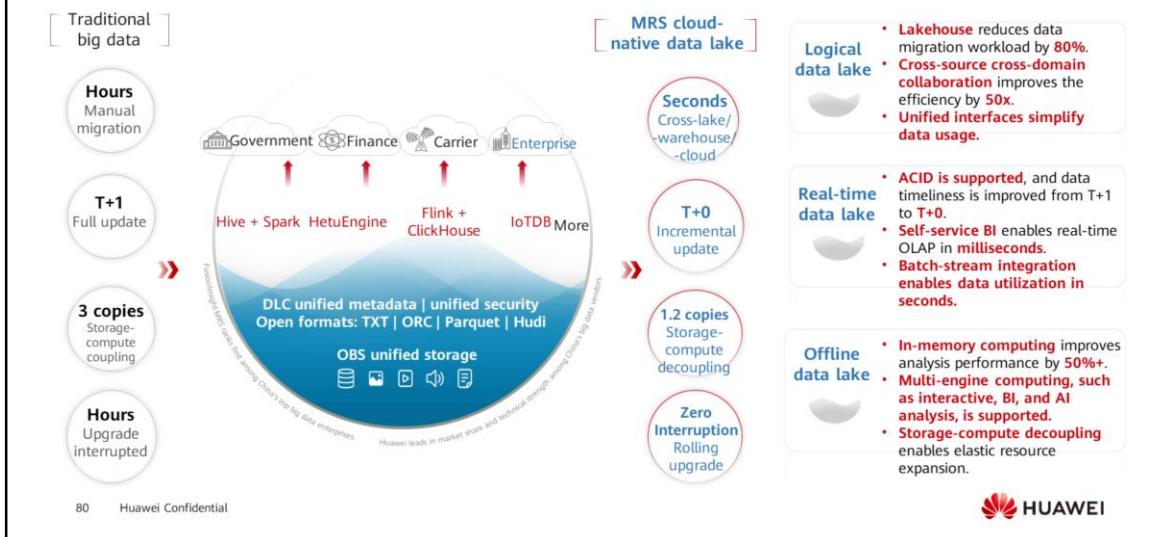


77      Huawei Confidential



- The FusionInsight solution provides cloud services such as the big data service MRS, Data Warehouse Service (DWS), Cloud Search Service (CSS), Graph Engine Service (GES), Data Lake Insight (DLI), and Data Lake Governance Center (DGC). It supports big data applications such as real-time analysis, offline analysis, interactive query, real-time retrieval, multimode analysis, data warehouse/mart, data access, and data governance of customers' full data. As a one-stop solution, it solves data issues in the analysis domain and unleashes the value of massive data, helping customers build one lake for one enterprise and one lake for one city.
- MRS builds three cloud-native data lakes with one architecture that is continuously evolving.
  - Logical data lake: enables collaboration across lakes, warehouses, and clouds, improving efficiency by over 30%.
  - Real-time data lake: supports millisecond Online Analytical Processing (OLAP), improving timeliness from T+1 to T+0.
  - Offline data lake: allows warehouses to be built within a lake to shorten analysis links, boosting analysis efficiency by over 10 times.
- DWS has the following features:
  - High performance: It outperforms competitors' products, such as Alibaba AnalyticDB and GBase 8a, in performance comparison tests.
  - High scalability: A cluster can contain 2,048 nodes and over 100 PB data.

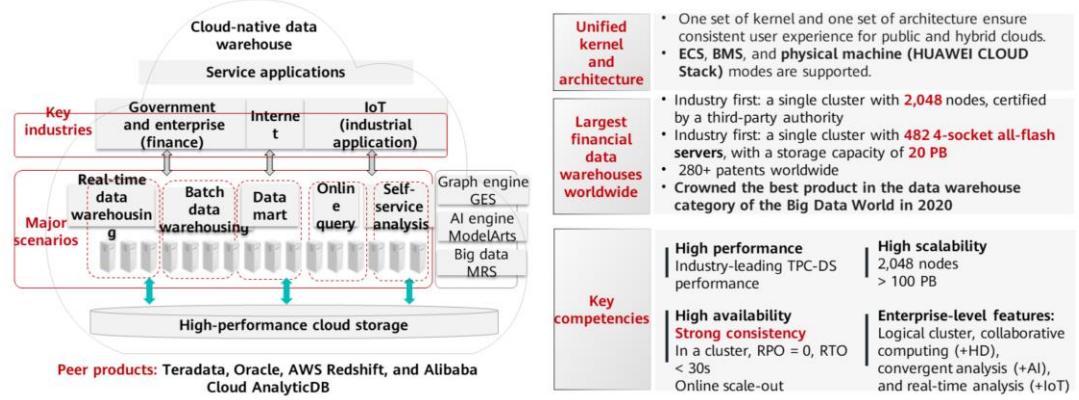
# MRS: Three Cloud - Native Data Lakes, One Architecture



- HUAWEI CLOUD FusionInsight MRS cloud-native data lake has evolved in the big data field in the last 10 years. It now provides capabilities of three data lakes with just one architecture.
  - For a logical data lake, collaborative analysis across lakes, warehouses, and clouds enabled by HetuEngine eliminates data silos and reduces data migration by more than 80%. It helps group enterprises deal with scattered data. For example, Industrial and Commercial Bank of China (ICBC) often requires cross-domain manual migration because it has more than 200 headquarters and branches. A logical data lake can be used to migrate data logically, reducing physical data migration.
  - A real-time data lake supports millisecond OLAP, improving timeliness from T+1 to T+0.
  - An offline data lake supports multi-engine computing, including interactive, BI, and AI analysis. Storage-compute decoupling enables unified data storage, and on-demand expansion of compute and storage resources, reducing TCO by 60%.
- FusionInsight MRS allows rolling upgrade of large clusters in batches and enables isolation of faulty nodes, ensuring smooth upgrades and service continuity. Take an application in the financial industry as an example. Rolling upgrade does not interrupt services at bank counters.

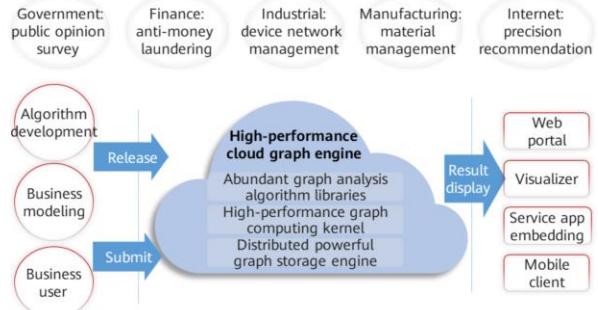
# GaussDB (DWS): Next-Generation Cloud Data Warehouse

- GaussDB(DWS) is a distributed database for data analysis and **hybrid transaction/analytical processing**. It supports both x86 and Kunpeng architectures and **row and column storage**, with the capabilities of **PB-level** data analysis, multi-mode analysis, and **real-time processing**. GaussDB(DWS) spans across the core systems of industries such as finance, government, and telecom.



- GaussDB(DWS), based on the GaussDB kernel, is the next-generation, all-scenario, cloud data warehouse of Huawei.
- It supports both x86 and Kunpeng architectures and row and column storage. It has the capabilities of PB-level data analysis, multi-mode analysis, and real-time processing, helping users make informed decisions in real time.

