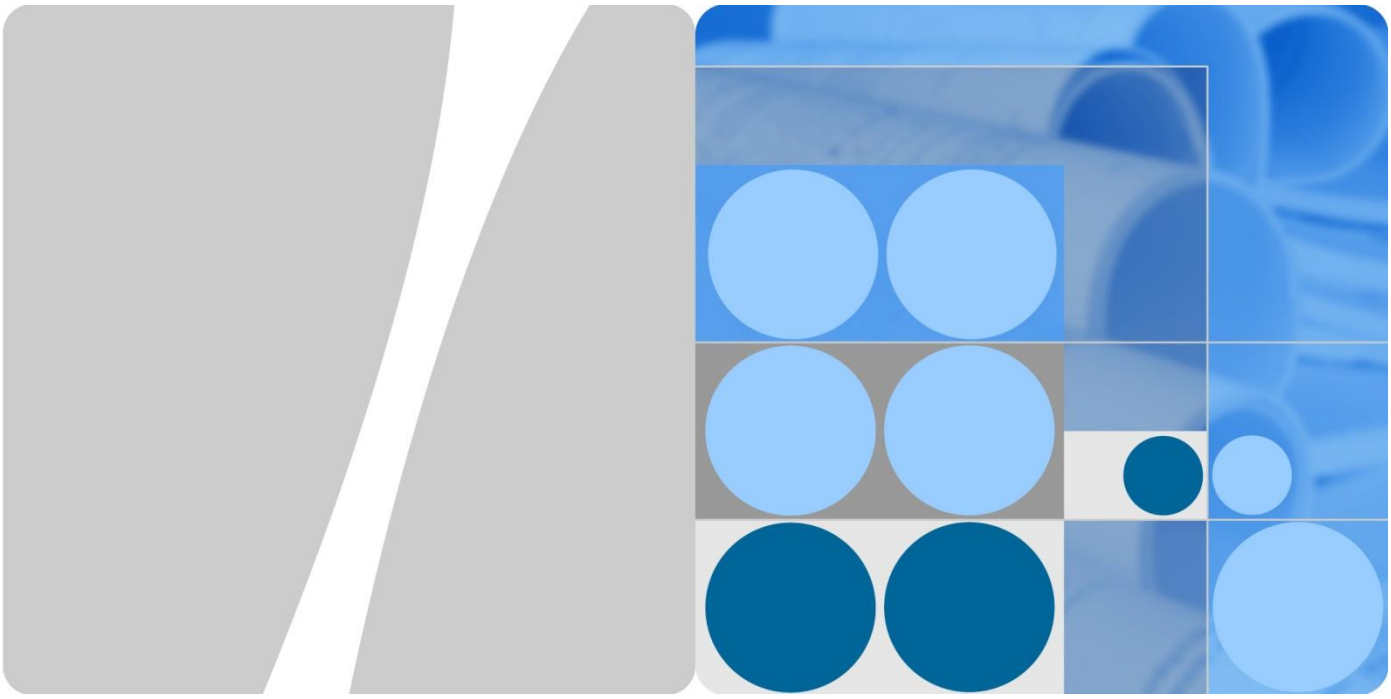


Huawei SD-DC² 2.0

Technical White Paper

Issue 02

Date 2016-02-29

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.

All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: http://e.huawei.com

Email: support@huawei.com

Change History

Issue	Description	Release Date	Prepared By	Approved By
01	This issue is the first official release.	2016-02-18		

Contents

1 Overview

The following table lists the main contents of this chapter.

Section	Content
1.1 Solution Positioning	Describes the positioning of the SD-DC^2 solution.
1.2 Solution Architecture	Describes the logical architecture of the SD-DC^2 solution.
1.3 Solution Deployment	Describes the logical deployment of the SD-DC^2 solution.

1.1 Solution Positioning

To address challenges for data centers and conform to technology development trends, Huawei launches the Service Driven Distributed Cloud Data Center (SD-DC^2) Solution. The SD-DC^2 solution supports physically discrete but logically unified resources, collaboration between the cloud and pipes, and service perception.

The SD-DC^2 solution provides an automatic management and virtualization platform to support fine-grained IT operation by using a converged architecture (with computing, storage, and network resources) as the fundamental unit of resource pools and constructing a software-defined networking (SDN) service perception network. The main idea of the SD-DC^2 solution is physical distribution and logical unification. The SD-DC^2 solution integrates data centers all over the world into a large data center which provides unified services. Such convergence improves IT efficiency for enterprises. Delocalization, software-defined data center construction, and automation are characteristics of the SD-DC^2 Solution. Logical unification has two meanings:

- All data centers and their resources are centrally managed, scheduled, and maintained on a rights- and domain-specific method. To support these capabilities, the SD-DC^2 solution must provide a unified O&M management platform.
- If distributed cloud data centers need to provide external services, a unified portal is required to display services, and a unified process is required to support service provisioning. In this case, the SD-DC^2 solution must provide a unified service platform.

The SD-DC^2 solution does not aim at improving efficiency and user experience of a single data center. Instead, it regards multiple data centers as an integrated one. Based on cross-data center management, resource scheduling, and disaster recovery (DR) design, the SD-DC^2 solution provides a cloud platform that migrates cloud resources across data centers, an O&M management system that centrally manages and schedules resources of multiple data centers, a large Layer 2 ultra-high bandwidth network, and software-defined data centers. The SD-DC^2 has an epoch-making, revolutionary data center architecture, which brings unprecedented benefits and user experience to customers. It provides the following benefits:

- Reduced total cost of ownership (TCO) and increased return on investment (ROI)

  The SD-DC^2 adopts virtualization technologies to ensure that software is independent of hardware and enable the infrastructure with low usage to provide elastic, automated, and secure computing resource pools. Resources can be allocated to applications on demand. The SD-DC^2 solution helps enterprises reduce operating expenditure (OPEX) costs by resource consolidation and automation. The adopted distribution technology logically unifies resources of multiple data centers, improving resource usage and reducing infrastructure investments. The SD-DC^2 solution provides disaster backup services and the cross-data center application migration capability with load balance to improve application availability and resource utilization. Improved availability and shortened downtime help enterprises save intangible costs. In addition, the SD-DC^2 solution enhances the availability of virtual machines (VMs) using the VM migration service. Encapsulation attributes of VMs and virtual disks and capabilities of obtaining VM status accelerate VM backup and restoration.

- Enhanced service agility, quick service rollout, and improved user satisfaction

  The SD-DC^2 solution allocates resources on demand using virtualization technology and supports all-round management and service automation. Self-service capabilities allow users to apply for computing, storage, and network resources on a per use basis. The SD-DC^2 solution quickly provisions and deploys services, and supports dynamic load balancing. The SD-DC^2 solution provides different service level agreements (SLAs) for different applications. It is capable of scaling out and scaling up based on user-defined scheduling policies. Such a flexible scalability ensures that the IT systems can quickly respond to service changes, making data centers transit from a costly center to a value-creating center.

- Fewer requirements for IT system management and maintenance resources

  The SD-DC^2 solution supports self-service capabilities. Users can apply for services by themselves, which minimizes dependency on the IT department. Automated workflows can be created based on standard processes, such as event management, problem management, change management, and release management. Centralized O&M, proactive management, and correlation of service requirements and IT processes eliminate failures and reduce manual operations, so that the O&M efficiency of multiple data centers is improved.

The SD-DC^2 provides the following key capabilities:

- Provides data center as a service (DCaaS) for tenants in the form of virtual data centers (VDCs).

  VDCs are an implementation of Software-Defined Data Center (SDDC). Resources of VDCs come from different resource pools of multiple physical data centers. VDC resources are classified into computing, storage, and network resources. VDC resource capacity is applied for by the VDC service manager or specified by the system manager. Resources are provided for users after applications are approved.

  Users can use VDC resources after they submit an application and the application is approved by the VDC service manager. The VDC service manager is responsible for service approval, service template management, service management, resource

configuration, resource provisioning, and self-service O&M. The VDC service manager performs life cycle management for services provided by a VDC. The VDC service manager can define services and publish the services to service catalogs to be applied for by users, review users' applications, and cancel published services. Access rights to VDC resources can be controlled. VDC networks can be defined by the VDC service manager. A VDC can be divided into multiple virtual private clouds (VPCs), each containing multiple subnets. VDCs provide multiple types of computing, storage, network, and application services at the infrastructure as a service (IaaS) layer.

The VDC service provides some self-service O&M capabilities, including viewing information about VDC alarms, performance, capacity, and topologies. The VDC service provides VDC resource metering information, enabling tenants to measure charging information.

- Provides cloud infrastructure that is optimized for various application scenarios.

  In different application scenarios, cloud infrastructure requirements vary. The SD-DC2 solution provides different types of infrastructure for different application scenarios to meet differentiated requirements of upper-layer applications and improve infrastructure efficiency and quick delivery capabilities. Cloud infrastructures are provided for four scenarios:

  - Standard virtualization scenario

    The common application virtualization infrastructure and desktop cloud infrastructure are provided.

  - High-throughput scenario

    The infrastructure supporting online analytical processing (OLAP) applications is provided. The infrastructure is optimized in aspects of storage and networks and supports high-performance network connections such as InfiniBand.

  - High scalability scenario

    The computing and storage convergence solution is adopted to provide rapid scaling capabilities for applications.

  - High-performance scenario

    To meet requirements of online transaction processing (OLTP) applications and replace midrange computers with x86 servers, servers adopt multiple Reliability, Availability, and Serviceability (RAS) technologies to improve reliability. Storage devices support input/output operations per second (IOPS) in millions, and servers supports response in microseconds.

- Supports unified and flexible management of cloud data centers.

  Resources of the SD-DC^2 solution come from multiple physical data centers. Diversified resources make management complex. To simplify management, the SD-DC^2 solution provides the following unified management functions:

  - Unified management of multiple data centers

    Resources from multiple data centers can be centrally accessed and managed.

  - Unified management of physical and virtual resources

    Physical servers, storage devices, and network devices as well as virtual resources are managed in a unified manner, and topologies between resources are displayed in a unified view.

  - Unified management of multiple virtualization platforms

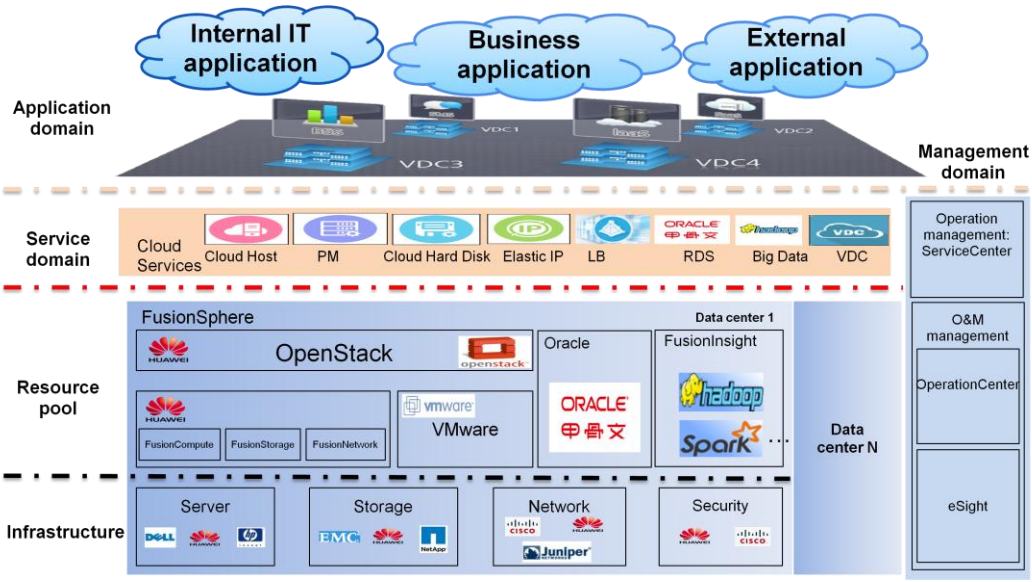
    Various virtualization platforms and technologies are managed in a unified manner.

  - Different SLA-based DR service capabilities

OpenStack architecture–based SD-DC^2 currently enables cloud hard disk backup services and active-passive services provisioning. It uses tenant SLAs to allocate DR service resources, ensuring multi-tenant self-service VM data security protection and business continuity DR capability.

1.2 Solution Architecture

The following figure illustrates the overall architecture of the SD-DC^2 solution.



The overall architecture consists of the infrastructure layer, resource layer, service layer, application layer, and management domain.

The following table describes each component of the SD-DC^2 solution architecture.

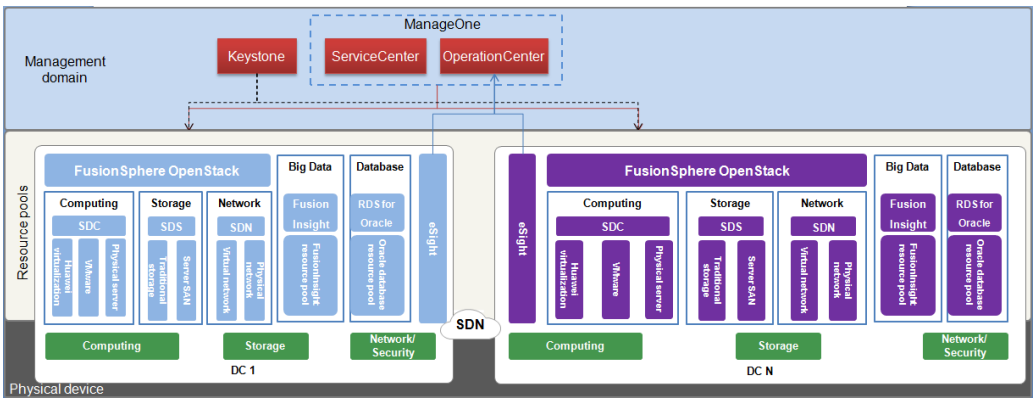
Functional Layer	Description
Infrastructure layer	The infrastructure layer consists of servers, storage devices, network devices, and security devices. It provides capabilities of constructing physical and virtual computing, storage, and network resource pools. The SD-DC^2 solution offers multiple types of infrastructure.
Resource pool	The resource pool layer manages virtual computing, storage, and network resources. The SD-DC^2 solution provides the capability of managing converged resource pools, heterogeneous virtualization platforms such as VMware and FusionSphere, and physical resource pools.

Functional Layer	Description
Service layer	Based on the O&M capabilities of the management layer, the service layer provides service catalogs to implement secondary resource operation for each service scenario. • Resources are flexibly allocated in the form of VDC to implement VDCaaS. • The VDC provides the cloud host service and the Elastic Block Storage (EBS) service to implement IaaS.
Management layer	The management layer provides unified management and resource scheduling for multiple cloud data centers and provides DCaaS based on VDCs. A VDC provides multiple types of cloud services. This layer also supports unified O&M of virtual and physical resources.
Application layer	At this layer, enterprises construct their application systems based on the service provided in the data center to meet customers' requirements.

1.3 Solution Deployment

The SD-DC^2 solution adopts OpenStack as the basic cloud management platform. With the support capabilities of OpenStack for heterogeneous virtual resources, the SD-DC^2 solution provides unified management and scheduling of multiple virtualization platforms, and implements converged resource pool capabilities. Based on OpenStack, a unified O&M management platform across multiple data centers is constructed, achieving the objectives of the SD-DC^2. The following figure shows the deployment of the SD-DC^2 components.

Figure 1-1 OpenStack-based SD-DC^2 component deployment



The figure shows the component connections in the OpenStack architecture. Keystone is deployed in the management domain to implement unified authentication for OpenStack instances. The OpenStack platform provides native capabilities to adapt to heterogeneous virtualization platforms, and supports multiple virtualization platforms such as VMware and FusionSphere.

Table 1-1 Components in the SD-DC^2 solution

Component	Description
ManageOne	ManageOne provides ServiceCenter and OperationCenter. ServiceCenter: implements unified service orchestration and automatic management based on cloud and non-cloud resources provided by resource pools. Including customizable heterogeneous and multi-resource-pool policies and orchestration, customizable enterprise service integration, resource pool management capabilities supplemented by third-party components. Especially automatic provisioning capabilities for heterogeneous traditional resources. OperationCenter: implements maintenance based on scenarios and visualized status, risk, and efficiency analysis for data center services. Proactively analyzes problems and works with the Service Center to implement data center self-optimization and self-healing based on analysis results.
eSight	Provides region-level O&M capabilities, including alarming, performance, monitoring, and topologies.
AgileController	Provides the network virtualization capability.
FusionCompute	FusionCompute virtualizes and pools computing, storage, and network resources.
FusionSphere OpenStack	FusionSphere OpenStack is an open-source cloud management system. It consists of multiple components, which are decoupled using Representational State Transfer (REST) interfaces and message queues. FusionSphere OpenStack can manage heterogeneous virtualization platforms, such as VMware and UVP. OpenStack consists of the following components: • Nova: virtual computing • Glance: image • Cinder: virtual disk • Neutron: virtual network • Swift: object storage • Keystone: authentication • Ceilometer: monitoring
FusionInsight	Functions as a big data resource pool to provide big data resources.
RDS for Oracle	Provides Oracle database services.

2 Key Features

About This Chapter

The following table lists the main contents of this chapter.