# GES: Integrated Graph Analysis and Querying

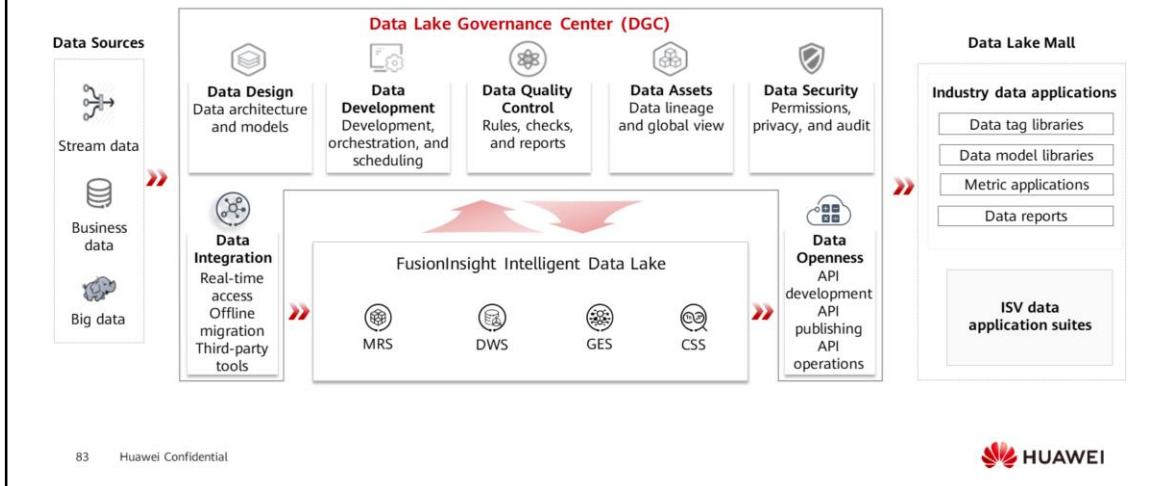


- 1 One-stop graph database and engine**
  - Integrated graph database and graph analysis engine
  - Comprehensive graph analysis and query capabilities provided through user-friendly GUI
  - China's first commercial native graph product with proprietary intellectual property rights
- 2 Integrated analysis and query**
  - A set of data is used for two purposes: query and analysis.
  - Mainstream graph query languages, Cypher and Gremlin, are supported, and native REST APIs and open-source APIs are available.
  - Over 30 high-performance algorithms are used for analysis and compute in multiple scenarios. More than 10 graph neural networks and graph embedding algorithms are provided.
- 3 Large scale and high performance**
  - Graphs with **over 10 billion vertices and 100 billion edges**
  - The query and algorithm performance is better than that of competitors in the industry. The 6-hop query response is within seconds. Many algorithms are excellent in large graph compute.
- 4 No-code visual analysis makes the GES easy to use**
  - Editing and entity drill-down are made simple with the intuitive GUI.
  - Wizard-based algorithm operations can be performed on the GUI, and the operation results and analytics are represented in an intuitive manner.
- 5 Huawei-developed kernel that has won international awards for multiple times**



- GES is a graph database technology that supports large-scale graphs with tens of billions of vertices and trillions of edges. The Huawei-developed kernel has won numerous international awards, such as the highest award at CAIS 2019 and the Excellent AI application award at IFTC 2020.

# DGC: One-Stop Data Development and Integration Management for 3x Higher Efficiency in Data Assetization



- DGC provides a full set of data governance functions, such as data integration, data design, data development, and data quality control, to generate and open data assets to upper-layer services.
- With DGC, enterprises can easily provide data to their industry partners and accumulate their own data assets.
- DGC is an internal product name used in Huawei. DGC and the FusionInsight Intelligent Data Lake are provided to users with the DAYU brand.
- Platform:
  - One-stop data development and integration management platform, 40+ heterogeneous data sources, development with simple drag-and-drop, multi-dimensional real-time search, and zero-code or low-code API development, achieving three times higher development efficiency
  - Cloud services developed for data architecture, data standards and specifications, data development, and data quality control based on Huawei's 10+ years' experience in data governance
- Ecosystem:
  - 100+ open APIs enable industry ISVs to implement quick integration and development.
  - 10+ partners provide industry data standards, models, metrics, and APIs.

## Quiz

1. CDN is a free cloud service.  
True  
False
2. Which of the following are the application scenarios for HUAWEI CLOUD CDN?  
A. Website acceleration  
B. File download acceleration  
C. VOD acceleration  
D. ECS running acceleration

- False. CDN is billed by traffic or bandwidth.
- ABC. CDN mainly accelerates applications. It cannot accelerate cloud servers.

# Summary

This course introduces database services, security services, CDN, and EI services of HUAWEI CLOUD, including:

- Relational and non-relational database types, and the application scenarios and key features of different databases.
- Basic concepts and importance of security services.
- Functions and working rules of the CDN and Enterprise Intelligence (EI) services.

After completing this course, you will have a comprehensive understanding of HUAWEI CLOUD and can better help enterprises accelerate cloud migration and business innovation.

# Recommendations

- Huawei Learning Website
  - <https://e.huawei.com/en/talent/#/>
- HUAWEI CLOUD Technical Support
  - <https://support.huaweicloud.com/intl/en-us/help-novice.html>
- HUAWEI CLOUD Academy
  - <https://edu.huaweicloud.com/intl/en-us/>

## Acronyms and Abbreviations

- AZ: availability zone
- APP: application
- API: application programming interface
- APT: advanced persistent threat
- CDN: content delivery network
- CPU: central processing unit
- CSA: cloud security alliance
- DDoS attack: distributed denial-of-service attack
- DDS: document database service
- DDM: distributed database middleware

## Acronyms and Abbreviations

- DAS: data admin service
- DWS: data warehouse service
- DEW: data encryption workshop
- EI: enterprise intelligence
- ELB: elastic load balance
- HA: highly available
- HSS: host security service
- IT: Internet technology
- IAM: identity and access management
- KMS: key management system

## Acronyms and Abbreviations

- LAMP: Linux+Apache+PHP+MySQL (a set of open-source software usually used to build dynamic websites)
- OLAP: online analytical processing
- OLTP: online transaction processing
- OBS: object storage service
- PITR: point-in-time recovery
- RTO: recovery time object
- UGC: user generated content
- VIP: virtual IP address
- WAF: web application firewall

# Thank you.

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

Copyright©2021 Huawei Technologies Co., Ltd.  
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.



# HUAWEI CLOUD O&M Basics



# Foreword

- HUAWEI CLOUD not only provides resource services to meet enterprise needs to migrate their service systems to the cloud, but also ensures the normal running of the service systems on the cloud to meet the enterprise governance requirements.
- This section will help you understand HUAWEI CLOUD O&M.

# Objectives

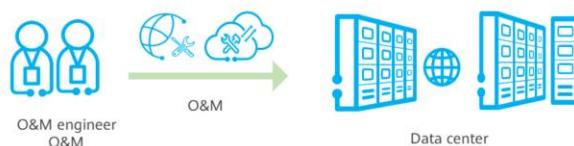
- Upon completion of this course, you will:
  - Gain basic knowledge about O&M, monitoring, and auditing.
  - Understand the positioning, principles, and usage of common governance services on HUAWEI CLOUD.

# Contents

- 1. O&M Basic Concepts and Principles**
2. Cloud Eye
3. Log Tank Service (LTS)
4. Cloud Trace Service (CTS)

## What Is O&M?

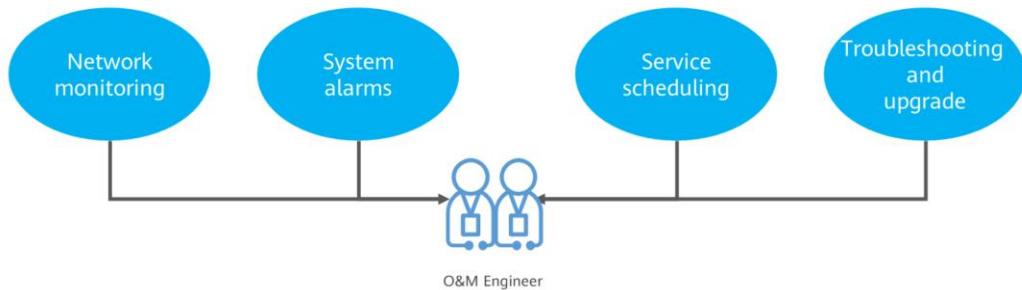
- O&M refers to operations and maintenance. It includes monitoring and managing devices and service systems to ensure services run normally. O&M also includes handling various problems and summarizing maintenance experiences to improve O&M efficiency and quality.
- O&M is essentially the operations and maintenance of devices and services such as servers and networks in each phase of their lifecycles, to achieve an optimum level of cost, stability, and efficiency.



- O&M focuses on various environments where the service system runs. It does not focus on programming, but on the use and management of these system platforms.
- In the ICT industry, those who perform O&M operations are typically referred to as O&M engineers.

## Responsibilities of O&M Personnel

- O&M personnel are responsible for planning information, networks, and services based on service requirements and ensuring the long-term stability and availability of services by using various means, including but not limited to the following:



- O&M personnel stabilize the infrastructure, basic services, and online services that the enterprise Internet services rely on, perform routine inspection to detect potential risks, optimize the overall architecture to prevent common operation failures, and connect multiple data centers to improve the DR capability of services. By using technical means such as monitoring and log analysis, O&M personnel can detect and respond to service faults in a timely manner to minimize service interruption and meet enterprise availability requirements for Internet services.
- To be an excellent O&M engineer, one needs to have comprehensive technical knowledge and troubleshooting experience, and have a strong sense of responsibility for their work.
- In the common organizational structure of the Internet industry, O&M, development, and testing are basic technical positions. In terms of the phase, development and testing personnel are engaged in the work before software or services are launched, while O&M (except O&M development) personnel are engaged in the work after the software or services are launched. O&M can be further classified into IT operations, network O&M, service O&M, and O&M development.