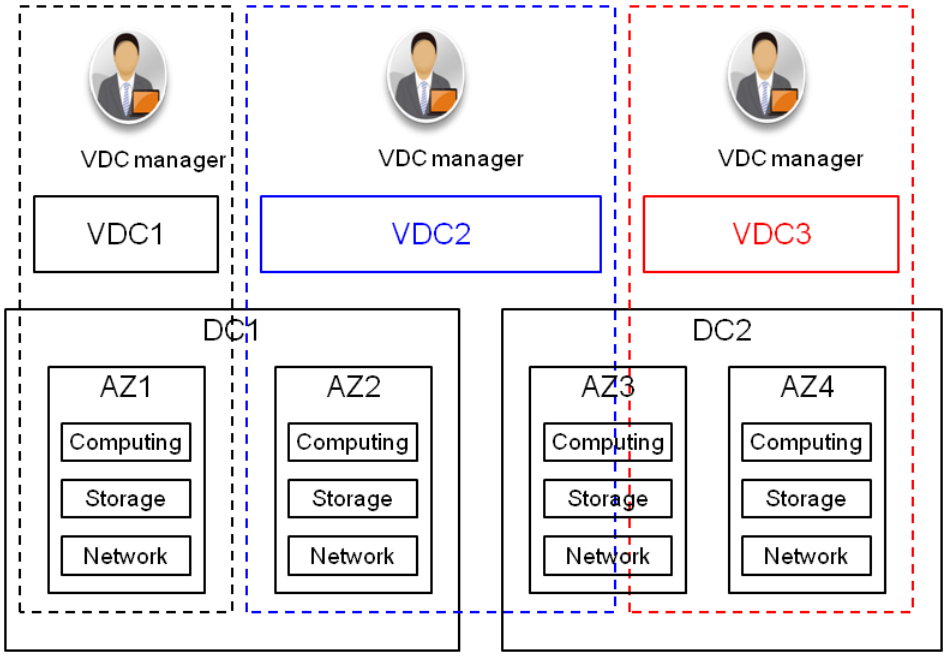
Section	Content
2.1 VDC	Provides an overview of VDC as well as describes its application scenarios and key features.
2.2 Unified Management	Describes the centralized management method applied in the SD-DC^2 solution.
2.3 OpenStack-based Architecture	Describes the Huawei OpenStack open cloud management platform.
2.4 Distributed Storage FusionStorage	Describes FusionStorage, the Huawei distributed storage system.
2.5 Big Data Service	Describes the Huawei FusionInsight solution.
2.6 SDN	Describes the Huawei SDN solution.
2.7 RDS	Describes the Huawei RDS solution.
2.8 Security Management	Describes the security planning in the SD-DC^2 solution.

2.1 VDC

2.1.1 Feature Overview

Virtual Data Center (VDC) is a technology used for logical isolation. It logically isolates physical resources into VDCs. A department and an organization can obtain the batch computing, storage, and network resource quota at a time by applying for a VDC. Within the resource quota, the VDC service manager can control the computing, storage, and network resources as needed.

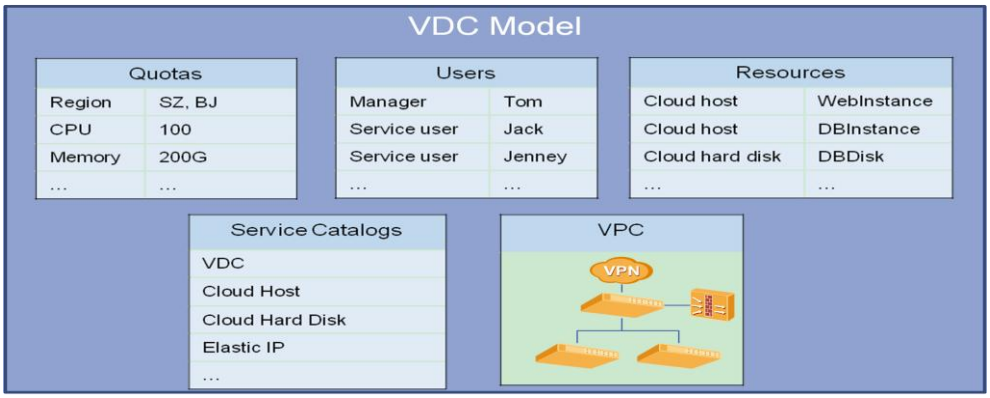
The following figure illustrates the structure of a VDC.



- VDC 1 corresponds to availability zone (AZ) 1 in DC 1. Cloud host and EBS resources are provisioned in AZ1.
- VDC 2 corresponds to AZ 2 in DC 1 and AZ 3 in DC 2. Cloud host and EBS resources are provisioned in AZ 2 and AZ 3.
- VDC 3 corresponds to AZ 3 and AZ 4 in DC 2. Cloud host and EBS resources are provisioned in AZ 3 and AZ 4.

  NOTE
  - VDC 1 is exclusive to AZ 1, and this must be planned by the system manager. ServiceCenter does not support exclusive AZs.
  - The quotas of VDC 2 and VDC 3 determine how many resources in AZ 3 are used by VDC 2 and VDC 3.

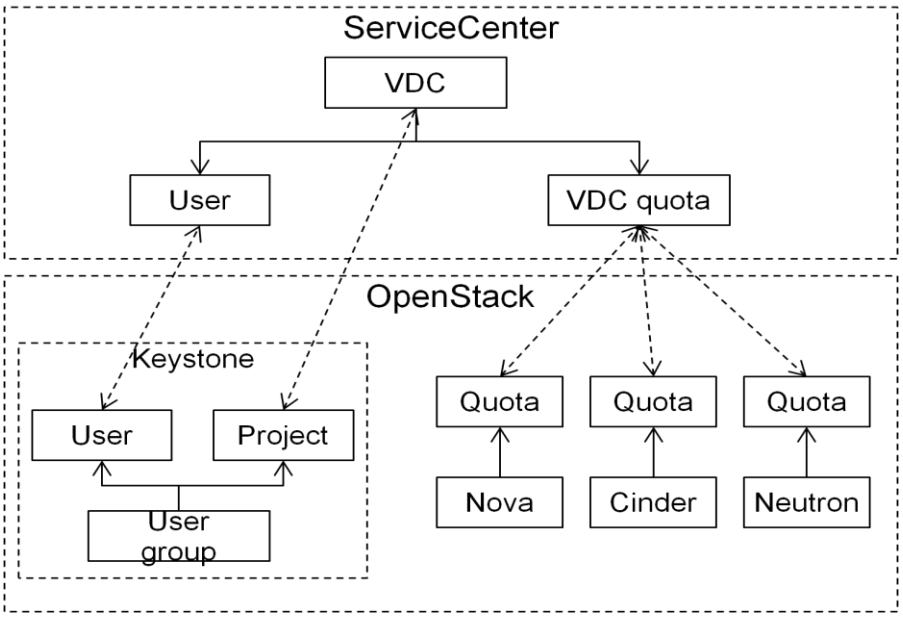
  A complete VDC involves quotas, users, resources, service catalogs, networks, and templates, as shown in the following figure.



- Quotas: A quota controls the upper limit of resources used by a VDC. It is defined by the VDC service manager when applying for a VDC, or by the organization manager when creating a VDC.

- Users: A VDC involves two user roles: VDC service manager and service user. The VDC service manager manages the users in the VDC and reviews service applications. Service users are the final users of resources.
- VPCs: A VPC is a logically isolated network environment, including sub-functions such as virtual routers (vRouters), networks, access control lists (ACLs), and VPNs.
- Service catalogs: A service catalog lists the services for which VDC users can apply. The system provides out-of-box services, including cloud host, EBS, elastic IP address, physical machine (PM), virtual load balancer (vLB), backup, disaster recovery, and FusionInsight. Managers can define VDC service catalogs for each individual department.
- Resources: Service resources that can be applied for by service users, including cloud host, EBS, and elastic IP address. Users can maintain and monitor the resources.

A VDC is a project in OpenStack, as illustrated in the following figure:



When the organization manager is creating a VDC on ServiceCenter, ServiceCenter creates a project in OpenStack and users in the project, and also configures the project quota in Nova, Cinder, and Neutron.

Two-level VDC management:

ServiceCenter provides VDC management in two levels: organization and VDC. The two-level management applies to government cloud and reselling by contractors.

In a government cloud scenario, an information center constructs a cloud resource pool for a government. This cloud resource pool will be used by each department of the government. Each department has its own service units to manage. In this scenario, the information center functions as the system manager, each department as an organization manager, and each service unit as a VDC service manager.

In a scenario of reselling by contractors, a carrier constructs cloud resource pools and leases them to enterprises. Some enterprises do not rent resources directly from the carrier but from a contractor. In this scenario, the carrier functions as the system manager, the contractor as an organization manager, and the enterprise as a VDC service manager.

The system manager creates organizations and organization managers, and specifies the quotas of organizations. The organization manager creates VDC and VDC service managers, and specifies the quotas of VDCs. The quota of a VDC cannot be greater than that of the owning organization.

Organizations and VDCs support the metering of resource usage.

2.1.2 Application Scenarios

In the private cloud of enterprises, leased resources must be independently managed, and networks must be isolated. Each VDC is an independent management entity that enjoys self-operation and self-maintenance capabilities. VDCs can be flexibly divided based on application scenarios of an enterprise.

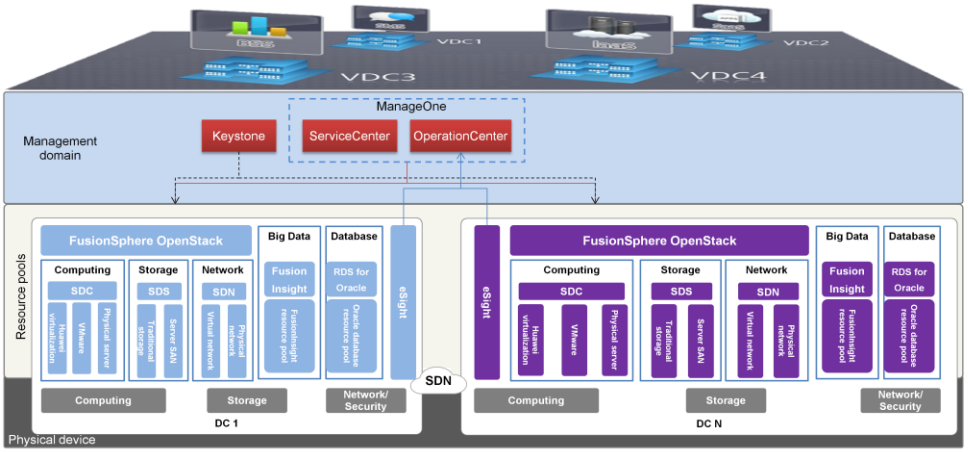
- VDCs are divided based on departments. Each department can independently manage its resources.
- VDCs are divided based on fields, such as the development VDC and testing VDC.

NOTE

Resources of a VDC come from one or multiple physical data centers. Customers can benefit from capabilities of VDCs, including self-service operation and maintenance and resource isolation management.

2.1.3 Deployment Architecture

Figure 2-1 VDC deployment diagram

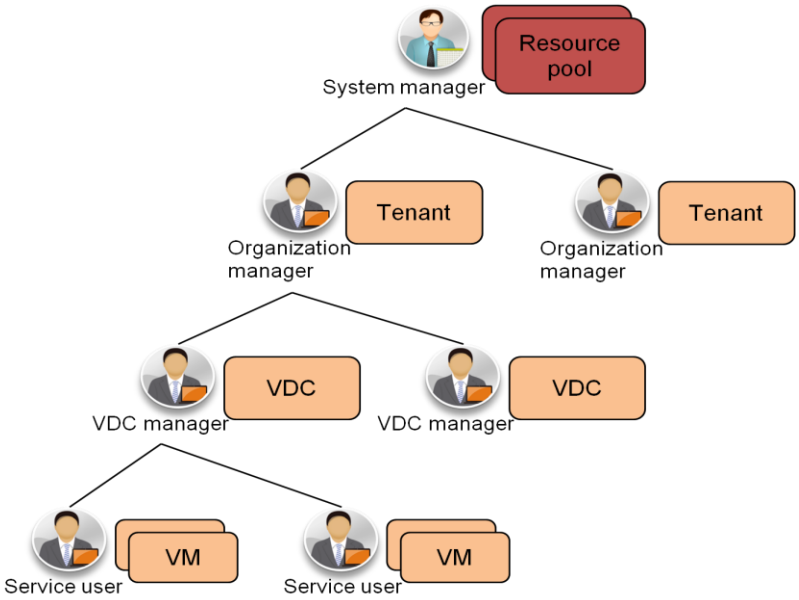


In the ServiceCenter+OpenStack deployment mode, OpenStack serves as the cloud management platform, and ServiceCenter provides VDC services. ServiceCenter connects to OpenStack through REST application programming interfaces (APIs) provided by OpenStack. VDC resources come from multiple OpenStack resource pools of multiple physical data centers. ServiceCenter provides the self-service maintenance capabilities of VDCs and OperationCenter provides the maintenance capabilities at the data center layer.

2.1.4 VDC Roles

The cloud service operation module contains users on the management side and users on the user side. The users on the management side are system managers. The users on the user side are organization managers, VDC service managers, and service users.

Figure 2-2 Relationship between each role

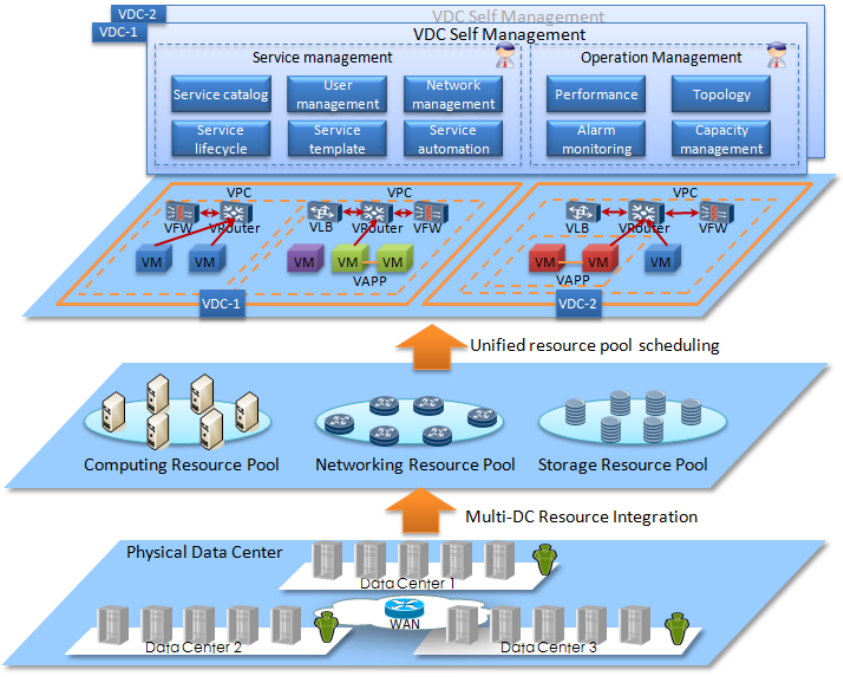


The following table describes VDC roles.

Role	Responsibility	Layer	Component
System manager	Responsible for operating and managing overall services of an SD-DC^2. • Manages organizations, including creating organizations and organization managers. • Maintains domain service catalogs, including defining services and managing service catalog permissions. • Managing resource pools, including accessing resource pools and managing AZs. • Managing system users. • Configuring system parameters. • Viewing operation logs.	Domain	ServiceCenter
Organization manager	• Managing VDCs, including reviewing VDC services and creating VDCs. • Maintains organizations' service catalogs, including defining services and managing service catalog permissions. • Manages organization users.	Organization	ServiceCenter

Role	Responsibility	Layer	Component
	• Views operation logs.		
VDC service manager	• Applies for VDCs. • Manages existing VDCs, including monitoring VDCs' capacity, applying for VDC expansion, and releasing VDCs. • Manages VDC users. • Manages VDC networks. • Maintains VDC service catalog. • Applies for, reviews, extends, changes, and releases services. • Uses and maintains VDC resources. • Views VDC operation logs.	VDC	ServiceCenter
Service users	• Applies for, extends, changes, and releases services. • Management application orders. • Uses and maintains existing resources.	VDC	ServiceCenter

2.1.5 Key Features



1. Unified Resource Management for Multiple Data Centers

A VDC obtains resources from resource pools in multiple physical data centers. Resource pools are provided by AZs. When an AZ is selected, a specific resource pool is selected. When creating a VDC, the manager selects AZs from a list based on requirements. When users apply for resources, resources are provisioned from the AZs.

Each AZ is divided into different Host Aggregates. Each Aggregate provides different SLAs. For example, the Aggregate that provides solid state drives (SSDs) is a high-performance Aggregate. The manager divides Aggregates based on SLA requirements. Aggregates are invisible to users. When applying for resources, users can specify SLA requirements, and the Scheduler selects resources from the Aggregate that meets the SLA requirements.

2. VDC Isolation

   VDCs provide management, network, and resource isolation capabilities.

   - Management isolation:

     After logging in to a VDC, users can apply for services provided by the VDC and use the services after their applications are approved by the VDC service manager. Each VDC provides independent user management, service management, template management, service catalog management, capacity management, O&M management, and approval process management capabilities. The VDC service manager manages the local VDC only without affecting the management of other VDCs. This ensures isolated VDC management.

   - Resource isolation:

     Resources include virtual resources, such as VMs and virtual disks. Each VDC independently manages its sources, and not resources can be shared between VDCs. Only the user who owns a certain resource can view information about the resource and operate the resource in a VDC, such as starting and stopping VMs and expanding capacity. Users of other VDCs cannot view information about the resource and operate the resource.

3. Quota Management

   VDC resources can be controlled using quotas. A quota is specified by an applicant when creating a VDC and approved by the system manager. A quota can also be specified by the system manager when creating a VDC. Quotas include the numbers of vCPUs, virtual local area networks (VLANs), VPCs, subnets, and VMs, memory size, and network bandwidth. If users apply for resources that exceed the quota, the VDC automatically rejects the application. The quota usage is displayed, facilitating capacity control for the VDC service manager.

4. User Management

   Each VDC allows for independent user management. The VDC service manager can assign VDC access rights to users. After being granted rights, users can log in to the VDC and apply for services of the VDC. A user can be granted permissions to access multiple VDCs.

5. Service Management

   The VDC service manager manages service catalogs and service life cycle. The VDC service manager can define service catalogs which include services that have been published and can be subscribed to by users. The VDC Service Manager defines services, including the service name, description, specifications, and properties, and then publishes the services to a service catalog. VDC users can browse the service catalog and apply for services from the catalog. The VDC service manager can cancel services that are not provided anymore. The canceled services are not displayed in the service catalog and cannot be applied for by users.

6. Template Management

A VDC provides multiple types of service templates that help the VDC service manager quickly define new services. Service configuration specifications and default values can be defined in service templates. Service templates provided by VDCs include VM templates and vApp templates. The service templates enable the VDC service manager to quickly create and deploy services.

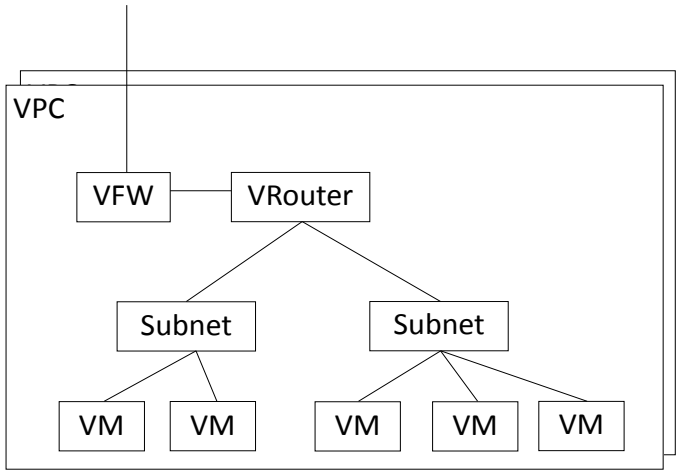
7. Service Automation

   Service automation supports automatic service provisioning and rollout. After users apply for services from service catalogs and their applications are approved, the service automation engine invokes related service implementation processes based on subscribed services. Service processes are implemented by the internal business process management (BPM) engine. The BPM engine automatically invokes REST interfaces of virtualization systems and ensures the process implementation sequence and results. The system provides default services processes covering VM, virtual disk, and virtual network device services. The manager can define new service provisioning processes to meet new service requirements.

8. Network Management

   VDC self-service network management is implemented by the basic capabilities provided by the virtualization layer. The following items support self-service management:

   A VPC provides a logically isolated virtual network environment on a physical data center for a VDC. Using the tenant portal of ServiceCenter, the VDC service manager can define a network environment for the VDC. The following figure shows a typical VPC structure:



   A VPC contains the following network components:

   - One virtual firewall (vFW)
   - One vRouter
   - Multiple network planes: intranet, routed network, and direct-connection network.

   - Network type
     - Intranet

       An intranet does not provide gateways. It has two layers and does not provide Layer 3 access. Intranets are for internal use only, and do not communicate with external network. For example, databases are deployed in intranets.

     - Routed network

A routed network provides VLAN and Layer 3 gateways. All routed networks in a VPC are interconnected to facilitate communication in different routed networks. Routed networks communicate with other networks. For example, web applications and the portal are deployed in a routed network.

A routed network is an internal network associated with routers. Two steps are required when creating a routed network: create an internal network, and then associate the internal network with the vRouters of the VPC. Therefore, after routers are deassociated from a routed network, that routed network becomes an internal network, and after an internal network is associated with routers, that internal network becomes a routed network.

- Direct-connection network

  A direct-connection network can connect VMs to external networks.

  The manager side of ServiceCenter has a type of network called external networks. A direct-connection network is on the tenant side. It is an existing public network inside an enterprise. It must be configured before cloud resource pools are deployed, and ServiceCenter automatically detects it on the manager side. External networks are classified into two types. One type of external networks is connected to the Internet, and the other type is used for VPC interconnection inside resource pools. After managers specify a type as planned, tenants can see those external networks.

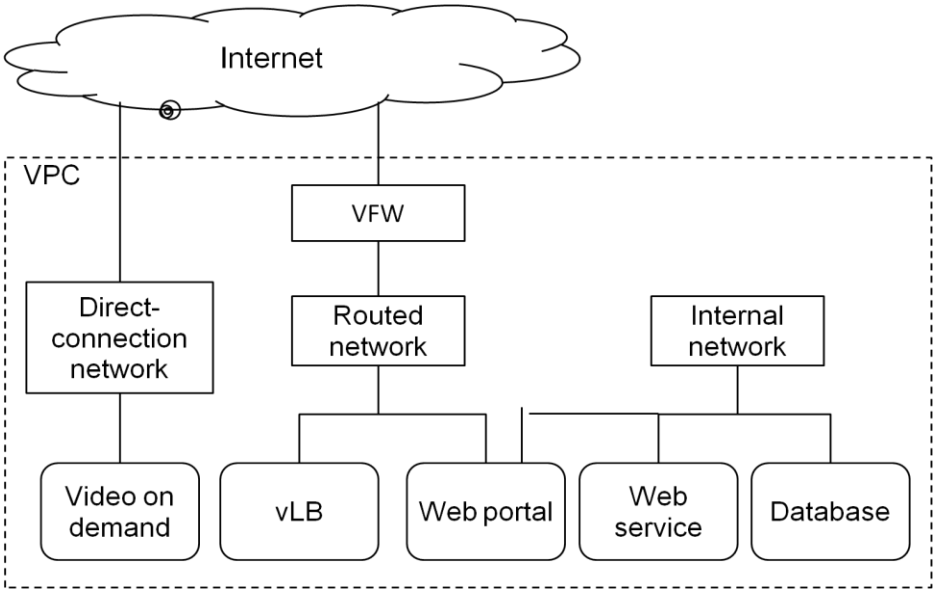
- IP address management

  Internal networks and routed networks support the following IP address assignment methods:

  - Dynamic Host Configuration Protocol (DHCP) dynamic assignment. OpenStack Neutron provides the DHCP capability. Each subnet has one software DHCP service to assign IP addresses to VMs.
  - Static ingestion. Using the capability of the virtualization layer, the virtual data center solution assigns an IP address to a VM when the VM is being created.

  Different VPCs have their own DHCP services and they can have overlapping address spaces. End users can then plan their internal addresses as desired regardless of how other departments select addresses. However, two networks with the same IP address cannot interwork, and duplicate IP addresses may complicate management. Therefore, do not plan duplicate IP address in a private cloud (allowed in a public cloud).

  The following figure illustrates how to use those types of networks.



For those services that communicate with external networks only, deploy them on a direct-connection network, for example, video on demand on the previous figure.

For those services that communicate with internal and external networks, deploy them on a routed network, for example, the Portal of a Web system.

For those services that provide internal access only, deploy them on an internal network, for example, the Web Service component and database component of a Web system. In addition, the Web Portal needs to interwork with the Web Service. Therefore, configure two network planes for the Web Portal: one routed network and one internal network.

Projects without AC SDN controllers do not support network automation and ServiceCenter does not provide a routed network. Therefore, as shown in the previous figure, deploy the Web Portal and vLB on a routed network.

- VPC service

  After creating a VPC, a VDC service manager can apply in the VPC for services of vRouter, network, elastic IP address, SNAT, ACL, and VPN.

  - vRouters: provide the routing function for VPC networks.
  - Elastic IP address: Elastic IP addresses are static, public IP addresses. Elastic IP addresses can be associated with cloud hosts. Tenants can use elastic IP addresses to access cloud hosts over the Internet. When the cloud host associated with an elastic IP address is faulty or needs to be upgraded, the elastic IP address can be mapped to another standby cloud host that is running properly. Services can be obtained from the standby cloud host without any modification of the configuration on the cloud host client, which reduces the impact on ongoing services.
  - SNAT: Source Network Address Translation (SNAT) is used to translate the source IP address contained in an IP packet into another public IP address, allowing internal users to access external networks using a shared public IP address. Normally, the internal network of an enterprise uses private IP addresses for internal communication, and it only uses a public IP address as its network egress. Therefore, when an internal network user sends a request to access an external network server, the access request packet can be sent to the server through the internal network egress, but the response packet from the

server cannot be sent to the internal user, because no route to the private IP address of the internal is available on the public network.

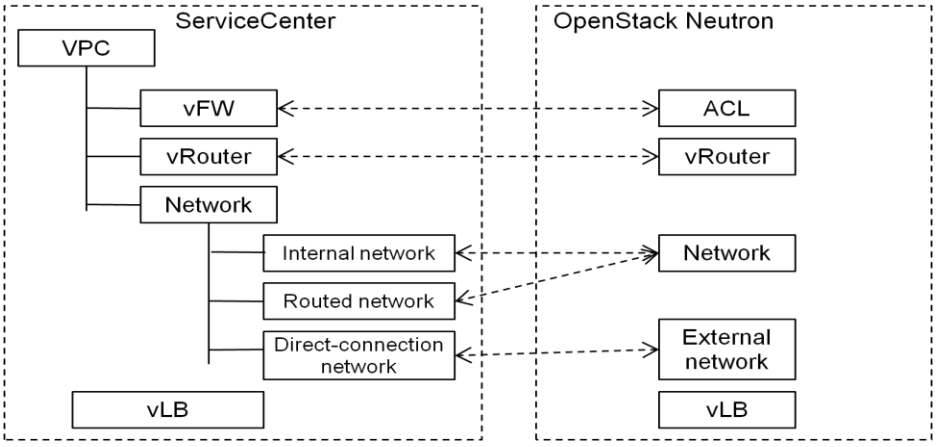
- Security group: A security group functions as a vFW of a cloud host to control the incoming and outgoing of network messages. Only authorized messages are allowed to pass.
- ACL: An ACL is a general tool for traffic matching. It can filter and match traffic in terms of IP addresses, ports, and protocols. It filters the traffic between networks in a VPC and between the Internet and the VPC.
- VPN: A remote network can be connected to a routed network in the VPC over the Internet through an Internet Protocol Security (IPsec) VPN tunnel to access VMs in the VPC remotely.

- Model mapping

  ServiceCenter provides the VPC function while OpenStack does not have VPC objects. For easy management, ServiceCenter creates a vRouter in OpenStack Neutron for each newly created VPC. However, ServiceCenter does not activate those vRouters at once until the VDC service manager creates vRouters in the VPC.

  A VPC network corresponds to a NetWork in OpenStack Neutron, and a cloud host network interface card (NIC) corresponds to a Neutron port.

  The following figure illustrates the VPC model mapping:



- Customer benefits of VPC
  - Network isolation
  - VPCs enable network isolation for VDCs. With VPCs, a department can have multiple isolated network domains. For example, a department has a production environment and a test environment. One VPC can be created for both of the two environments to isolate the two environments.
  - Efficient management of VLANs and IP addresses
  - In general, VLAN information and IP addresses are recorded in soft or paper copy. The recorded information may be inconsistent with the actual information. If errors occur, the errors are difficult to isolate.

    VLANs and IP addresses are allocated to VPCs as resources, which enables clear recording of the time when VLANs and IP addresses are applied, their purposes, resource usage, and remaining amount, improving management efficiency and reducing errors.
  - Secure connections

Organizations can access VPCs over a VPN, improving interconnection security for enterprises' internal networks. The organization manager can control connections between applications and external networks, and assign different elastic IP addresses to various applications, meeting specific application requirements.

9. VDC Metering Management

   The VDC metering capability includes the metering of CPUS, memory, disks, VPCs, elastic IP addresses, security groups, and VM resource quotas. The VDC quota metering feature provides metering data for the service settlement. Managers and tenants use the metering data to perform the offline charging and settlement. VDC quotas that can be metered include the numbers of vCPUs, elastic IP addresses, VPCs, VMs, and security groups, memory capacity, and disk space. The VDC quota metering feature provides functions such as querying VDC metering statistics of a specified period and exporting metering statistics.

10. Self-service O&M Management
    - Capacity management

      The VDC service manager can view resource usage and capacity usage of a VDC. Usage of CPUs, memory, disks, and bandwidth can be queried. The VDC service manager can set capacity thresholds. When the resource usage exceeds the threshold, the system reports an alarm.

    - Performance management

      The VDC service manager can view the performance data of resources in a VDC on OperationCenter, including performance indicators such as CPUs, memory, disks, and networks.

2.1.6 Cloud Service List

After logging in to the VDC self-service portal, users can view multiple types of preset cloud services in service catalogs, including:

- Cloud host

  The cloud host service enables users to use cloud hosts like using local PCs. When applying for the cloud host service on the self-service portal, users can select VM templates and specify VM specifications (such as storage and memory capacity) and submit the application. Users can use log in to cloud hosts using assigned IP addresses after their applications are approved. Users can start, shut down and expand as well as expand and reduce capacity of cloud hosts.

- EBS

  Elastic Block Storage (EBS) provides block storage services to VMs. The EBS service enables users to expand storage space on demand. The OSs on user VMs access block storage space by volume. The EBS service allows users to expand or reduce block storage capacity.

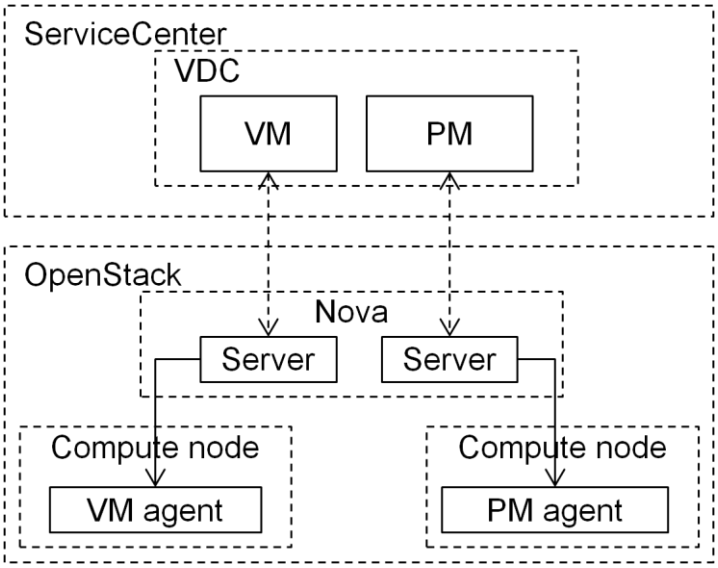
- PM

  The physical machine (PM) service is typically used for applications to which cloud hosts are not applicable. For example, big data analysis services deployed on PMs have a smaller disk I/O and network I/O latencies than those deployed on cloud hosts.

  The system manager defines service specifications and parameters of PMs for tenants to apply for. The domain service manager then reviews those applications. Tenants can log in to the servers using information about the assigned servers.

  For OpenStack Nova, PMs and VMs are all server objects. They are differentiated using the tag field of server (this tag is used by the system and is invisible to users). Each

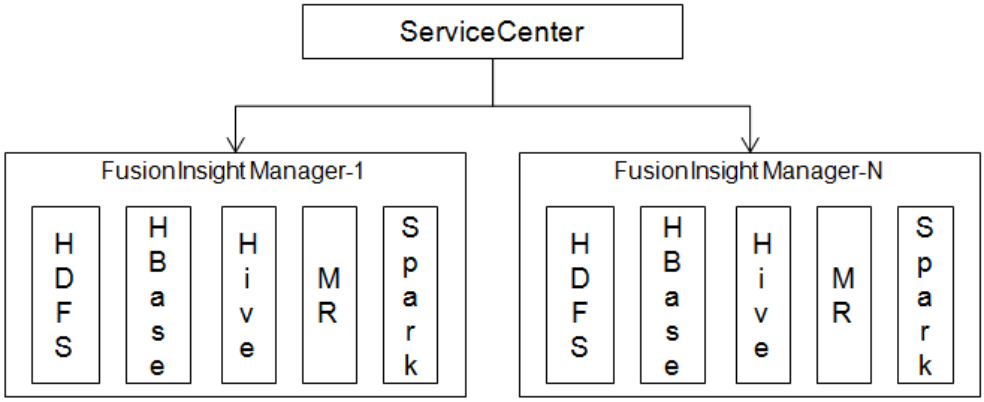
compute node of OpenStack requires the Nova agent. Therefore, the system manager needs to plan compute nodes for provisioning PM and VM services, and deploy them in different host groups. See the following figure.