## Classification of O&M Personnel

- As the number and complexity of devices, operating systems, and applications deployed in ICT data centers increase, enterprises have higher requirements on O&M. This calls for specialized O&M. Common O&M positions are as follows:

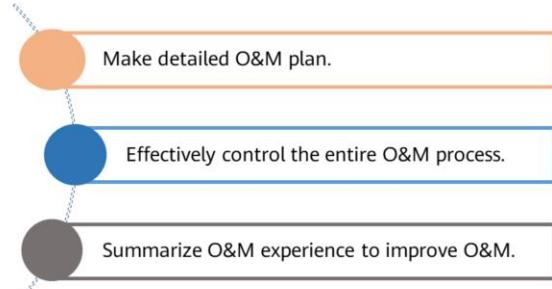


- Hardware O&M:
  - Equipment room planning (including equipment room location selection, network deployment, and server deployment planning)
  - Network system maintenance (including network adjustment, capacity expansion, bandwidth monitoring, and network QoS)
  - Server management (procurement, receipt, deployment in cabinets, system installation, delivery, and maintenance)
- System O&M:
  - O&M based on OS usage includes system optimization and performance monitoring
- Database O&M:
  - Software installation, configuration optimization, backup policy selection and implementation, data restoration, data migration, troubleshooting, preventive inspection, and other services for user databases
- Application O&M: mainly refers to the O&M of services used by users to ensure the stability of services in the case of continuous iteration.
  - Change management, ensuring system stability in the process of continuous iteration
  - Fault management, including application monitoring, fault locating, fault rectification, and application optimization
  - Resource management: ensuring that the application system runs properly with optimal resources, and evaluating whether capacity expansion is

required to meet future service needs.

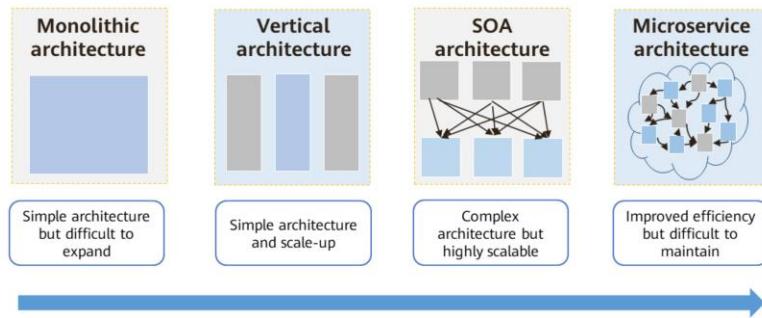
## ICT O&M Principles

- We have learned about various O&M positions. How should ICT O&M personnel go about their jobs? O&M seems simple. However, to better serve enterprise business systems, we must first understand the overall principles of ICT O&M.



## O&M Challenges Brought About by Evolution of IT Architecture

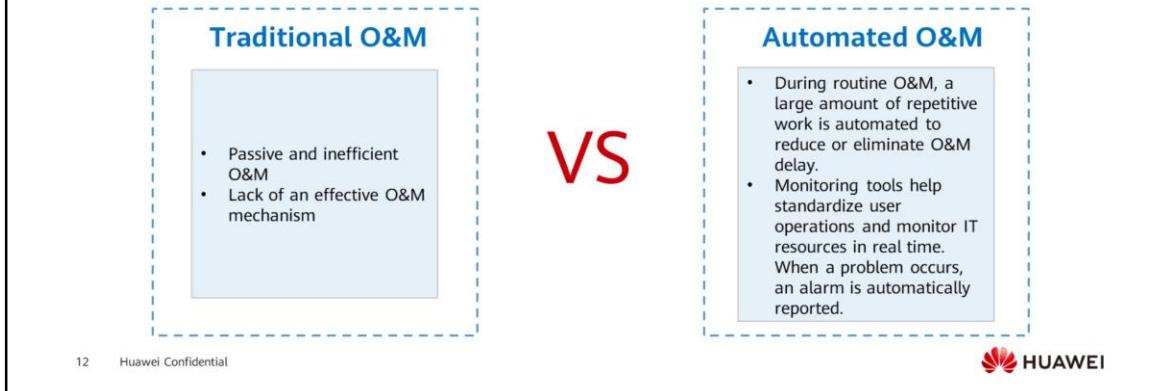
- The IT architecture becomes more and more complex, and O&M personnel face many challenges.



- The IT architecture evolves more and more complex. In an enterprise, development and O&M are usually two independent departments with different work objectives and technical directions. When a project needs to be completed by the two departments, their communication is not smooth, the progress is delayed, and the enterprise efficiency is greatly reduced. Therefore, the entire system architecture needs to evolve continuously, moving from traditional O&M to automatic O&M. This will help break down the barriers between O&M engineers, development engineers, and quality assurance engineers, and form an efficient work system.
- Characteristics of the monolithic architecture: All functions are integrated in one project. The architecture is simple, the development cost in the early phase is low, and the development period is short. Therefore, this architecture is ideal for small-scale projects. However, as the small projects grow larger, it is difficult to develop, expand, and maintain the monolithic architecture.
- Characteristics of the vertical architecture: Projects using the monolithic architecture are vertically divided. The project architecture is simple, the development cost in the early phase is low, and the development period is short. The vertical architecture is ideal for small-scale projects. Vertical splitting ensures that small projects cannot become too large.
- Characteristics of the SOA (Service-Oriented Architecture): Repeated common functions are extracted as components to provide services for each system. Projects (or systems) communicate with services through WebService or remote procedure call (RPC). The SOA architecture improves the development efficiency, and the system reusability and maintainability. Cluster and architecture optimization solutions can be formulated based on the characteristics of different services.

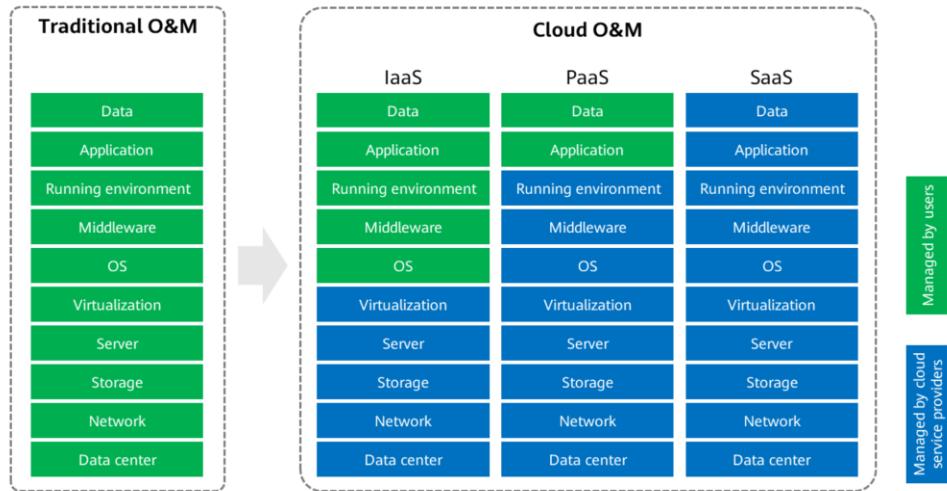
## Era of Automated O&M

- Traditional manual O&M is being gradually replaced by automated O&M platforms. Responsibilities of O&M personnel and development personnel are converging. The concept of integrated O&M and development (DevOps) is becoming more and more popular and is being used by most enterprises.



- After more than a decade of development, IT operations is now facing a new direction: automation, which is an inevitable result of IT technology development. Nowadays, the complexity of IT systems requires digital and automated O&M. Automated O&M refers to the automation of daily and repeated work in IT operations and the transformation from manual work to automation. Automation is the sublimation of IT operations. IT operations automation is not only a maintenance process, but also a management improvement process. It is the highest level of IT operations and also the development trend in the future.
- DevOps is a group of processes, methods, and systems that are used to promote communication, collaboration, and integration between development, technical operation (O&M), and quality assurance (QA) departments. DevOps greatly reduces the gap between O&M and development and the delivery time.

## O&M Changes in the Cloud Era

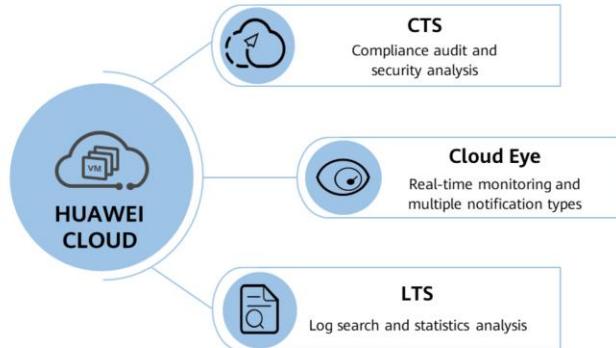


13      Huawei Confidential



- Compared with traditional O&M, cloud O&M greatly reduces the enterprise O&M costs. The O&M management services provided on the public cloud enable users to complete routine O&M at little or no cost. All these services are based on automatic O&M technologies.