- FusionInsight

  The FusionInsight service provides distributed file storage (Hadoop distributed file system, HDFS), bill query (Hadoop database, Hbase), offline analysis (Hive), batch processing (MapReduce), and in-memory computing (Spark).

  The following figure illustrates the principle of FusionInsight.



  The system manager needs to construct a FusionInsight resource pool and deploy five FusionInsight services in the pool.

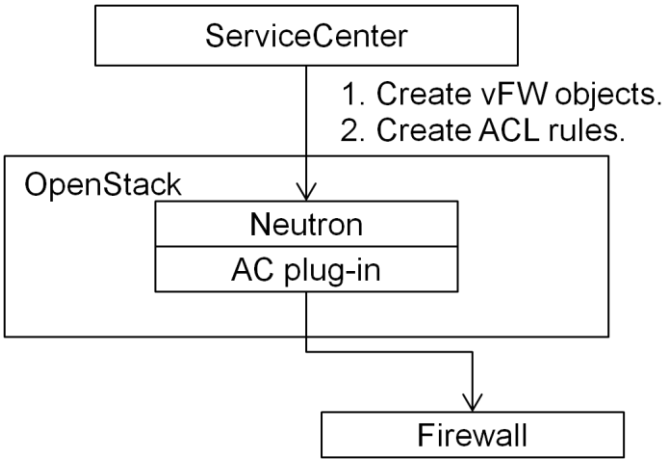
  VDC users apply for FusionInsight services on ServiceCenter and specify the quotas and data access permissions. After the VDC service manager approves the application, ServiceCenter calls the FusionInsight Manager interface to create FusionInsight tenants and users. The VDC manager then assigns data access permissions and associates FusionInsight services for FusionInsight users.

  After they successfully apply for FusionInsight services, VDC users can view the service names and download service certificates in the FusionInsight console of ServiceCenter. Then VDC users can use those service names and certificates to access those services from FusionInsight clients.

FusionInsight Manager supports five types of FusionInsight service clients. VDC users can request the system manager to download those clients in FusionManager and share them with VDC users.

- vFW

  vFWs provide isolation protection for VMs in a VPC. Each VPC has one vFW, and a vFW contains one vFW instance and a maximum of 1000 ACL rules (created by the VDC service manager).



  The vFW service is provided by Neutron, the Neutron plug-in of the AC SDN controller, and a hardware firewall.

  On ServiceCenter, the VDC manager creates a vFW and adds ACL rules. The requests for creating the vFW and adding ACL rules are sent to Neutron, and then Neutron uses the AC plug-in to implement related functions on the hardware firewall.

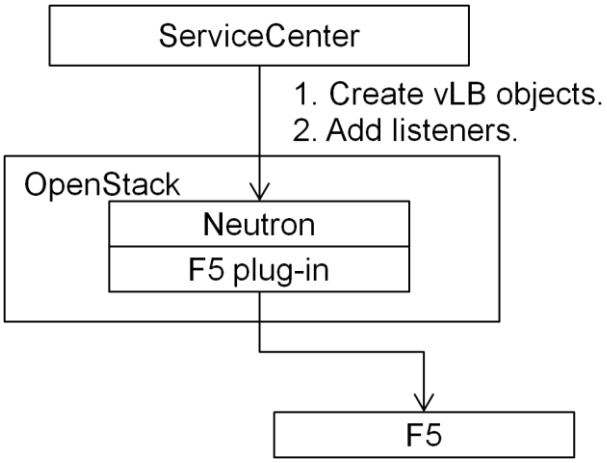
  The following table describes the ACL rule parameters:

Parameter	Description
Protocol	Any, TCP, UDP, or ICMP.
IP version	IPv4 or IPv6.
Source IP address	Source IP address for rule control.
Destination IP address	Destination IP address for rule control.
Source port	The source port is a single port or a port range (format: P1:P2; range: 0-65535).
Destination Port	The destination port is a single port or a port range (format: P1:P2; range: 0-65535).
Priority	Position of the rule in the policy. It starts from 1. If it is not bound to a policy, the rule is blank.
Policy	Allow or Deny.

- vLB

  Virtual Load Balancer (vLB) is a service that automatically distributes access traffic to multiple VMs to balance the load. It enables you to achieve greater levels of fault tolerance in your applications and expand application service capabilities.

The following figure illustrates the principle of vLB:



vLB is provided by Neutron, the F5 plug-in and an F5 device.

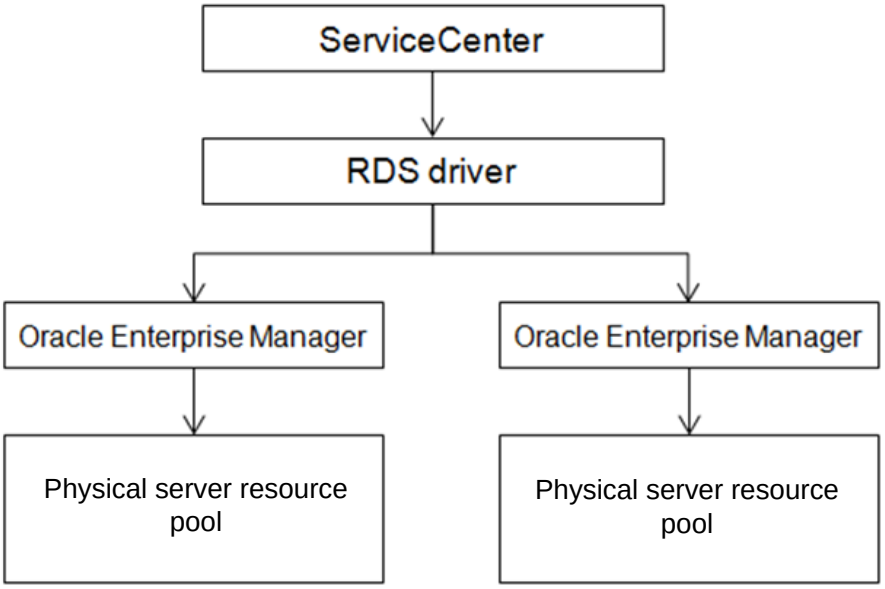
On ServiceCenter, users apply for vLBs on demand and add listeners. Listener parameters include protocol types, front-end network, back-end network, route policies, maximum number of sessions, and health check information. Protocol types include TCP, HTTP, and HTTPS. Route policies include round-robin and least connection. Health check information includes check paths, timeout duration, check period, maximum number of retries, and protocol types.

After the VDC manager approves the vLB service application, ServiceCenter delivers the application to Neutron. Then Neutron uses the F5 plug-in to create a vLB service on the F5 device.

- RDS for Oracle

  With the RDS for Oracle service, tenants can apply for Oracle database instances on demand.

  The following figure illustrates the principle of the RDS for Oracle service.



Multiple PMs form a resource pool for deploying Oracle database instances. Oracle Enterprise Manager takes the resource pool into management as a cluster. Also, Oracle

Enterprise Manager creates Oracle instances on PMs according to user-defined parameter settings. Therefore, to use the RDS for Oracle service, users need to purchase the Oracle Enterprise Manager component.

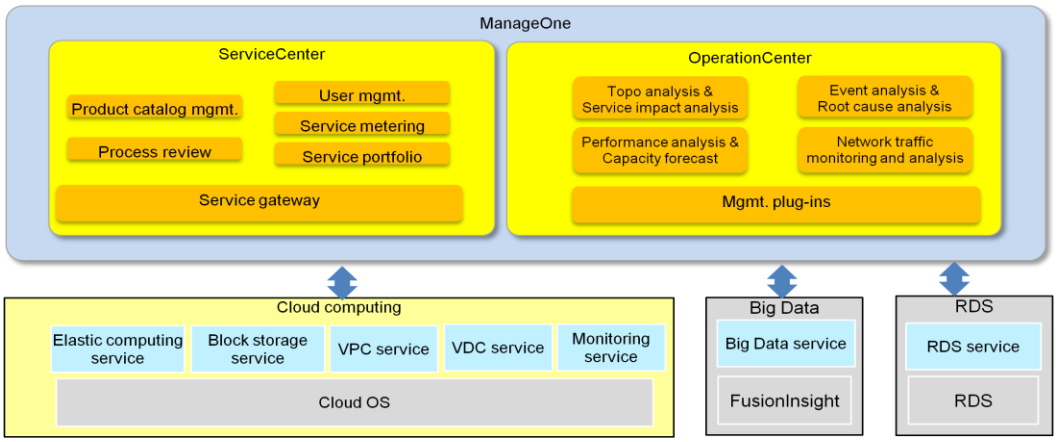
The RDS driver encapsulates the Oracle Enterprise Manager interface into a service for ServiceCenter.

Users apply for the RDS For Oracle service on ServiceCenter. They can apply for a single Oracle instance or a RAC cluster, and configure the computing and storage quotas of a specific database. ServiceCenter calls the RDS driver interface to create an Oracle database instance.

2.2 Unified Management

2.2.1 Overall Architecture

Figure 2-3 Overall architecture of the SD-DC^2 management subsystem



ManageOne is a management software suite for cloud data center management. It has two components, ServiceCenter and OperationCenter.

- ServiceCenter

  Uses cloud and non-cloud resources in resource pools to provide customizable data center services and unified service orchestration and automatic management capabilities.

- OperationCenter

  Implements O&M operations based on scenarios and visualized status, risk, and efficiency analysis for data center services, and works with ServiceCenter to implement data center self-optimization and self-healing based on analysis results.

2.2.2 Application Scenarios

The unified management capabilities of the SD-DC^2 apply to the following scenarios:

- Unified multi-data center management

  Multiple physical data centers must be managed in a unified manner.

- Unified physical and virtual resource management

Physical and virtual resources of a data center must be managed in a unified manner. For example, physical and virtual resources are centrally monitored and managed in a topology.

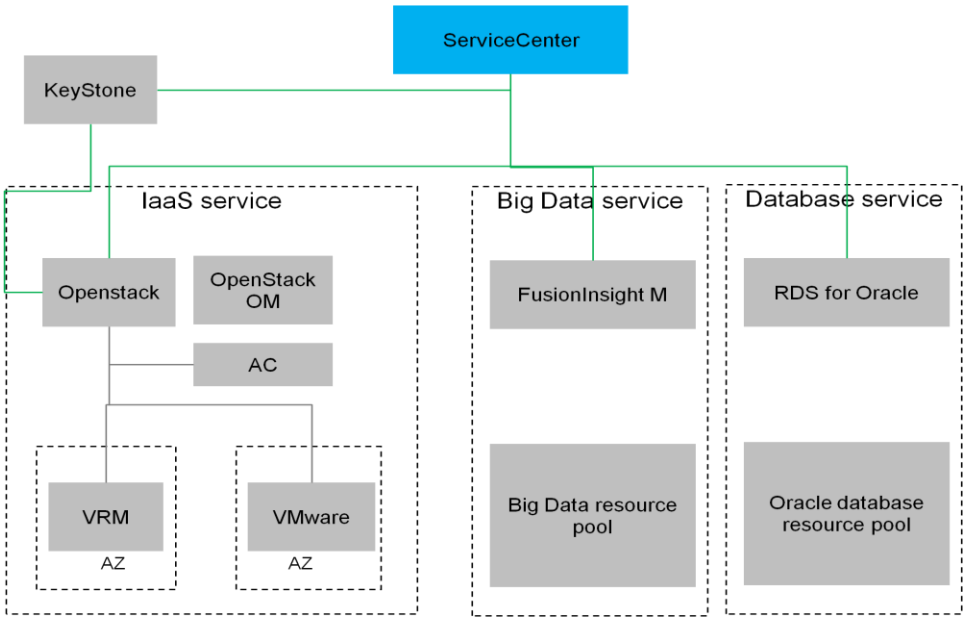
- Unified heterogeneous resource pool management

  Heterogeneous virtualization platforms in a data center must be managed in a unified manner. For example, the vSphere and Huawei UVP virtualization platforms are managed in a unified manner.

2.2.3 Deployment Architecture

2.2.3.1 Operation Architecture

Figure 2-4 Architecture of the operation management subsystem



- ServiceCenter

  It is responsible for tenant management and cloud service management. ServiceCenter is deployed on a remote VM in only one data center.

- OpenStack

  It cloudifies the infrastructure resource pools of the local data center. OpenStack is deployed at each data center. All OpenStack and ServiceCenter systems share the same Keystone.

- FusionInsight Manager

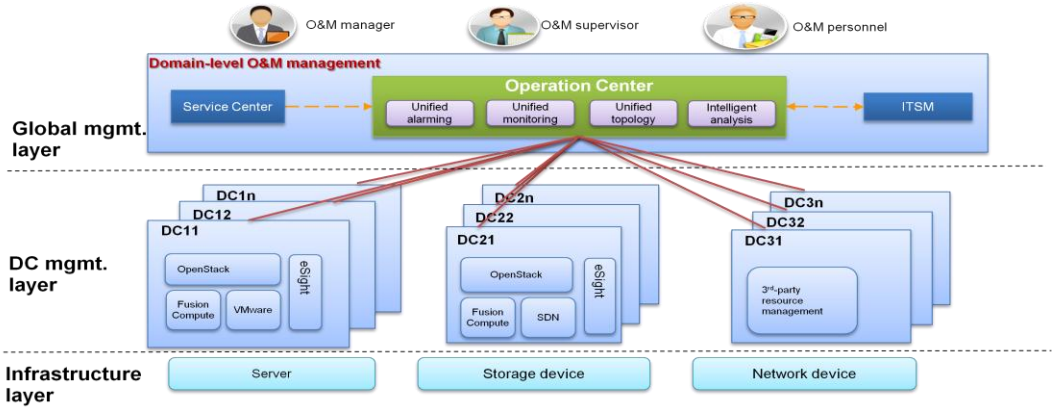
  It manages FusionInsight resource pools, and connects to ServiceCenter to provision the FusionInsight service.

- RDS for Oracle

  It provides the RDS for Oracle service. It works with Oracle Enterprise Manager.

2.2.3.2 O&M Architecture

Figure 2-5 Architecture of the O&M management subsystem



- A local cloud resource management system and a physical device O&M system are deployed in each data center to manage local O&M operations, configuration, and monitoring data collection.
- The domain-level unified O&M management system OperationCenter is deployed on the central node to integrate cloud resource monitoring information and non-cloud resource monitoring information from each data center and perform unified O&M management.
- OperationCenter also supports the correlated analysis of O&M monitoring data and service-related data, and provides value-added functions such as the root cause analysis, service impact analysis, traffic exception analysis, capacity analysis and planning, and service preventive maintenance inspection (PMI) scheduling.

2.2.4 Key Features

2.2.4.1 Private Cloud Operation

1. User management

   User management includes the following functions in a private cloud scenario:

   - VDCs

     virtual data center (VDC) is a basic unit for using virtual resources on ServiceCenter. A VDC is managed by VDC managers. The system manager can set the resource scope and resource quota of a VDC.

     A manager who creates VDCs can create VDC service managers or set other managers to VDC service managers.

   - Domains

     A domain is used to divide resources into logical groups for domain-specific management. Domain-specific management enables different managers to perform operations on resources in different domains.

   - Users

     Managers can add, delete, modify, and query users on the graphical user interface (GUI). Managers can modify users' phone numbers, email addresses, and description.

     The manager can assign a user different operation permissions in different domains by setting the roles of the user.

- Roles

  For ease of use, the system provides default roles to develop services. Users can also add roles as required.

- The system also allows roles to be added, deleted, modified, and queried. Managers can define different roles to assign rights to users. The management operations include:

- Domain service manager: is the ServiceCenter system manager who has the highest permission. Responsibilities of the domain service manager are as follows:

  Manages ServiceCenter, including configuring interconnections between ServiceCenter and resource pool management systems, assigning permissions, and checking software status to ensure that ServiceCenter can run properly.

  Manages services, including managing VDCs, managing domain service catalogs and users, and monitoring domain capacity.

  - Organization service manager: manages all services of the organization.
  - VDC service manager: is responsible for VDC service management, including VDC service applying for, extending, and releasing, VDC service user management, and VDC service catalog management.
  - Service user: is the end user who uses services. The service user can apply for and use services.

- Security policy management

  Users can define security policies, such as account lock policy and password rules, to ensure system security.

2. Service management

   Services refer to usable services provided for tenants. Service management provides the following functions:

   - Service catalog management

     Service managers can define services in a service catalog, such as the service name, icon, and parameters (for example, VM specifications and VM OS types).

     On ServiceCenter, a public service catalog can be defined to be used by all VDCs by default. The domain service manager can publish the public service catalog to specific VDCs through a white list.

     ServiceCenter also provides default service catalogs.

   - Service management

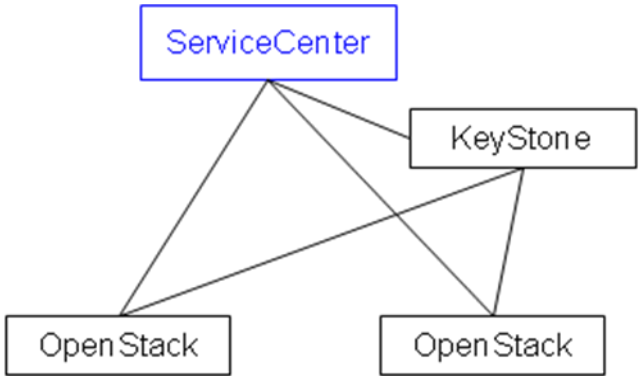
     Services can be applied for, changed, approved, released, and maintained, and application orders can be managed on ServiceCenter.

3. Multi-DC resource management

   ServiceCenter can have resource pools that are scattered in different locations. Those resource pools support OpenStack only, and are connected to FusionCompute and VMware virtual resource pools using OpenStack.

   A VDC can use resources of multiple resource pools. A VDC user can apply for cloud host and cloud hard disk services from the nearest resource pool.

   The following figure illustrates the deployment of ServiceCenter:



2.2.4.2 Unified O&M Management

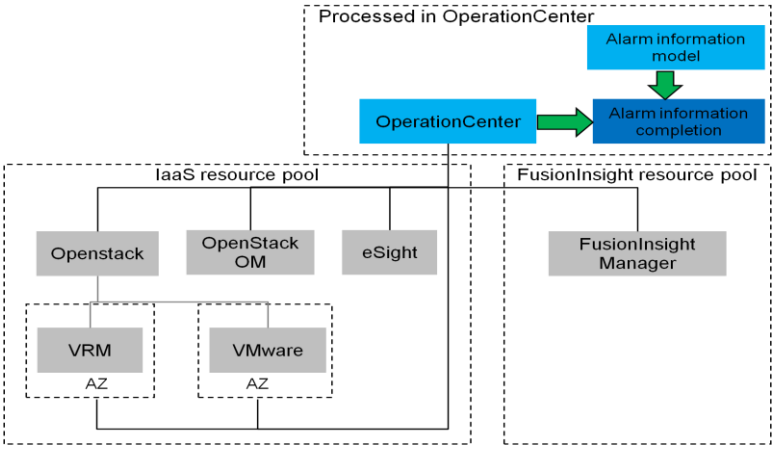
1. Centralized Alarm Management

   OperationCenter collects alarm information of managed objects and allows users to manage alarms. The following topological views are provided:

   - Alarm information about different devices and virtual resources is reported to OperationCenter for unified management.
   - Alarms reported by different alarm objects have different or missing attributes. OperationCenter completes alarm information based on the unified alarm model.
   - In OperationCenter, users can acknowledge alarms, query alarms by type, transfer alarms to work orders, send alarm notifications, and clear alarms, facilitating management.

   Automatic Alarm Information Completion

   The following figure shows the process of automatic alarm information completion.



   The unified alarm information model includes the following information about alarm objects:

   (1) Location information
   - Province
   - Data center
   - Equipment Room
   - Region
   - Cloud platform

- Cluster

(2) Object type information

- Object code
- Object type
- Object subtype
- Original device number (DN) of an object in the monitoring management software
- Alarm object name
- Object name in the monitoring management software
- Logical data center to which an object belongs

(3) Management system information

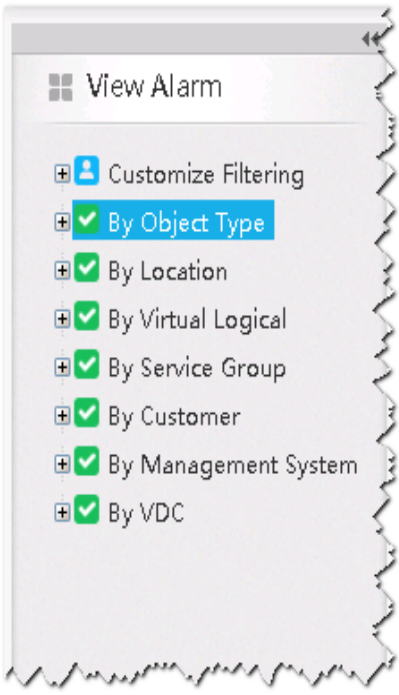
- Owning resource pool management system
- Owning monitoring management software (alarm, performance, and object)
- Others
- Service group
- Customer information
- Last update time
- Logical zone
- Additional information.
- Customized information

Flexible Alarm Handling

OperationCenter supports the following alarm handling operations:

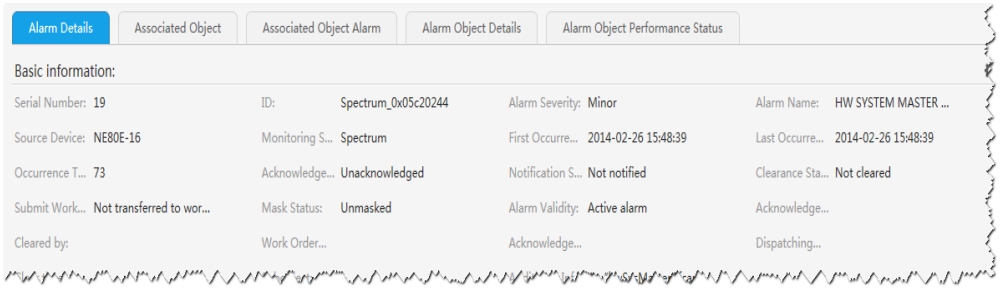
Query alarms from different dimensions, as shown in the following figure.

Figure 2-6 Querying alarms from different dimensions



Handle alarms in different ways, such as masking alarms and transferring alarms to work orders.

Query alarm related information, as shown in the following figure.



2. Centralized Performance Management

OperationCenter collects performance information of managed objects and allows users to manage performance information. The centralized performance management involves the following aspects:

- Real-time performance monitoring

  Monitoring or maintenance personnel can learn device performance in real time and discover performance problems. OperationCenter allows users to save performance indicators of devices they are concerned about to views. Users can select the views when querying the information the next time, which improves operation efficiency.

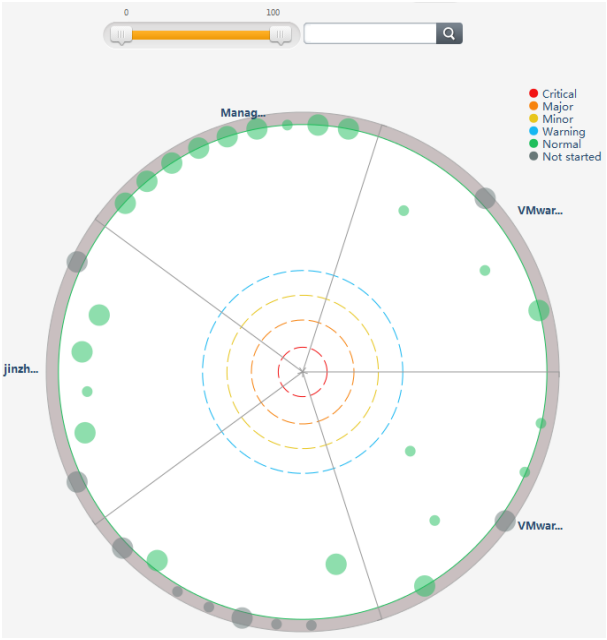
- Top N performance display

  OperationCenter provides Top N views for key indicators. Devices with Top N indicator values are displayed allowing users to identify performance bottlenecks.

- VM sky map

  A VM sky map displays the usage of VMs in the data center. ManageOne allows you to search for the information about a single VM, view VM information by data center, virtualization platform, and cluster, and query the resource information by resource usage.

The following figure shows a VM sky map:



3. Centralized Topology Management

   OperationCenter displays data of managed objects on topologies, enabling monitoring and maintenance personnel to learn the overall condition of a data center. The following topologies are provided:

   - Physical location topology
   - A physical location topology displays devices based on physical locations of devices. Users can learn device locations from a physical view, facilitating fault location.
   - Service application topology
   - A service application topology displays information about managed objects based on services provided by the objects. When exceptions occur in a service, uses can learn information about the object that provides the service.
   - Virtual logic topology
   - On the virtual logic topology, managed objects are logically classified, including physical and virtual resources. A virtual logic view allows users to view the resources by resource pool or zone.
   - Network IP topology
   - A network IP topology displays device information based on physical links. The data of this topology includes the data obtained from other management systems.
   - VDC topology

     This topology displays information about devices of a VDC. The data of this topology is synchronized from other management systems.

   In OperationCenter, uses can also view the physical location topologies of multiple data centers managed by OperationCenter to learn the connections between multiple data centers.

   Permission- and domain-specific management of multiple data centers is supported. That is, different managers manage different data centers. Each manager can only view the topologies of data centers they manage and expand the topologies to learn the hierarchy of data centers.

4. Asset management

   OperationCenter stores asset object information in the Managed Object DataBase (MODB), manages assets using a unified model, and supports query of alarm, performance, and topology information about assets, improving management efficiency.

   OperationCenter collects asset information in two ways:

   - Collects asset information using management software.
   - Manual input

     When asset information cannot be collected using management software, manually import asset information to OperationCenter.

   Asset object management allows users to perform the following operations:

   View object-related information when querying object information, such as alarm and performance information, as shown in the following figure.

   Querying asset information:

- Manage asset information.

  Change asset locations, define asset types, and synchronize asset information to third-party systems.
- Set associated information about assets.

  Associate assets with customers and IP addresses so that you can learn the IP address of an asset object and to whom the asset is assigned.

5. Report management

   OperationCenter provides various report types and generation modes, meeting O&M requirements.

   You can view reports in two ways:

   - Manual reports

     View reports immediately. After a task for manually creating a report is successfully executed, you can view the generated report, export the report, and send the report to specified users by email.
   - Periodic reports

     View the reports generated at the specified period. After a periodic report task is successfully created, OperationCenter generates a report based on the data of the specified period, saves the report, and sends the report to specified users by email. You can choose **Reports** > **Periodic Reports** to view and export the generated periodic reports.

   Characteristics of reports are as follows:

   - Common graphs, such as broken line charts, pie charts, and column graphs are supported.
   - Reports of data center devices, such as servers, VMs, network devices, and storage devices, are supported.
   - Reports can be exported in PDF, Word, and Excel formats.
   - Report logos can be changed, and products support report extension, meeting customers' personalized needs.
   - Reports can be sent to specified email, allowing O&M personnel to learn the data center situation in office.

6. Unified portal

   OperationCenter provides a unified O&M management portal. The portal performs the following functions:

   - Unified entries to other management systems

     - If monitoring or maintenance personnel need to log in to other management systems (such as SDM and eSight), they can click links in **My Workbench** > **My links** on the OperationCenter home page to go to the GUIs of other management systems.

- In addition, **My Workbench** provides the following functions:
  (1) Displays the work orders for the current user to process in the **My To-Do** area.
  (2) Provides a system configuration wizard and quick entries to common setting tasks.
  (3) Provides quick entries to common operation tasks.

– Dashboard

The dashboard displays the general information of a data center. Users can select the following information to be displayed on dashboards:

- Data center–level capacity snapshot information
- Top 5 VMs in a data center in terms of CPU, memory, disk I/O, and NIC I/O
- Cluster-level capacity snapshot information in the data center
- Top 5 physical devices in the data center in terms of CPU, memory, disk I/O, and network I/O

7. Maintenance Permission Management

OperationCenter provides the permission management function for O&M accounts and centralized registration function.

The following concepts exist in OperationCenter permission assignment:

– Roles

A role is a collection of rights, which is used to authorize users. Different roles are given different rights and assigned to users or user groups, so that the users or user groups have corresponding operation rights. Authorizing users or user groups by role ensures orderly rights management.

– Object domains

An object domain is a domain of objects of a certain type. Objects are classified by their locations.

An object domain is specified for users or user groups to implement permission management because the users or user groups have operation rights only on the objects in the domain.

– Users

A user is a management account. A user has corresponding rights after a role and an object domain are specified for the user.

– User groups

A user group is a collection of management accounts. A user group has corresponding permissions after a role and an object domain are specified for the user group. Users in a user group have operation permissions in managed objects specified by the user group.

OperationCenter supports two authentication modes: local authentication and LDAP authentication.

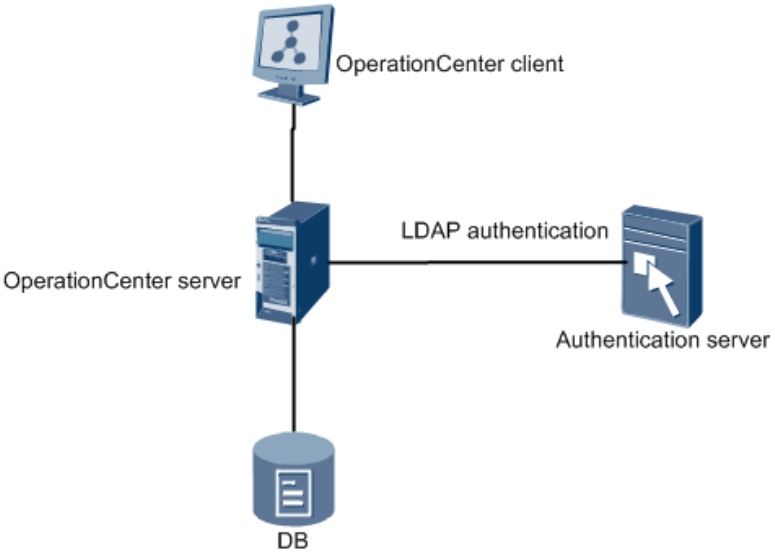
– Local authentication

In local authentication mode, user data is stored in the local database of OperationCenter and used for authentication. A user who logs in to OperationCenter is called a local user.

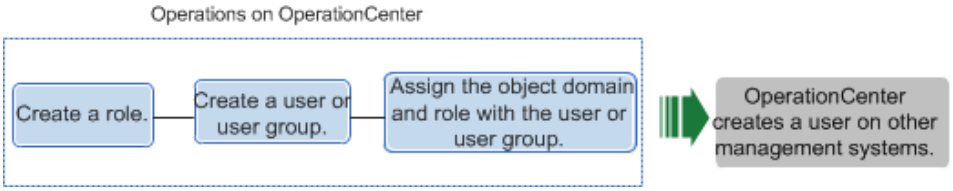
– LDAP authentication

In LDAP authentication mode, user data is stored in the LDAP server and used for authentication. In this case, a user who logs in to OperationCenter is called an LDAP user.

The following figure illustrates the LDAP authentication mode.



The following figure shows the permission assignment process.



Log in to OperationCenter, and create a role. Create a user or user group, and associate an object domain and role with the role. Permissions are then allocated. Users created in OperationCenter can perform operations only in OperationCenter.

OperationCenter creates users and assigns roles to them in other systems.

NOTE

You can configure links of other management systems in the **My Links** area on the OperationCenter home page. Then you can click the link to access other management systems.

8. Cloud resource capacity management

   OperationCenter provides the following cloud resource capacity management capabilities:

   – Capacity analysis

     OperationCenter displays capacity snapshot views and capacity trends at different levels based on physical locations and virtual logic.

     - Capacity views of data center–level CPU, memory, disk, public IP address, and VLAN
     - Capacity views of resource-level CPU, memory, disk, public IP address, and VLAN
     - Capacity views of AZ-level CPU, memory, and disk

NOTE

An AZ is a logical zone of physical resources (including computing, storage, and network resources). Devices in an AZ communicate with each other at Layer 2. From the end users' point of view, shared disks created in an AZ can be attached to any VM in the AZ, and VMs (that are not using local storage) can be migrated between virtual hosts in the AZ.

- Capacity views of cluster-level CPU, memory, and disk
- Capacity views of equipment room–level CPU, memory, disk, public IP address, and VLAN

- Capacity warning

  Capacity thresholds can be flexibly defined based on service requirements. When the capacity exceeds the threshold, alarms of different severities are generated prompting the manager to expand the capacity.

- Capacity statistics report

  Instant and periodic capacity reports can be generated. Periodic reports are sent to specified users periodically.

9. Full text search

   Users can search for the following information in OperationCenter using the full text search function:

   - Resource object information
   - Alarm information
   - Events
   - Log information
   - Online help information

10. Cloud data center health analysis

    OperationCenter supports health analysis for a data center, an area in a data center, computing resource pools, computing clusters, servers, cloud hosts, storage resource pools, storage devices, and network devices. Through health analysis, monitoring personnel, maintenance personnel, and the O&M supervisor can learn the health status of data centers and the factors that affect the data center health, and prevent faults.

    The following assessment is implemented for computing, storage, network, and virtual resources of a cloud data center.

    - Health of computing, storage, network, and virtual resources of a cloud data center is assessed at different levels based on current loads, load trends, and alarms.
    - Risks of computing, storage, network, and virtual resources of a cloud data center are assessed at different levels based on the remaining capacity, predicted resource consumption (remaining time), and pressure.
    - Efficiency of computing, storage, network, and virtual resources of a cloud data center is assessed at different levels based on density, low-efficiency running status, light load distribution, and port usage to discover inefficient devices.

2.2.5 Roles and Typical Processes

2.2.5.1 O&M

Roles

Table 2-1 describes the roles required in database O&M.

NOTE

The roles described in Table 2-1 are predefined in OperationCenter.