## Common O&M Services on HUAWEI CLOUD

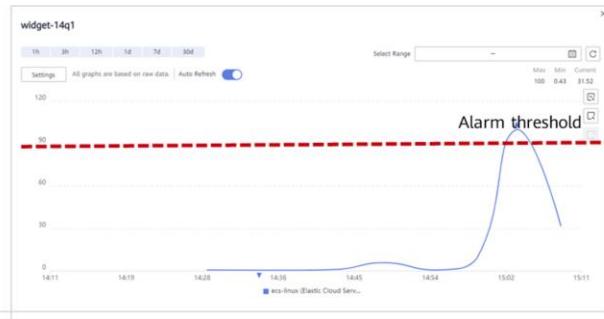


# Contents

1. O&M Basic Concepts and Principles
- 2. Cloud Eye**
3. Log Tank Service (LTS)
4. Cloud Trace Service (CTS)

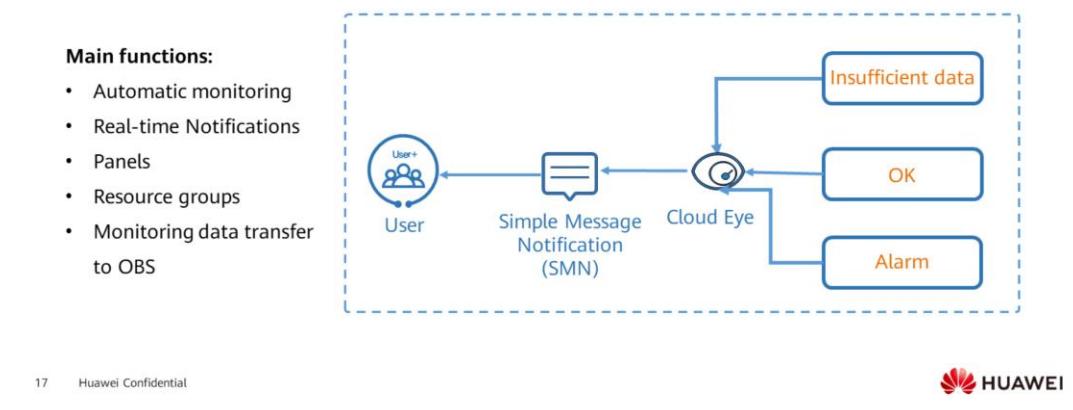
## Why Is Monitoring Required?

- Monitoring helps identify potential risks. Through monitoring, we can learn about the running status of the enterprise network. Once a security risk is detected, O&M personnel can be informed of the risk in a timely manner, so that they have time to mitigate the risk. This prevents the service system from being affected and resolves issues at the earliest.



## What Is Cloud Eye?

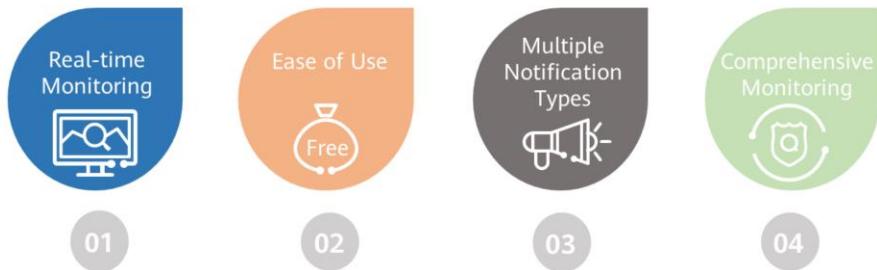
- **Cloud Eye** is a multi-dimensional monitoring service. You can use Cloud Eye to monitor resources, set alarm rules, identify resource exceptions, and quickly respond to resource changes.



- Cloud Eye provides the following functions:
  - Automatic monitoring: Monitoring starts automatically after cloud resources such as Elastic Cloud Servers (ECSs) or Auto Scaling (AS) groups are created. After you deploy a cloud service, you can view its running status and set alarm rules on the Cloud Eye console.
  - Real-time notification: You can enable Simple Message Notification (SMN) when creating alarm rules. When the cloud service status changes and metrics reach the thresholds specified in the alarm rules, Cloud Eye notifies you by text messages, emails, or by sending HTTP/S messages to servers. In this way, you can monitor the cloud resource status and changes in real time.
  - Panels: Panels enable you to view cross-service and cross-dimension monitoring data. Panels display key metrics centrally, providing an overview of the service operating status and allowing you to check monitoring details when troubleshooting.
  - Resource group: A resource group allows you to add and monitor correlated resources and provides a collective health status for all resources that it contains.
  - OBS dump: Raw data for each metric is kept for only two days on Cloud Eye. If you need to retain data for longer, enable Object Storage Service

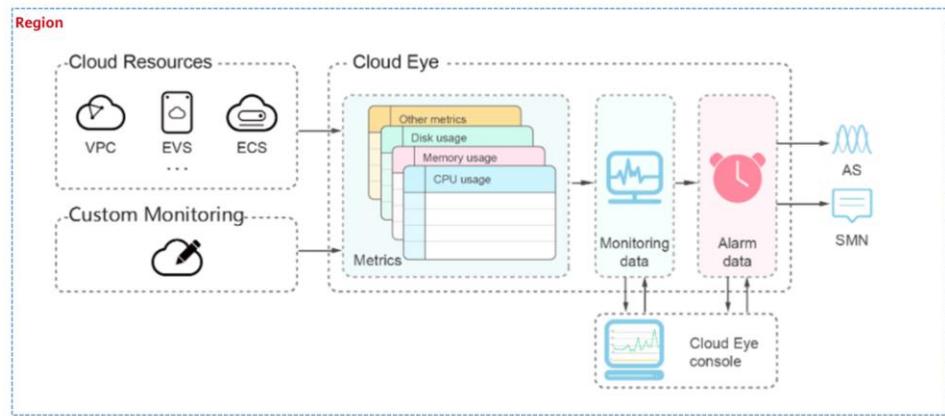
(OBS), and raw data will be automatically synchronized to OBS and saved.

## Cloud Eye Advantages



- Real-time monitoring: Cloud Eye visualizes metrics of cloud resources and services in real time.
- Ease of use: Cloud Eye allows you to easily analyze metrics aggregated using different methods. You can view metrics data for up to six months.
- Multiple notification types: When a metric reaches the threshold, Cloud Eye notifies you by emails or text messages, allowing you to keep track of the running status of cloud services. Cloud Eye can also send HTTP/HTTPS messages to an IP address of your choice.
- Comprehensive monitoring: You can derive actionable insights from detailed metrics for your cloud servers. In addition, you can use the Agent, APIs, and SDKs to monitor custom metrics.

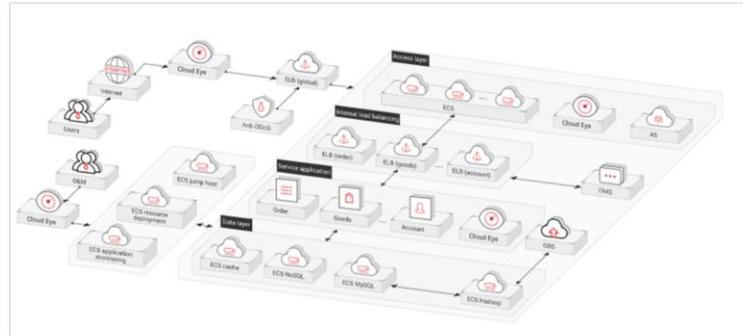
# Cloud Eye Architecture



# Application Scenario - E-Commerce Websites

## Characteristics of E-Commerce Websites

- Editable Monitoring Panel**  
Provides you with a comprehensive view of key system monitoring information.
- Alarm-Triggered Scaling**  
The system automatically expands or reduces capacity based on the service traffic volume and configured alarm rules.
- Comprehensive Server Monitoring**  
Enables monitoring of customized network traffic metrics to prevent network bottlenecks.

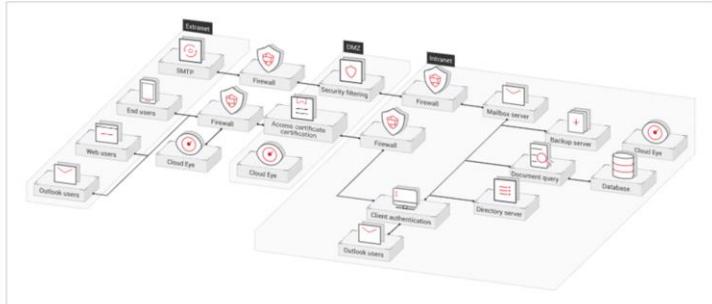


- In the solution shown in the figure, Cloud Eye effectively monitors the service running status and works with other services to dynamically adjust resources. Cloud Eye can also monitor the status of cloud services, networks, and applications and send alarm notifications.

## Application Scenario - Enterprise Offices

### Characteristics of Enterprise Offices

- Alarm-Triggered Expansion**  
ECS expansion is automatically triggered by alarms that are generated if ECS usage reaches the configured threshold.
- Log Monitoring**  
Logins are monitored in real time and malicious login requests are rejected to ensure security.
- Comprehensive Server Monitoring**  
Enables monitoring of customized network traffic metrics to prevent network bottlenecks.



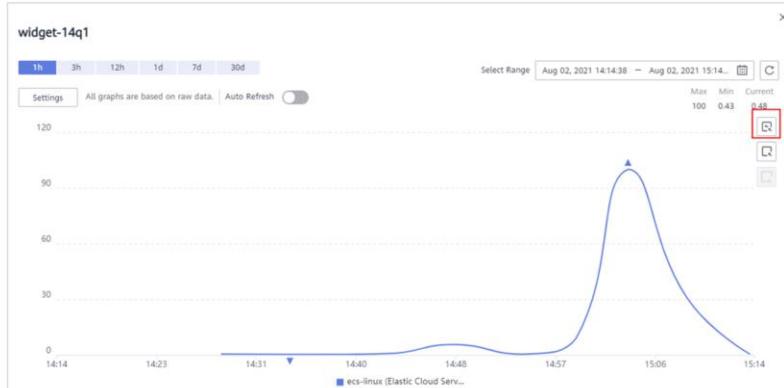
21 Huawei Confidential



- In the solution shown in the figure, Cloud Eye monitors the status of each service.

## Panels

- You can use panels to view core metrics and compare the performance data of different services.



22     Huawei Confidential



- Panels allow you to compare performance data of different services from different dimensions. You can create 20 panels, add 24 monitoring graphs to each panel, and add 20 items to each graph.

## Metrics

- This is the core concept of Cloud Eye. A metric refers to a quantitative value of a resource dimension on the cloud platform, such as the ECS CPU usage and memory usage. A metric is a time-dependent variable that generates a certain amount of monitoring data over time. It helps you understand the changes over a specific period of time.

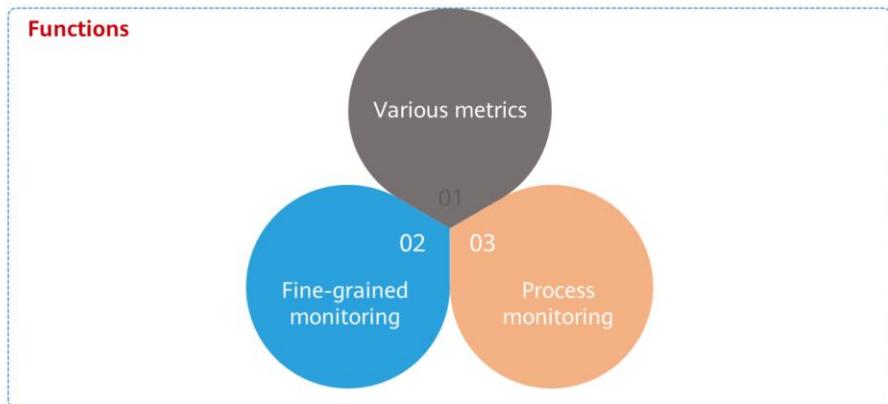


23      Huawei Confidential



## Server Monitoring

- Server monitoring comprises basic monitoring, OS monitoring, and process monitoring for servers.



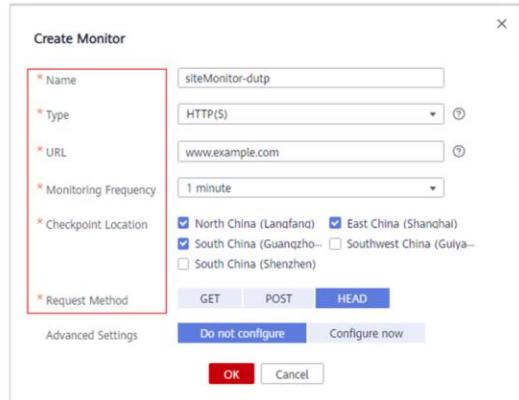
24      Huawei Confidential



- Server monitoring comprises basic monitoring, OS monitoring, and process monitoring for servers.
  - Basic monitoring provides Agent-free monitoring for basic ECS and BMS metrics.
  - OS monitoring provides fine-grained OS monitoring for servers, and it requires the Agent (a plug-in) to be installed on all servers that will be monitored.
  - Process monitoring is used to monitor active processes on hosts. By default, Cloud Eye collects CPU usage, memory usage, and number of opened files of active processes.
- Functions:
  - Server monitoring provides more than 40 metrics, such as metrics for CPU, memory, disk, and network, to meet the basic monitoring and O&M requirements for servers.
  - After the Agent is installed, data of Agent-related metrics is reported once a minute.
  - CPU usage, memory usage, and number of opened files used by active processes give you a better understanding of the ECS and BMS resource usages.

## Website Monitoring

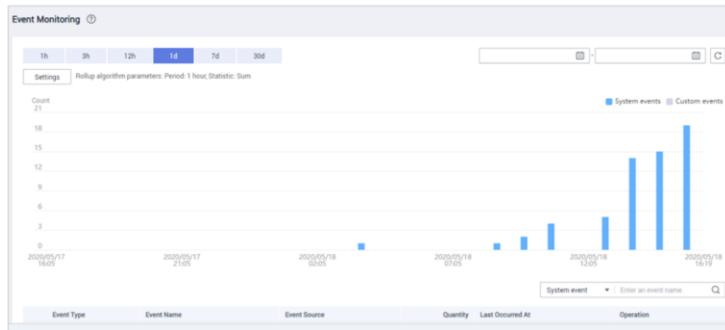
- Website monitoring is to continuously monitor remote server statuses, such as availability and connectivity, by simulating real users' access to remote servers.
- Website monitoring can detect availabilities of domain names and IP addresses, access response time, and packet loss rate, and generate alarms based on monitoring results.



- Website monitoring is free. The website monitoring function is available in the CN North-Beijing1 region. If you want to use this function in other regions, ensure you have the CES FullAccess permissions configured in project **cn-north-1 [CN North-Beijing1]**.
- Advantages:
  - You can create, modify, disable, enable, or delete monitors.
  - The configuration is simple and quick, allowing you to improve efficiency and save resources that you would otherwise use to configure complex open-source products.
  - You receive notifications of website exceptions in real time.

## Event Monitoring

- In event monitoring, you can query system and custom events reported to Cloud Eye through the API. You can create alarm rules for both system and custom events. When specific events occur, Cloud Eye generates alarms for them.



## Custom Monitoring

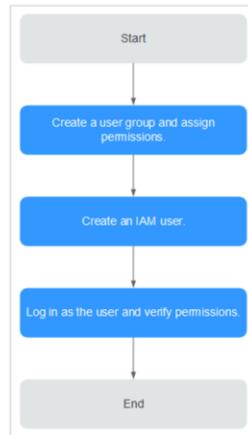
- The Custom Monitoring page displays all custom metrics reported by users. You can use simple API requests to publish collected monitoring data of those metrics to Cloud Eye for processing and display.



# Permissions Management

## Configuration:

- Create a user group on the IAM console, and assign the **CES Administrator**, **Tenant Guest**, and **Server Administrator** policies to the group.
- Create a user on the IAM console and add the user to the created group.
- Log in to the Cloud Eye console as the created user, and verify that the user only has the **CES Administrator** permission.

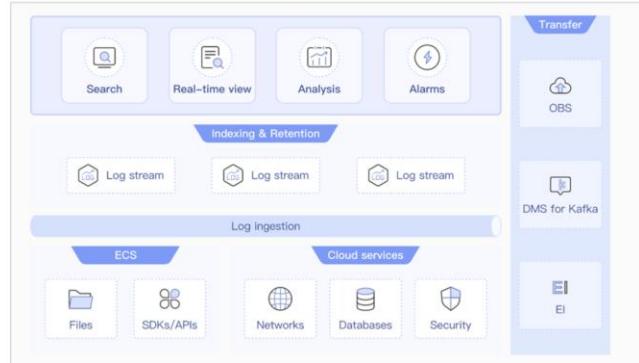


# Contents

1. O&M Basic Concepts and Principles
2. Cloud Eye
- 3. Log Tank Service (LTS)**
4. Cloud Trace Service (CTS)

## Why Logs Matter?

- Logs are files generated by system processes and record important system information. They provide useful details for fault location and program commissioning.