Table 2-1 O&M roles

Role	Responsibility	Task
Monitoring personnel	Monitor data centers, identify faults, and instruct the maintenance personnel to rectify the faults in a timely manner.	• Identify faults in a timely manner. • Perform preliminary analysis and fault locating. • Dispatch work orders to the maintenance personnel for troubleshooting. • Trace the troubleshooting process.
Maintenance personnel	Rectify faults to ensure service provisioning in data centers.	• Rectify faults as soon as possible. • Periodically learn the health status of data centers and make analysis for optimization. • Input and manage resources.
System manager	Perform routine operation and maintenance on the system to ensure proper running of the system.	• Configures system functions, such as configuring alarming. • Manages system security, including managing users, security policies, and logs. • Performs routine operation and maintenance to ensure proper running of the management software.

In addition to the preceding roles, the role of the O&M supervisor is also involved during maintenance. This role manages O&M personnel of the entire data center and is responsible for the entire data center, including:

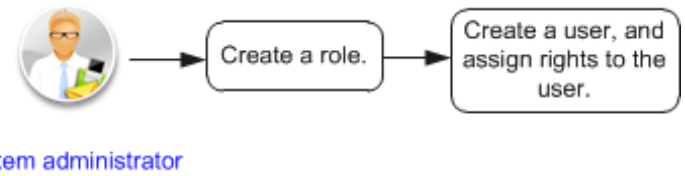
- Learn the data center health status in real time.
- Learn handling progress and results of major problems and complaints.
- View O&M reports periodically.

NOTE

This role of the O&M supervisor is not predefined in OperationCenter, which can be created and granted permissions as required.

Permission assignment

The following figure shows the permission assignment process.



The system manager creates roles in OperationCenter and assigns permissions to the roles.

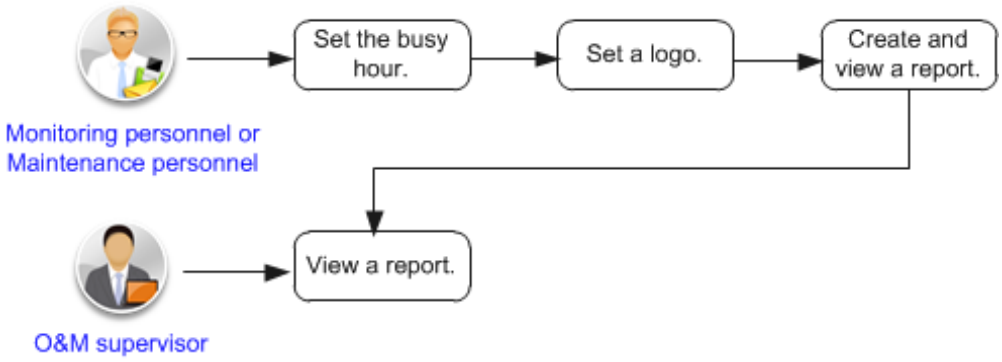
The system manager creates users in OperationCenter as well as specifies user roles and objects the users can operate.

NOTE

OperationCenter provides the centralized registration function. After users are created in OperationCenter, if OperationCenter connects to other management systems (such as eSight), OperationCenter will create the same user accounts on other management systems.

Configuring and viewing reports

The following figure shows the process for configuring and viewing reports.



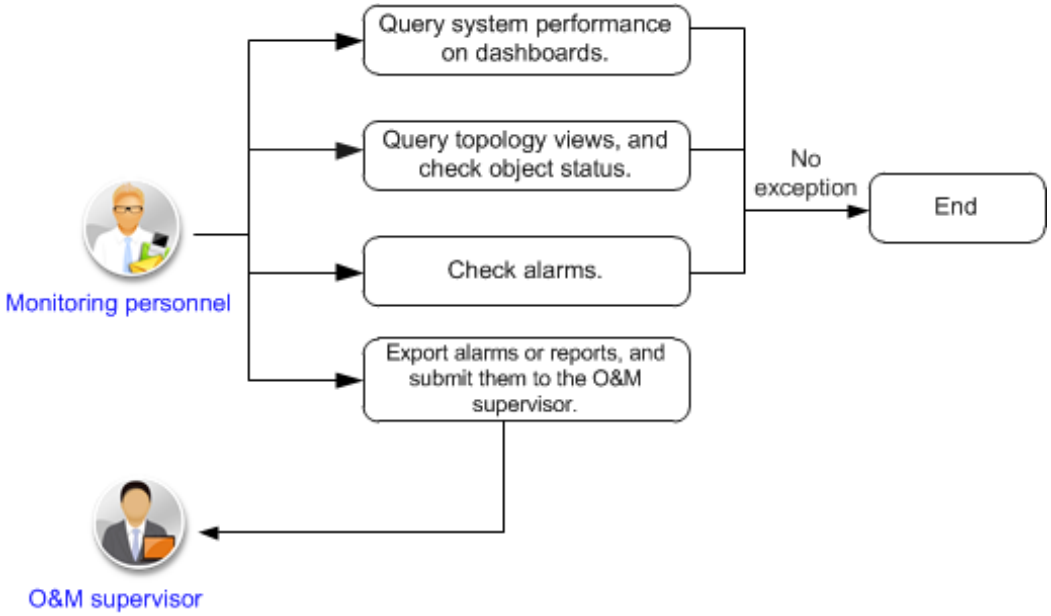
Monitoring or maintenance personnel set report information (such as logos and busy hours for collecting statistics) in OperationCenter.

Monitoring or maintenance personnel create tasks for generating instant or periodic reports (the time period when statistics are collected can be set to busy hours) in OperationCenter, and view instant reports.

The O&M supervisor views reports when needed.

Routine monitoring

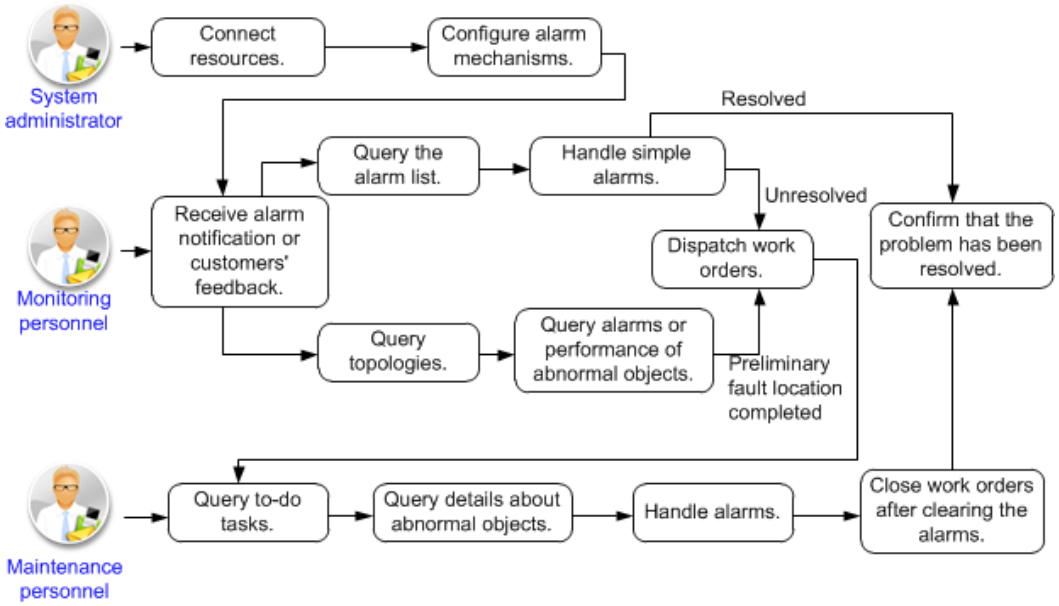
The following figure shows the process of routine monitoring.



- Monitoring personnel monitor the entire data center, including viewing the data center's overall information on the **Dashboard** tab page, viewing the status of topological nodes on topology views, and viewing alarms.
- Monitoring personnel export alarm information or reports and report the data center running status to the O&M supervisor.

Troubleshooting

The following figure illustrates a troubleshooting process.



NOTE

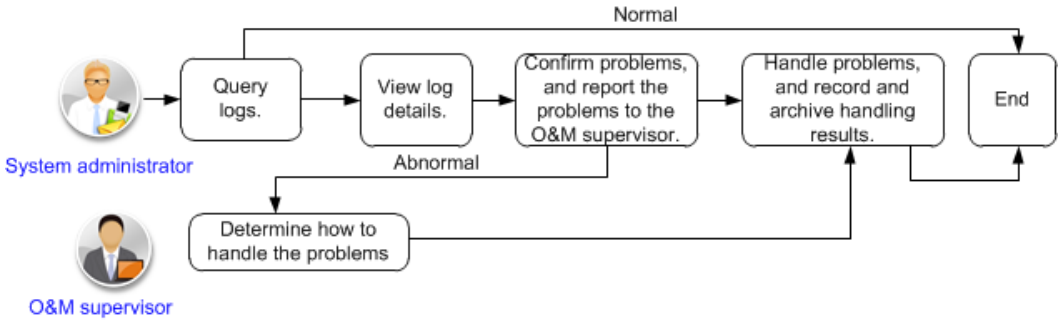
During deployment or maintenance, the system manager connects managed resources to OperationCenter and configures alarm mechanisms (such as alarm notification).

- After receiving problem feedback, monitoring personnel view the alarm list in OperationCenter and implement preliminary locating and handling. If the trouble ticket system (TTS) is deployed, monitoring personnel transfer alarms that they cannot handle to TTS. If TTS is not deployed, monitoring personnel notify maintenance personnel of problems in other ways (for example, by phone).
- Monitoring personnel can quickly locate alarms on topologies.
- After receiving tasks, maintenance personnel log in to OperationCenter to view alarm object information and handle the alarms based on their maintenance experience.
- Process alarms.

  After alarms are cleared, maintenance personnel close work orders.
- Monitoring personnel confirm that problems are resolved.

Log audit

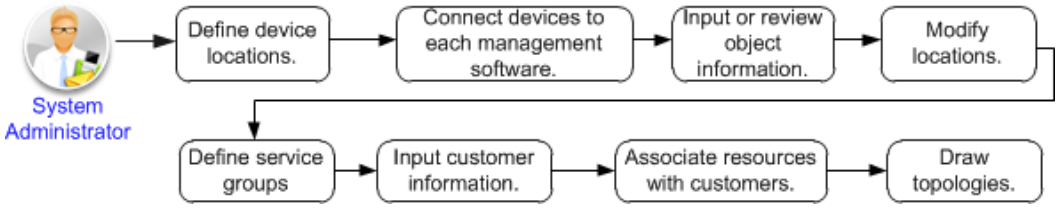
The following figure shows a log audit process.



- The system manager views security logs on OperationCenter to check whether the data center is running properly. If no exception occurs, no further action is required.
- After detecting abnormalities, the system manager confirms problems and reports the problems to the O&M supervisor. After the O&M supervisor makes handling decisions, the system manager or security manager resolves the problems and records handling results.

Configuring resource access

The following figure describes how to configure resource access.



- Define the location. The system manager defines the physical location of the resource to be connected.
- The system manager connects managed devices (resources) in each management system, and configures the interconnection information for OperationCenter and the management systems.
- OperationCenter collects device information from each management system.
- Add or approve objects.
  - Manually input object information when object information cannot be collected from management systems.
  - Review new device information when the information is collected from unreliable management software.
  - Modify locations of new devices.

  Perform this operation when object locations need to be set or modified.
- Define a service group.

  Perform this operation when dimensions of objects need to be defined to facilitate resource, alarm, and event query based on dimensions.
- Add customers.

  Manually enter the customer information to be associated with resources. This supports customer-related O&M information analysis and maintenance.
- Associate resources.

Perform this operation when resource information needs to be associated with IP addresses and customers to facilitate maintenance.

- Draw topologies.

  Draw topological connections among multiple data centers.

2.2.5.2 Operation

Roles

The following table describes roles used for the operation process.

Table 2-2 Operation roles

Role	Responsibility	Task
VDC service operator	A service operator is an end user who uses VPC services.	• Applies for, changes, extends, and releases services. • Queries and maintains resources that have been applied for. • Queries application orders.
System super manager	The system super manager is the ServiceCenter system manager who has the highest permission. The system super manager has the following responsibilities: • Manages ServiceCenter, including configuring interconnections between ServiceCenter and resource pool management systems, assigning permissions, and checking software status to ensure that ServiceCenter can run properly. • Manages services, including managing domain service catalogs and users as well as monitoring domain capacity.	• Connects resource pools to ServiceCenter. • Manages domain service catalogs. • Manages domain software packages, scripts, and VM templates, as well as reviews VDC service applications. • Queries resources provisioned in the domain. • Monitors domain capacity. • Manages domain users.
Organization manager	Manages all services of the organization.	• Manages the service catalogs of the organization. • Creates, modifies, deletes, and queries VDCs in the organization. • Sets VDC quotas of the organization. • Manages users in the organization.

Role	Responsibility	Task
VDC service manager	The VDC service manager is responsible for managing VDC services, including VDC applying for, extending, and releasing, VDC service user management, VDC application template management, and VDC service catalog management.	• Applies for, changes, extends, and releases a VDC. • Queries resources that have been applied for in the VDC. • Queries all application orders in the VDC. • Manages service users in the VDC. • Approves service applications of service users. • Creates resources in the VDC and allocates resources to service users. • Manages the VDC service catalog. • Manages the VDC software library. • Monitors VDC capacity.

Connecting Cloud Resource Pools

The following figure shows the process for connecting cloud resource pools.

Connect resource pools.

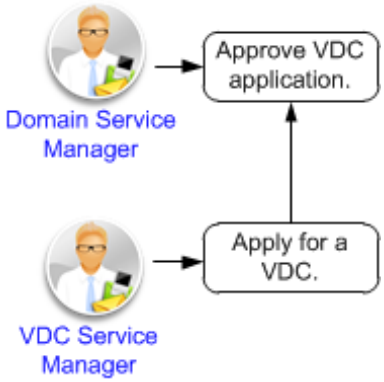
The domain service manager connects cloud resource pools to ServiceCenter to facilitate unified management and resource use. A cloud resource pool can be connected only to one ServiceCenter. Otherwise, exceptions will occur in the system.

Create an organization.

Organization is the unit of allocating resources on ServiceCenter. For example, all virtual resources of an enterprise can be consolidated into an organization for unified management. The resources used by each department can be consolidated into a VDC. The organization manager manages the resources of the organization, whereas the VDC manager manages the sources of the VDC. Level- and permission-specific management is then achieved.

Applying for a VDC

A VDC service manager applies for a VDC from the organization manager using the service mode to manage resources (such as VMs) in the VDC.



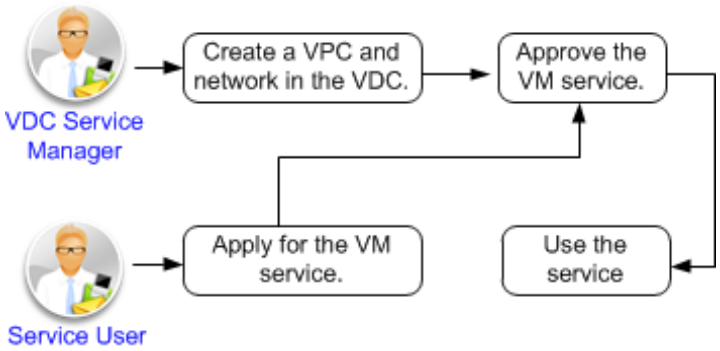
The VDC service manager applies for a VDC service as required.

The organization manager reviews the VDC service application submitted by the VDC service manager. After the application is approved, the system automatically provisions VDC resources.

Applying for a VM

A service user applies for a VM instance of specified specifications to run application programs on the VM and provide services. The following figure shows the process for applying for a VM.

NOTE
The VDC service manager must create VM templates in FusionManager before users apply for VMs.



The VDC service manager creates a VPC and a network in the VDC.

VPCs can provide secure and isolated network environments for users. In a VPC, users can define a virtual network that is equivalent to a traditional network to deploy VM and application instances. Users can use a shared VPC or define a VPC and a network.

VDC users select templates to apply for VM services as required.

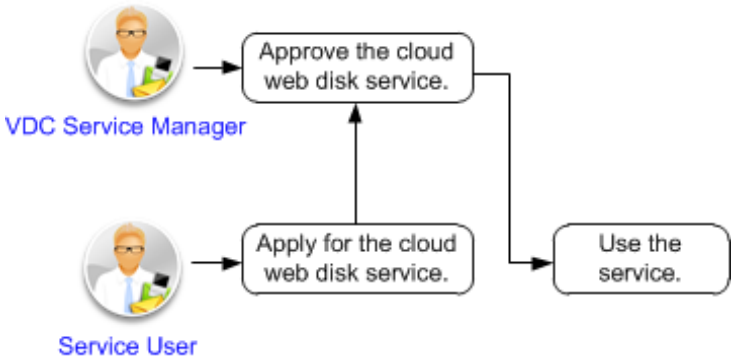
The VDC service manager reviews VM services applied for by VDC users. This step can be skipped if the services do not need to be approved.

VDC users use VM services.

Applying for a Cloud Hard Disk

Cloud hard disks provide persistent and highly reliable block storage services. Applying for a cloud hard disk is a process for applying for creation of a cloud hard disk.

The following figure shows the process for applying for cloud hard disks.



VDC users apply for the cloud hard disk service.

The VDC service manager reviews the application.

VDC users use the cloud hard disk service.