30      Huawei Confidential



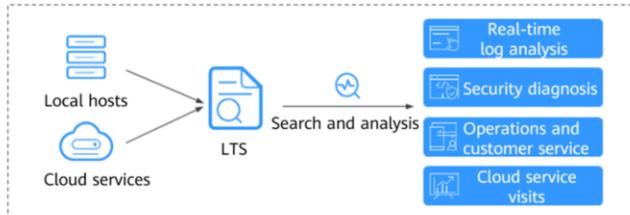
- O&M log platform:
  - LTS provides centralized log management. You can collect application and middleware logs distributed across your VMs and manage these large amounts of logs all in one place.
- Advantages:
  - Lifecycle management: You can manage logs throughout their lifecycle, from collection, search and analysis, to aging and archival.
  - Massive and fast: Hundreds of terabytes of logs can be ingested per day, and log search is fast. You can get responses in seconds even when there are gigabytes of logs involved.
  - Cost-effective: Pay-per-use billing allows you to adapt to changing demands without overcommitting budgets.

## What Is Log Tank Service?

- Log Tank Service (LTS) collects logs from hosts and cloud services for centralized management, and processes large volumes of logs efficiently, securely, and in real-time. LTS provides you with the insights needed for optimizing availability and performance of cloud services and applications. It allows you to make faster data-driven decisions, perform device O&M easily, and analyze service trends.

### Major functions:

- Real-time log collection
- Log query and real-time analysis
- Log monitoring and alarms
- Log transfer



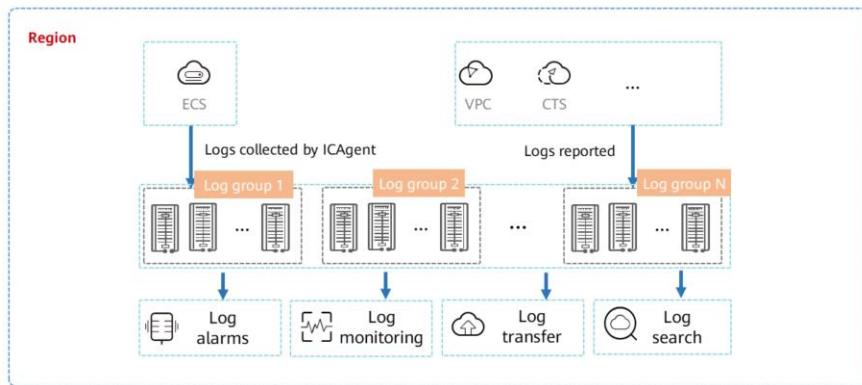
- Real-time log collection
  - LTS collects logs from hosts and cloud services in real time and displays them on the LTS console in an intuitive and orderly manner. You can query logs or transfer logs for long-term storage.
  - You can define log structuring rules so LTS will extract logs that are in a fixed format or share a similar pattern based on the rules. Then you can use SQL syntax to query the structured logs.
- Log query and real-time analysis
  - Collected logs can be quickly queried by keyword or fuzzy match. You can analyze logs in real time to perform security diagnosis and analysis, and obtain operations statistics, such as cloud service visits and clicks.
- Log monitoring and alarms
  - LTS works with Application Operations Management (AOM) to count the frequency of specified keywords in logs retained in LTS. For example, if the keyword ERROR occurs frequently, it can indicate that services are not running normally.
- Log transfer
  - Logs reported from hosts and cloud services are retained in LTS for 7 days by default. You can set the retention period to 1 to 30 days. Logs older than the retention period will be automatically deleted. For long-term storage, you can transfer logs to Object Storage Service (OBS) and Data Ingestion Service (DIS).

## LTS Advantages

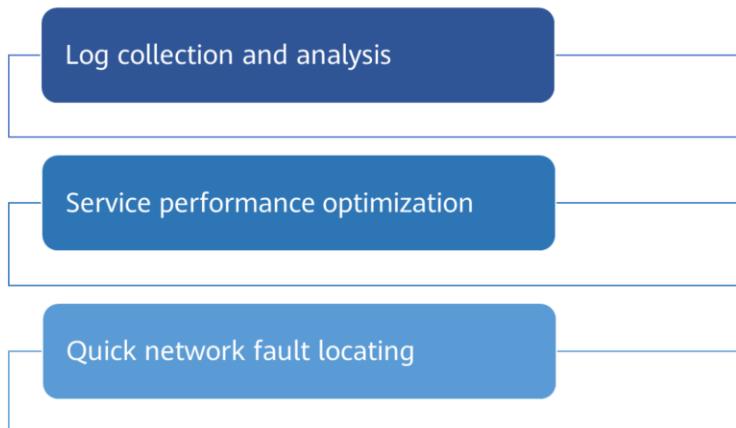


- One-stop management: You can easily manage your logs with this all-in-one platform. You can ingest, store, transfer, and search for logs in LTS. Advanced functions are also available for you to structure or visualize logs, or run SQL queries.
- Scale and speed: LTS allows you to manage petabytes of logs with ease. It can ingest logs at up to 200 TB/day. You can obtain results in seconds even when searching gigabytes of logs, or when running SQL aggregate queries in a significantly large number of logs.
- Secure and reliable: LTS keeps your data secure with HTTPS encryption and rights- and domain-based control. It has an availability of 99.95%.
- Cost-effective: LTS helps you save on maintenance, and there are no upfront commitments. You can scale up resources at any time to meet spikes in log volume, but you only pay for what you use.

## LTS Architecture



## LTS Scenarios



- Log collection and analysis
  - Without proper management, there are too many logs of hosts and cloud services, which are difficult to query and are cleared periodically. Using LTS, collected logs are displayed on the console in a clear and orderly manner for fast query, and can be stored for a long time if necessary. Collected logs can be quickly queried by keyword or fuzzy match. You can analyze logs in real time to perform security diagnosis and analysis, and obtain operations statistics, such as cloud service visits and clicks.
- Service performance optimization
  - Performance and quality of website services play an important role in customer satisfaction. By analyzing the network congestion logs, you can identify the performance bottlenecks of your websites, and take measures such as improving website caching policies or network transmission policies to optimize performance.
- Quick network fault locating
  - Network quality is the cornerstone of service stability. LTS centralizes logs from different sources, helping you detect and locate faults in a timely manner and enabling backtracking. For example, you can quickly locate an ECS that causes an error, such as an ECS with excessive bandwidth usage. In addition, you can judge whether there are ongoing attacks, leeching, and malicious requests by analyzing access logs, and locate and rectify faults as soon as possible.

## Using LTS: Basic Concepts

A log group is the basic unit in LTS for log management. You can set log retention duration for a log group.

The screenshot shows the LTS console's Log Management interface. It displays a table with one row for a log group named 'lts-group-k07x'. The table columns include 'Log Group Name/ID', 'Log Retention Duration', 'Created', 'Creation Type', and 'Operation'. The 'Log Group Name/ID' column shows 'lts-group-k07x' and '7e4fc1b9-4b16-4bf6-8f37-0e0d30ca13f2'. The 'Log Retention Duration' column shows '7'. The 'Created' column shows 'User'. The 'Operation' column has a 'Modify' button. A red box highlights the 'Log Group Name/ID' column.

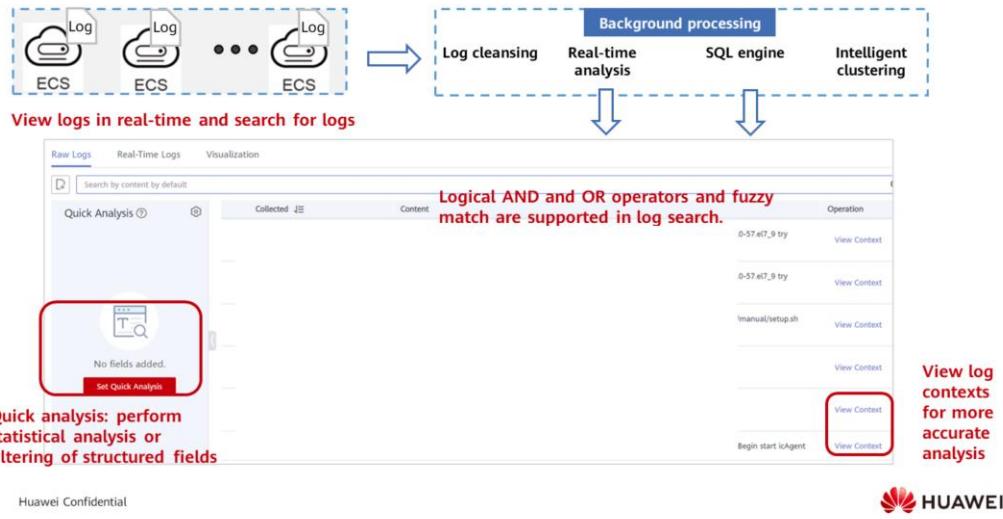
A log stream is the basic unit for log read and write. You can create log streams in a log group for finer log management.