2.3 OpenStack-based Architecture

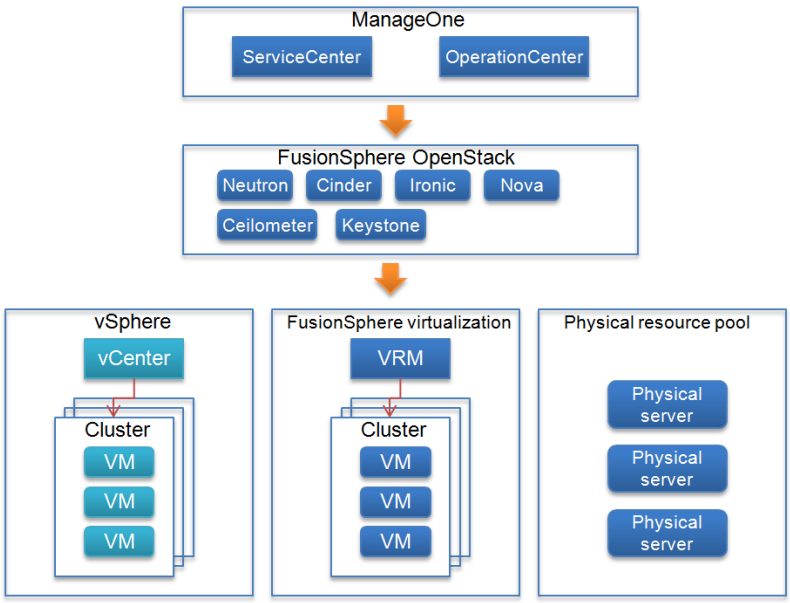
2.3.1 Application Scenarios

The Huawei cloud management platform implements open architecture based on OpenStack. As the most active open-source project, OpenStack involves the participation of multiple vendors. OpenStack can be expanded using plug-ins, supporting third-party devices and heterogeneous virtualization platforms. In this mode, customers are not bound to the infrastructure and virtualization platform of a single vendor, and the open northbound APIs of OpenStack can be easily invoked and managed by third parties. Based on OpenStack architecture, Huawei implements the support for converged resource pools. Converged resource pools apply to the following scenarios:

- Hybrid deployment of heterogeneous virtualization platforms
  Unified management of heterogeneous virtualization platforms is supported. Resource provisioning capabilities are provided for heterogeneous resource pools such as Huawei FusionSphere and VMware.
- Hybrid management of physical servers and VMs
  In some cloud data centers, high-performance and high IOPS applications (such as databases) are deployed on physical servers, and common-performance applications (such as middleware applications) are deployed on VMs.

2.3.2 Logical Architecture

Figure 2-7 Logical architecture of a distributed cloud data center



OpenStack architecture provides support for distributed data centers. The Huawei FusionSphere resource pool, third-party resource pools (such as VMware vSphere), and physical resource pools can be constructed separately, but they are managed by Huawei FusionSphere OpenStack in a unified manner. The ManageOne cloud management platform provides unified monitoring and resource provisioning capabilities for heterogeneous resource pools. The selection of resources is determined according to the resource quality of service (QoS) requirements and resource types (such as vSphere or FusionSphere VM).

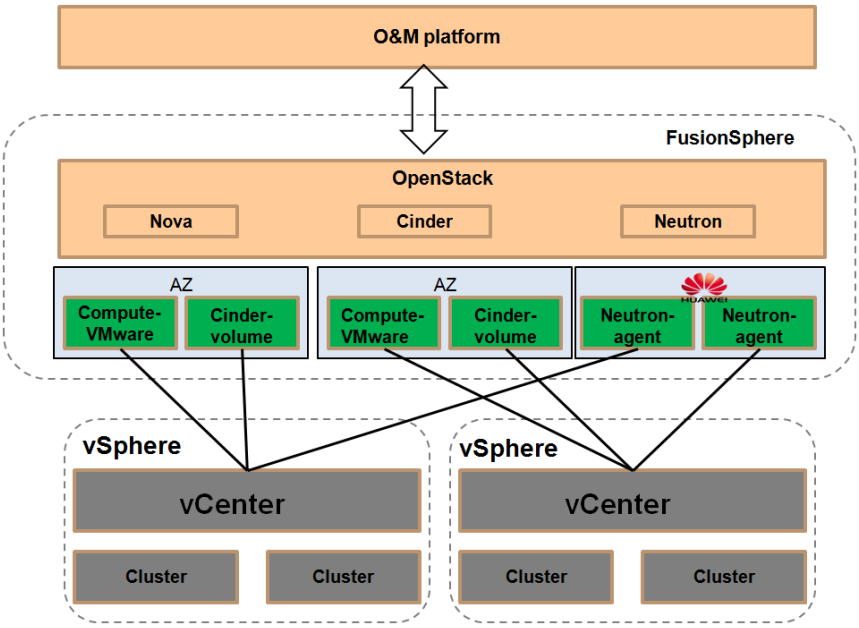
2.3.3 Key Features

1. Unified Management of Heterogeneous Resource Pools

   The Huawei FusionSphere solution provides unified management of heterogeneous virtual resource pools. The heterogeneous management capabilities of Huawei OpenStack provide management of the VMware vSphere virtual resource pool and the Huawei FusionSphere virtual resource pool. Different virtualization platforms are converged into a resource pool that features physical distribution and logical unification. Huawei OpenStack provides OpenStack-based northbound APIs, facilitating management operations such as resource creation and deletion.

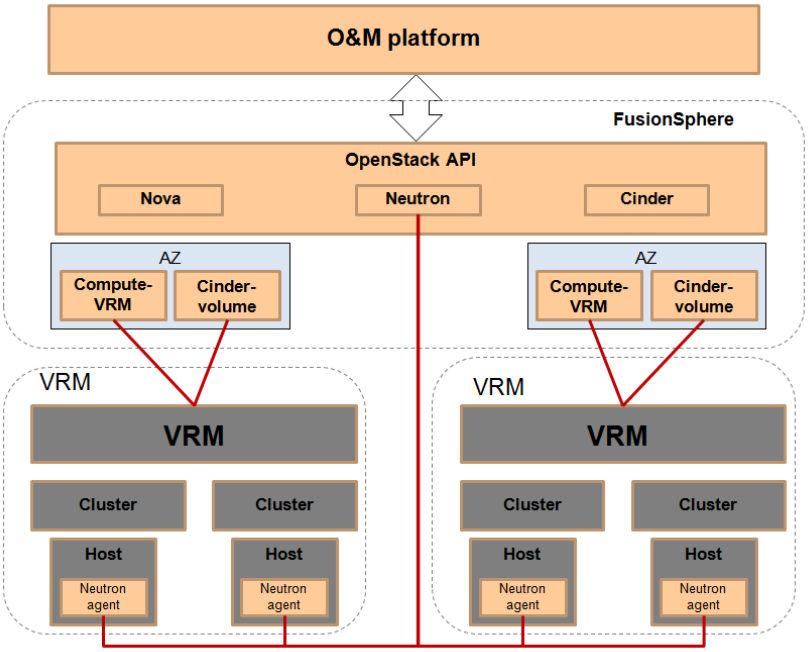
   The ManageOne cloud management platform supports heterogeneous virtual resource management in a VDC. By selecting a resource pool type, tenants can apply for the creation of different VMs and manage different virtualization platforms.

Figure 2-8 Heterogeneous VMware management solution



The previous figure shows the Huawei heterogeneous VMware vSphere management mode. By adding plug-ins to the computing component Nova, storage component Cinder, and network component Neutron, Huawei OpenStack connects to northbound APIs of VMware vCenter and implements life cycle management of VMware resources.

Figure 2-9 Huawei OpenStack management solution for FusionSphere



The previous figure shows the Huawei OpenStack management solution for the FusionSphere virtualization platform. The management mode is similar to that of VMware. By adding plug-ins to Nova, Cinder, and Neutron, Huawei OpenStack connects to the Huawei VRM, and invokes APIs to implement life cycle management of FusionSphere resources.

2. Huawei OpenStack Enhancements

   To resolve the disadvantages of the open-source OpenStack, Huawei implements some enhancements to meet commercial requirements. Huawei OpenStack implements the following enhancements:

   - Enhanced reliability

     Huawei OpenStack implements enhanced reliability based on the open-source OpenStack, including the HA capabilities for all components where single points of failure may occur (such as LBaaS Agent and L3 Agent), eliminating single points of failure from the entire system. Huawei OpenStack also provides monitoring capabilities for the health status of component processes, supporting restart of faulty components and quick restoration of the system. In addition, Huawei OpenStack supports backup of management data. If the system is faulty, management data of the entire system can be restored using backup files.

   - Enhanced manageability

     In terms of manageability, Huawei OpenStack aims to simplify OpenStack maintenance. Huawei OpenStack supports the one-click installation and deployment and the role-based automatic multi-node concurrent installation, improving maintenance efficiency using automatic deployment on the Web UI. Huawei OpenStack supports smooth upgrades and upgrade processes such as upgrade assessment and upgrade confirmation, implementing strict upgrade procedures. Upgrade rollbacks provide reliability assurance for upgrade operations. Huawei OpenStack also supports the collection of system diagnosis information, such as system run logs and operation logs, and supports remote fault locating. Huawei OpenStack supports refined monitoring and fault alarming capabilities, and provides multiple monitoring indicators.

   - Enhanced support for infrastructures by adding plug-ins

     Expansions aim to enhance the usability, reliability, compatibility, and automatic management level of the standard OpenStack, and provide an OpenStack-based cloud platform solution for commercial use. All expansions and enhancements are implemented based on the native standard plug-ins and driving mechanisms of OpenStack. The OpenStack main code is not modified. Drivers from other vendors can be seamlessly integrated with the Huawei OpenStack solution for commercial use, ensuring the openness of OpenStack. The enhanced functions can be easily migrated to OpenStack of a later version when OpenStack upgrades.

3. Resource QoS Management

   Resource QoS management aims to meet customers' requirements on the resource QoS, and provide services of a balance between cost-effectiveness and QoS. To implement resource QoS management, SLAs must be clearly defined for resources. An SLA is an agreement between the service provider and customers to ensure the performance and reliability of services to be provided at certain costs. SLAs are provided to ensure a balance between QoS and costs.

   The SD-DC^2 solution supports the QoS tag group defining for resources. Multiple tags can be added to a resource. The following table provides an example.

Host Group	Tag
Host group 1 (Aggregate 1)	SSD = true, Reliability = high, GPU = true

Host Group	Tag
Host group 2 (Aggregate 2)	Reliability = medium, security = true

Multiple tags are defined for the two host groups to describe the QoS features of the host cluster. For resource selection, the manager can define flavor tags. The flavor tags include resource QoS tags. For example, the high-reliability flavor can be defined as the selection of VMs in host group 1 and the high-security flavor can be defined as the selection of VMs in host group 2. During VM creation, users can select the QoS according to the flavor tags. For example, if a user requires a high-security VM, the scheduling algorithm creates a VM in host group 2 for the user.

2.4 Distributed Storage FusionStorage

2.4.1 Application Scenarios

The FusionStorage applies to the following scenarios:

- Cloud resource pool scenario

  The resource pool scenario applies to large- and medium-sized enterprises or organizations, for example, carriers, finance, and oil sectors, which have high requirements for storage capacity. In the resource pool scenario, FusionStorage pools general-purpose x86 servers deployed in a large-scale cloud computing data center to build up large-scale storage resource pools and provide standard interfaces for accessing block storage data. FusionStorage supports various Hypervisors and cloud platforms, such as Huawei FusionSphere, VMware vSphere, and open-source OpenStack. FusionStorage consolidates scattered resources into centralized storage resource pools, allowing the cloud data center to have strong elastic scheduling capabilities, improving storage resource utilization, and simplifying management.

  Traditional SAN devices cannot achieve linear expansion of performance and capacity. If customers use traditional SAN devices to build storage resource pools, one data center may have SAN devices of different models or from different vendors. Resource usages may be imbalanced among multiple SAN devices, and resources cannot be managed in a unified manner or elastically scheduled. Frequent application data migrations among multiple SAN devices increase O&M costs.

- Database scenario (high IOPS, high bandwidth)

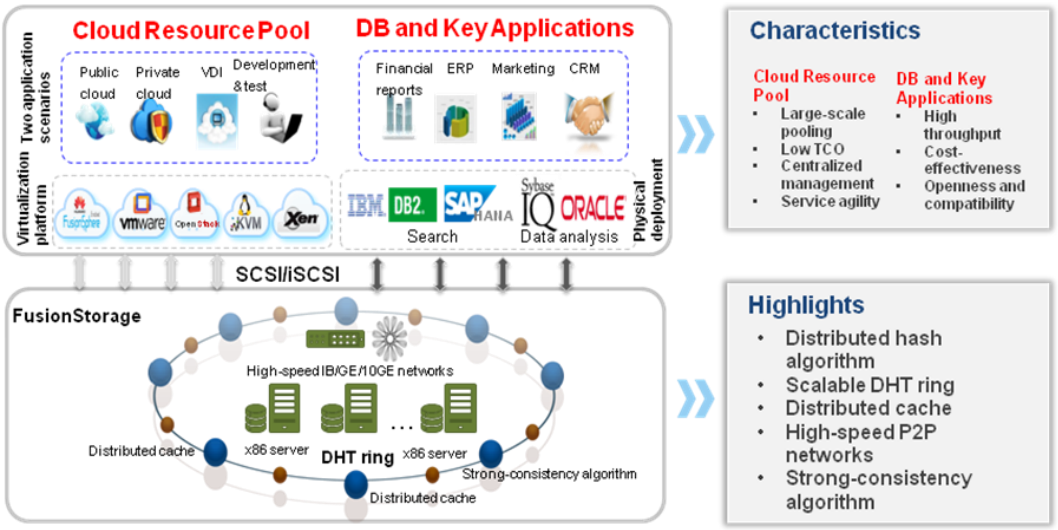
  The architecture consisting of traditional SAN storage and midrange computers has performance bottlenecks due to centralized engines, and therefore cannot provide high I/O bandwidth and scalability. FusionStorage uses the P2P blocking-free switching technology and does not have the performance bottlenecks that may be caused by centralized engines. Therefore, FusionStorage completely eliminates the bandwidth bottlenecks that may occur between computing and storage nodes.

  - FusionStorage uses the distributed design and allows all I/O operations to be parallelly processed.
  - FusionStorage supports distributed cache. Compared with traditional SAN storage, FusionStorage increases the cache capacity by N times and improves the hit rate and I/O efficiency for accessing data hotspots.

FusionStorage supports the high-speed, low-latency InfiniBand network, which helps eliminate network bottlenecks. Therefore, FusionStorage can be used in database applications,

for example, bandwidth-demanding online analytical processing (OLAP) database applications and IOPS-demanding online transaction processing (OLTP) database applications.

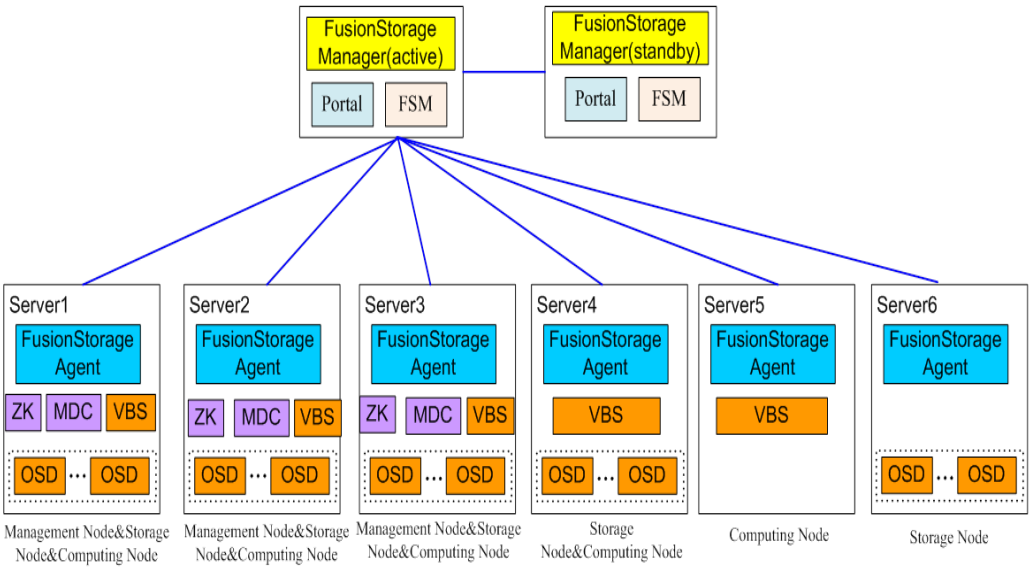
Figure 2-10 Application scenarios of FusionStorage



2.4.2 Logical Architecture

The following figure shows the logical deployment of the FusionStorage system.

Figure 2-11 Logical deployment of FusionStorage



FusionStorage Manager: FusionStorage management module, which is deployed on two hosts to work in active/standby mode and provides O&M functions, such as alarm reporting, monitoring, logging, and configuration.

FusionStorage Agent: FusionStorage agent module, which is deployed on each host to enable the host to communicate with the FusionStorage Manager node. A FusionStorage Agent node can collect the monitoring and alarm information of its host, receive upgrade packages during a component upgrade on the host, and perform the upgrade.