The screenshot shows the LTS console's Log Management interface for a specific log group ('lts-group-k07x'). It displays a table with one row for a log stream named 'lts-stream-c026'. The table columns include 'Log Stream Name/ID', 'Created', 'Creation Type', 'Custom Metric Filter', and 'Operation'. The 'Log Stream Name/ID' column shows 'lts-stream-c026' and 'bf0beff9e9-4984-4387-9d41-da7a2cb1c236'. The 'Created' column shows 'User'. The 'Operation' column has a 'Search' button. A red box highlights the 'Log Stream Name/ID' column.

ICAgent is the log collection tool of LTS. If you want to collect logs from hosts, install ICAgent on the hosts.

- Log groups can be created in two ways. They are automatically created when other HUAWEI CLOUD services are connected to LTS, or you can create one manually on the LTS console.
- Data is written to and read from a log stream. You can configure logs of different types, such as operation logs and access logs, to be written into different log streams. ICAgent will package and send the collected log data to LTS on a log-stream basis. To view logs, you can go to the corresponding log stream and query them. In short, the use of log streams greatly reduces the number of log reads and writes and improves efficiency.
- ICAgent is the log collection tool of LTS. If you want to use LTS to collect logs from a host, you need to install ICAgent on the host. Batch ICAgent installation is supported if you want to collect logs from multiple hosts. After ICAgent installation, you can check the ICAgent status on the LTS console.

## Using LTS: Querying Logs



- Advantages:
  - High-performance database for log storage
  - PB-level storage capacity and high throughput
  - Log transfer to OBS
  - Log analysis in context

## Using LTS: Viewing Real-Time Logs

- You can view logs in real time on the Real-Time Logs tab, where the logs are updated every five seconds.

The screenshot shows a user interface for viewing logs. At the top, there are three tabs: 'Raw Logs', 'Real-Time Logs' (which is currently selected and underlined in blue), and 'Visualization'. To the right of the tabs is a dropdown menu set to 'Last 1 hour'. Below the tabs is a 'Log Content' section containing the following log entries:

```
Aug 2 15:52:42 ecs-linux kernel: psmouse serio1: VMMouse at isa0060/serio1/input0 lost sync at byte 1
Aug 2 15:52:42 ecs-linux kernel: psmouse serio1: VMMouse at isa0060/serio1/input0 - driver resynced.
Aug 2 15:52:43 ecs-linux kernel: psmouse serio1: VMMouse at isa0060/serio1/input0 lost sync at byte 1
Aug 2 15:52:43 ecs-linux kernel: psmouse serio1: VMMouse at isa0060/serio1/input0 - driver resynced.
```

At the bottom right of the log content area are two buttons: 'Clear' and 'Pause'.

- Logs are reported to LTS once every minute. You may wait for up to 1 minute before the logs are displayed on the **Real-Time Logs** tab. In addition, you can control log display by clicking **Clear** or **Pause** in the upper right corner.
  - If you click **Clear**, displayed logs will be cleared from the real-time view.
  - If you click **Pause**, loading of new logs to the real-time view will be paused. After you click **Pause**, the button changes to **Continue**. You can click **Continue** to resume real-time loading of logs.

## Using LTS: Structuring Logs

- After you add extraction rules, LTS uses these rules to convert raw logs to a structured format, facilitating execution of SQL queries.

Select a log event as a sample and extract fields.

2020-01-08\_17:17:36.420 [http-nio-8083-exec-9-txid = b63545cd7a69f2aa1] INFO c.h.c.api.PersistenceRestController - 0 add 3421122341 to cart.

You can extract fields from log texts in any format and easily generate regular expressions for log parsing on the UI.

Statement	Description	Example
GROUP BY	Group results of one or more columns (generally used with aggregate functions).	select * group by (year),(month)
LIKE	Search for the specified pattern in a column in the WHERE clause.	count(*)
WHERE	Specify the selection condition.	count (<column name>)

Function	Description	Example
min (<column name>)	Calculate the minimum value in a column.	select min(num)
max (<column name>)	Calculate the maximum value in a column.	select max(num)
avg (<column name>)	Calculate the average value in a column.	select avg(num)
sum (<column name>)	Calculate the sum of values in a column.	select sum(num)

Common SQL syntax and aggregate functions are supported.

- LTS supports SQL queries on structured logs. You can set rules for LTS to structure raw logs and run SQL queries one to two minutes after the setting.

## Using LTS: Visualizing Logs

- You can visualize SQL query results in tables, trend charts, bar charts, or pie charts.



# Using LTS: Collecting Statistics and Configuring Alarms

Step 1 Create a statistical rule

Rule Type: Keyword  
Rule Name: Statistics\_Rule  
Keyword: ERROR  
Description:  
Log Bucket: Log-Bucket1

Step 2 Check the curve chart of keyword statistics

Interpolation Mode: null

Step 3 Set a threshold

Threshold Preview: Average, Statistical cycle: 1 minute

Step 4 Connect to the alarm center

Rule Name	Log Bucket	Rule Type	Keywords	Metric
count-error	count-error	Keywords	error	...

Keyword alarms support HTTP, email, and SMS notifications.

- LTS does not provide alarms itself. The alarms are actually configured on Application Operations Management (AOM).

# Contents

1. O&M Basic Concepts and Principles
2. Cloud Eye
3. Log Tank Service (LTS)
- 4. Cloud Trace Service (CTS)**

## What Is Auditing?

- Auditing is the process of gathering and analyzing evidence to evaluate an enterprise's financial statements, drawing conclusions and producing reports on the compliance of the statements to generally accepted standards, and communicating the audit results to stakeholders. An audit in the information and communications technology (ICT) industry is mainly an examination of the entire lifecycle of information systems.

Audits on enterprises will usually compare the following two aspects:



## Purpose of Auditing

- Auditing is to check whether the information presented in an enterprise's financial statements is fair and accurate, helping the enterprise operate in a healthy manner. In the ICT industry, audits usually aim to examine whether information systems are running healthily.



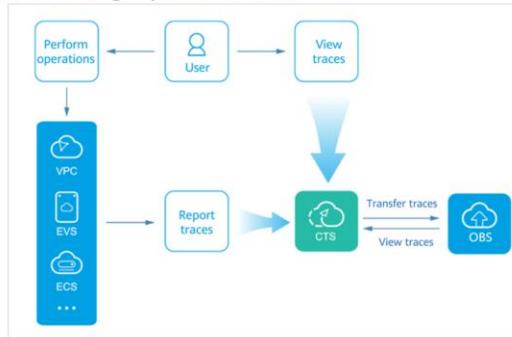
43      Huawei Confidential



- If you want to migrate your services to the cloud, you will need to ensure the compliance of your own service systems and that of the service systems and resources of the cloud vendor you choose.
- CTS plays an important role in HUAWEI CLOUD's own compliance. The service records operations of almost all services and resources in HUAWEI CLOUD, and carries out security measures such as encryption, disaster recovery, and anti-tampering to ensure the integrity of traces (operation records) during their transmission and storage. In addition, you can use CTS to design and implement solutions that help you obtain compliance certifications for your service systems.

## What Is Cloud Trace Service?

- Cloud Trace Service (CTS) records operations on cloud resources in your account. You can use the operation records (called traces in CTS) to perform security analysis, track resource changes, conduct compliance audits, and locate faults. You can also transfer the traces to Object Storage Service (OBS) in real time to store them for a longer period of time.



## CTS Advantages

### Traditional auditing

- Standardized audit processes cannot be carried out in traditional IT environments. Unauthorized API calls and console operations on servers, databases, OSs, and other resources cannot be systematically recorded in real time. System configuration changes are manually documented by IT staff.
- Audit records are manually documented. They do not have copies and cannot be stored for a long time.

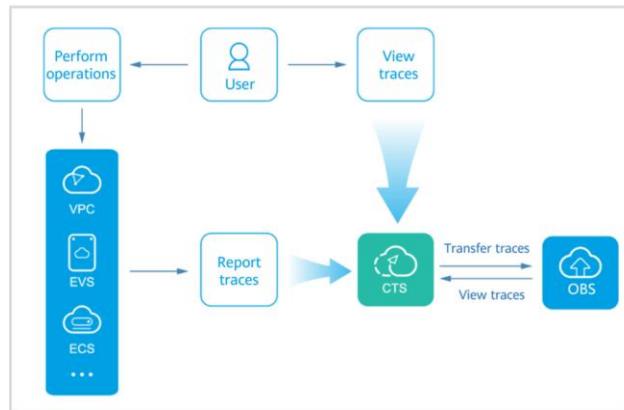
VS.

### CTS

- All operations on cloud resources can be systematically recorded in real time. Manual efforts are not required.
- CTS can regularly merge traces into time-stamped files and transfer the files to OBS buckets for highly available and cost-effective long-term storage.

## CTS Architecture

- **Major functions:**
  - Trace collection
  - Trace query
  - Trace transfer
  - Trace file encryption



- Trace collection: CTS records operations performed on the console, API calls, and system-triggered actions.
- Trace query: You can query traces of the last seven days on the CTS console by multiple filters, such as trace type, trace source, resource type, user, and trace status.
- Trace transfer: Traces can be periodically transferred to OBS buckets for long-term storage. Traces are merged into trace files corresponding to specific services during transfer.
- Trace file encryption: Trace files can be encrypted using keys provided by Data Encryption Workshop (DEW) during transfer.

## CTS Basic Concepts

### Tracker

You need to enable CTS before using it. A tracker is automatically created when CTS is enabled. The tracker automatically identifies all cloud services you are using and records all operations performed on the services.

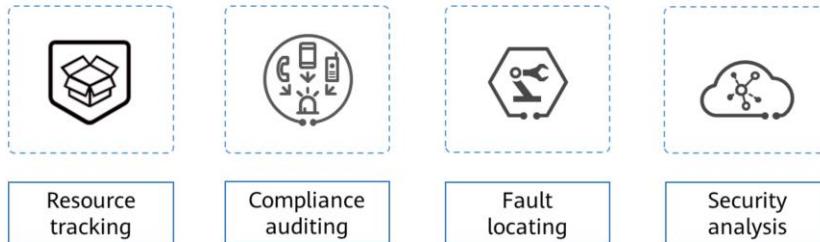
### Trace

Traces are operation records captured and stored by CTS. Traces help you identify when a specific operation was performed by a specific user on a specific resource.

- Management traces
  - Traces reported by cloud services
- Data traces
  - Traces of read and write operations reported by OBS

- You can create a data tracker on the CTS console to record operations on data.

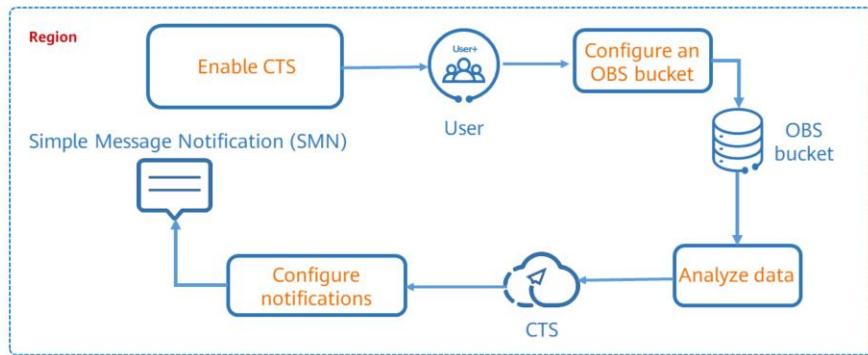
## CTS Scenarios



- Resource tracking: You can get a comprehensive history of changes made on cloud resources throughout their lifecycle.
- Compliance auditing: CTS provides a history of operation records and security capabilities, making it easy to comply with internal policies and regulatory standards.
- Fault locating: If a fault occurs, you can use filters to quickly search for unusual operations. This accelerates troubleshooting and reduces manpower requirements.
- Security analysis: Each trace records details of an operation. You can identify when an operation was performed by a specific user and the IP address from which the operation was performed. This helps you detect unauthorized operations and analyze resource changes. You can also configure email or SMS notifications for key operations.

## Using CTS: Security Analysis

- Each trace records details of an operation. You can identify when an operation was performed by a specific user and the IP address from which the operation was performed. You can perform security analysis and user behavior pattern analysis based on traces and configure notifications for key operations.



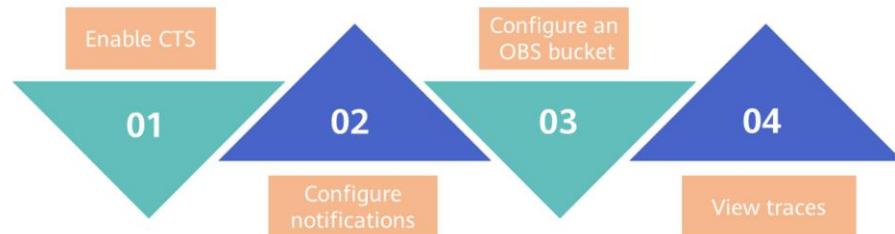
49 Huawei Confidential



- The process of security analysis is as follows:
  - CTS records all operations under your account after you enable CTS.
  - The traces are stored in an OBS bucket.
  - The data analytics component can download traces from buckets for analysis.
  - Analysis results can be set as triggers for sending notifications using SMN.

## Using CTS: Resource Change Tracking

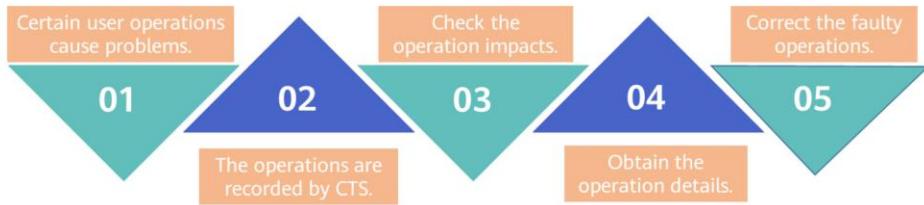
- CTS records resource changes and the change results, allowing you to track and analyze resource usage statistics.



- The process of resource change tracking is as follows:
  - All changes on cloud services are recorded by CTS.
  - You can configure notifications for key operations.
  - Change records are permanently saved.
  - You can query traces for details about resource changes.

## Using CTS: Fault Locating

- If a fault occurs, you can view CTS traces to figure out the cause and rectify the fault quickly. For example, you can quickly determine that the deletion of a system volume during configuration led to a failure in ECS capacity expansion.



- The process of fault locating is as follows:
  - A user performs operations that cause problems.
  - All operations are recorded by CTS.
  - You can search related traces by resource name and check the operation impacts.
  - You can obtain the operation details, including the time and the user who performed the operations.
  - You can correct the faulty operations based on the obtained information.

## Using CTS: Compliance Auditing

- CTS records operations and allows you to query operation records, making it easy to comply with internal policies and regulatory standards. This helps you meet the requirements of IT compliance certifications (for example, certifications for financial cloud and trusted cloud).

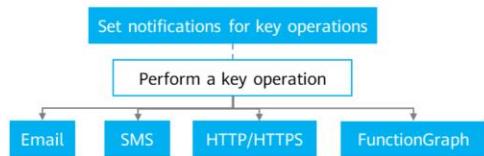


## Using CTS: Key Event Notifications

- Traces can be set as triggers for notifications sent to emails, mobile phones, and system interfaces or as triggers to invoke FunctionGraph functions.

Major functions:

- You can be alerted of changes to core system components, networking, and security configurations so that risks can be detected and mitigated as soon as possible.
- Traces collected by CTS can be synchronized to your own audit systems through HTTP/HTTPS notifications for independent auditing.
- FunctionGraph can be triggered by traces to execute specific functions.



## Quiz

1. Cloud Eye is a free cloud service.  
True  
False
2. Which of the following are scenarios of CTS?
  - A. Resource tracking
  - B. Compliance auditing
  - C. Fault locating
  - D. Security analysis

- True. Cloud Eye is free. You can use Cloud Eye to monitor and manage your purchased cloud services.
- ABCD

## Summary

- O&M services play an important role in ensuring that platforms are secure and operate normally. We can use CTS to better manage platforms, and use Cloud Eye to monitor platforms in real time. With LTS, we can obtain logs in real time and evaluate and eliminate potential risks.

# Recommendations

- Huawei Learning
  - <https://e.huawei.com/en/talent/#/>
- HUAWEI CLOUD Help Center
  - <https://support.huaweicloud.com/intl/en-us/help-novice.html>

## Acronyms and Abbreviations

CTS: Cloud Trace Service

ECS: Elastic Cloud Server

IT: Internet technology

ICT: Information and communications technology

IAM: Identity and Access Management

LTS: Log Tank Service

OBS: Object Storage Service

SOA: Service-oriented architecture

## Acronyms and Abbreviations

SQL: Structured query language

VPC: Virtual Private Cloud

# Thank you.

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

Copyright©2021 Huawei Technologies Co., Ltd.  
All Rights Reserved.

The information in this document may contain predictive  
statements including, without limitation, statements regarding  
the future financial and operating results, future product  
portfolio, new technology, etc. There are a number of factors that  
could cause actual results and developments to differ materially  
from those expressed or implied in the predictive statements.  
Therefore, such information is provided for reference purpose  
only and constitutes neither an offer nor an acceptance. Huawei  
may change the information at any time without notice.