ZooKeeper (ZK): A FusionStorage system needs three, five, or seven ZK processes to form a ZK cluster, which provides the arbitration service for the Metadata Controller (MDC) cluster during the election process. Ensure that at least three ZK processes are deployed and more than half are active.

Metadata Controller (MDC): A metadata control component of FusionStorage for controlling distributed cluster node status, data distribution rules, and data rebuilding rules. A FusionStorage system needs at least three MDC nodes, which form an MDC cluster. During the system startup, the ZooKeeper cluster elects the active MDC node, which then monitors the status of other MDC nodes. If the active MDC node fails, the ZK cluster then elects another active MDC node from all proper MDC nodes. Each storage resource pool has its own MDC node. If this MDC node fails, the active MDC node specifies another properly running MDC node to host this pool. One MDC node can manage up to two storage resource pools. The MDC process can start on each storage node. An MDC process automatically starts if the user adds a storage resource pool. One FusionStorage system allows up to 96 MDC processes.

Virtual Block Storage (VBS): Deployed on all hosts as a VBS cluster to manage volume metadata and provide the distributed storage access point (AP) service over SCSI or iSCSI interfaces, enabling computing resources to access distributed storage resources. The VBS node communicates with all Object Storage Device (OSD) nodes deployed in the storage resource pool that is accessible to the VBS node. Therefore, the VBS node can access all hard disks in the storage resource pool. Each host has one VBS process running by default. However, you can deploy multiple VBS processes on one host to improve host I/O performance. During the VBS startup, the active MDC node communicates with the VBS nodes and elects the active one.

Object Storage Device (OSD): Deployed for each disk on all servers to perform specific I/O operations. One host can have multiple OSD processes deployed. When SSD cards are used as the main storage, multiple OSD processes can be deployed on one SSD card to maximize the SSD card usage and performance. For example, one 2.4 TB SSD card supports a maximum of six OSD processes, each of which manages the I/O operations for 400 GB of space.

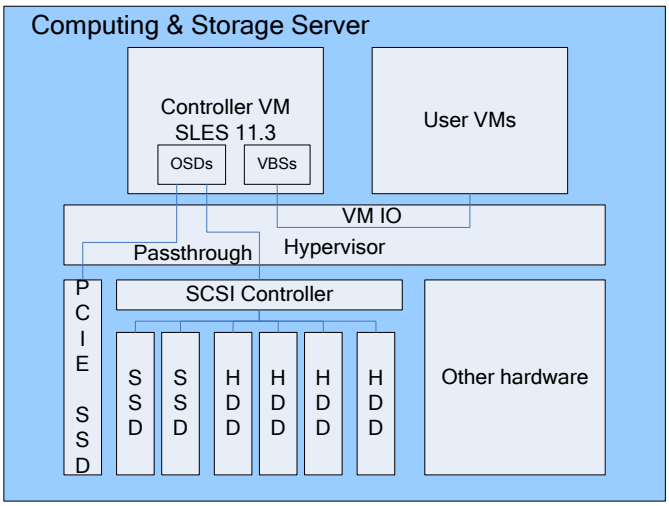
2.4.3 Software Deployment

FusionStorage deployment requires at least three servers. This section describes how to deploy each module of FusionStorage:

VBS/OSD Deployment

VBS and OSD nodes support the converged or separated deployment of computing and storage resources. OSD nodes must be deployed on at least three servers, whereas VBS nodes can be deployed based on service requirements.

1. Converged deployment of computing and storage resources
   (1) Deployment for the VMware Hypervisor



OSD and VBS nodes are deployed on the controller VM. PCIe SSDs of the server or HDDs and SSDs under the SCSI controller pass through to the OSD nodes on the controller VM for storage media management.

The VBS node uses iSCSI interfaces to provide ESXi with the block storage service. When configuring an iSCSI target for ESXi, configure VBS iSCSI service ports for the local server first and configure VBS iSCSI service ports for other servers to serve as the multipathing backup service ports. User VMs on the local server preferentially access the storage services provided by the VBS node on the controller VM of the local server.

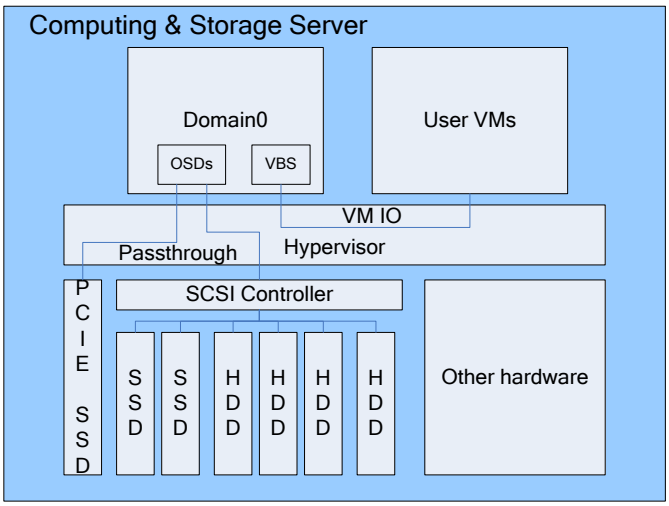
If the controller VM or VBS node on the local server malfunctions, ESXi automatically switches to the standby iSCSI service port to continuously provide storage access services for user VMs. After the controller VM or VBS node on the local server recovers, ESXi automatically switches back to the iSCSI service port provided by the local VBS node.

In addition, the system requires a storage device with an independent storage channel to install ESXi and the controller VM.

Controller VM specifications:

- 8 vCPUs
- Sufficient memory, storage, and network resources. For details about resource requirements, see section 2.4.5.1 "Hardware Platform."

(2) Deployment for the Huawei UVP Hypervisor



The OSD and VBS nodes are deployed in Domain 0. PCIe SSDs of the server or HDDs and SSDs under the SCSI controller pass through to the OSD node in Domain 0 for storage media management. User VMs on a server access the storage services provided by the VBS node of the Domain 0 VM on the server.

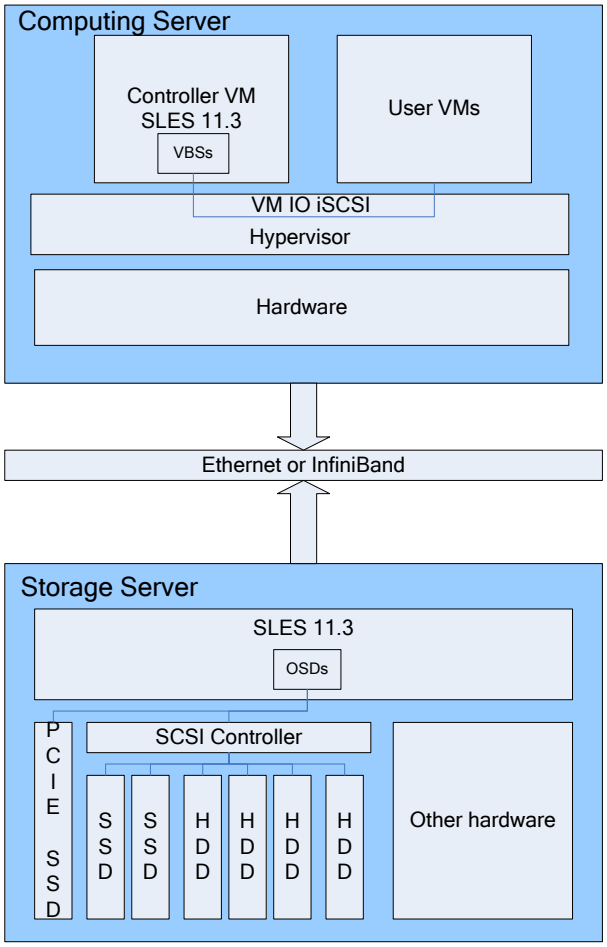
Huawei UVP and Domain 0 can be installed on one or two HDDs attached to the SCSI controller, or be installed on a storage device with an independent storage channel.

Resource requirements of FusionStorage in Domain 0:

- 4 vCPUs
- Sufficient memory, storage, and network resources. For details about resource requirements, see section 2.4.5.1 "Hardware Platform."

2. Separated deployment of computing and storage resources

   (1) Deployment for the VMware Hypervisor



The OSD node is deployed on an independent storage server that runs the SLES 11.3 OS and manages PCIe SSDs of the server or HDDs and SSDs under the SCSI controller. The VBS node is deployed on the controller VM.

When configuring an iSCSI target for ESXi, configure VBS iSCSI service ports for the local server first and configure VBS iSCSI service ports for other servers to serve as the multipathing backup service ports. User VMs on the local server preferentially access the storage services provided by the VBS node on the controller VM of the local server.

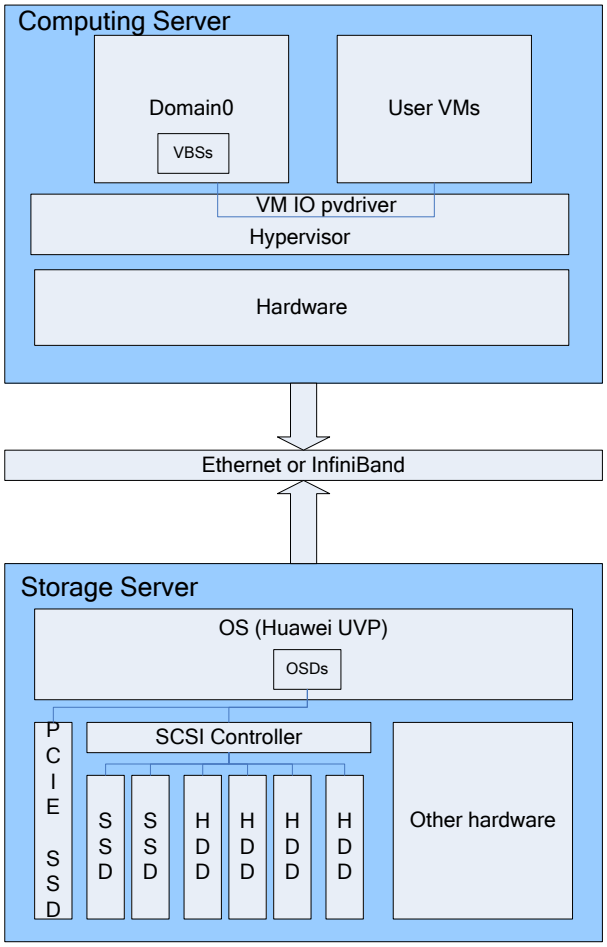
If the controller VM or VBS node on the local server malfunctions, ESXi automatically switches to the standby iSCSI service port to continuously provide storage access services for user VMs. After the controller VM or VBS node on the local server recovers, ESXi automatically switches back to the iSCSI service port provided by the local VBS node.

ESXi and the controller VM are deployed on the storage devices of the local server.

Controller VM specifications:

- vCPUs
- Sufficient memory, storage, and network resources. For details about resource requirements, see section 2.4.5.1 "Hardware Platform."

(2) Deployment for the Huawei UVP Hypervisor



The OSD node is installed on an independent storage server that runs Huawei UVP OS and manages PCIe SSDs of the server or HDDs and SSDs under the SCSI controller. The VBS node is deployed in Domain 0. User VMs on a server access the storage services provided by the VBS node of the Domain 0 VM on the server. Huawei UVP and Domain 0 are installed on the storage device of the server.

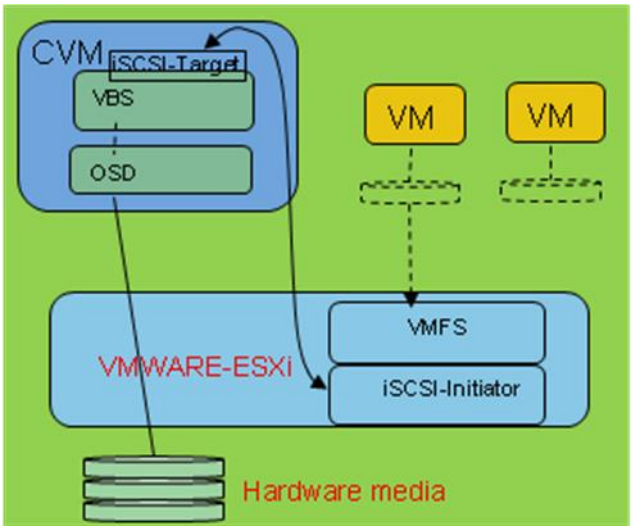
Resource requirements of FusionStorage in Domain 0:

- 4 vCPUs
- Sufficient memory, storage, and network resources. For details about resource requirements, see section 2.4.5.1 "Hardware Platform."

2.4.4 Key Features

2.4.4.1 SCSI/iSCSI Block Interfaces

The VBS node of FusionStorage provides block interfaces over SCSI or iSCSI. SCSI interfaces enable the host accommodating the VBS node to access storage resources. SCSI interfaces apply to FusionStorage deployed on physical servers or FusionStorage deployed in the FusionSphere or KVM solution. iSCSI interfaces enable other hosts or VMs that do not have the VBS nodes deployed to access storage resources. iSCSI interfaces apply to the VMware vSphere and Microsoft SQL Server clusters.



The SCSI protocol supports SCSI-3 persistent reservations and non-persistent reservations. The persistent reservations are used in HANA clusters, whereas the non-persistent reservations are used in MSCS clusters.

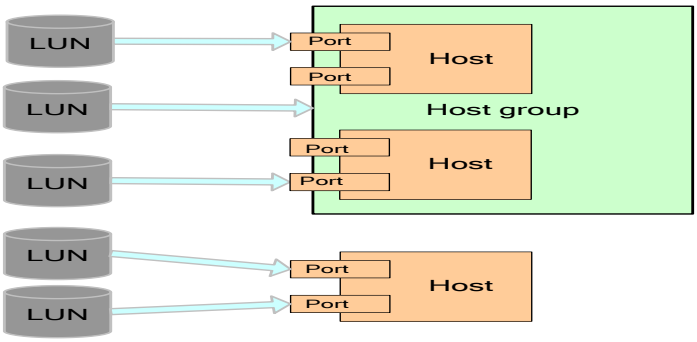
In the FusionStorage system, the VBS node offers an iSCSI target. Computing resources on a server can connect to the iSCSI target using the server initiator to access block storage resources on FusionStorage. Furthermore, FusionStorage supports the following standards to ensure secure access over the iSCSI protocol:

Challenge Handshake Authentication Protocol (CHAP) identity authentication to ensure trustful and secure access from clients. The CHAP protocol periodically authenticates the identity of the peer end through three-way handshake when the links are initially set up and after links are available. CHAP provides protection against replay attacks from the peer through the use of an incrementally changing identifier and of a variable challenge-value. It limits the time for being exposed to an attack.

LUN Masking to authorize a host to access LUNs. For a SAN storage device, hosts use LUNs as local storage devices, and therefore data maintenance is performed on the hosts. In this case, isolate hosts' access to LUNs, preventing one host from damaging data of another host. LUN Masking binds LUNs to host bus adapter (HBA) world wide names (WWNs) and ensures that the LUNs can be accessed only by specified hosts or host groups. One host can have multiple LUNs, and one LUN can be bound to multiple hosts. In virtualization scenarios where advanced LUNs are demanded, configure one host to have multiple LUNs. If cluster systems, such as Oracle RAC, require shared volumes, bind one LUN to multiple hosts.

The core functions of LUN Masking are implemented by the mapping among ports, hosts, host groups, and LUNs.

Figure 2-12 Mapping required for implementing LUN Masking



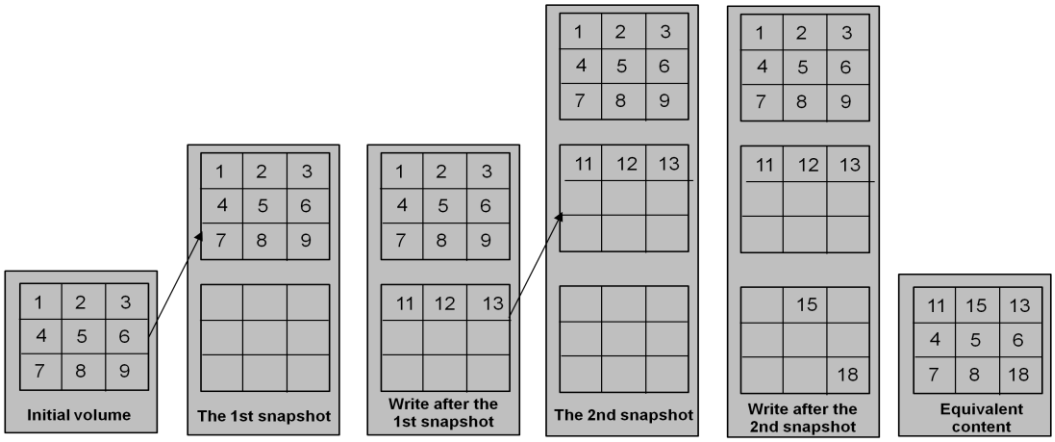
LUN Mapping binds LUNs to ports on storage devices so that hosts can access different LUNs using these ports. LUN Mapping can be used if a storage system concurrently provides data storage services for multiple applications and the hosts of these applications locate in different geographical areas.

2.4.4.2 Snapshot

FusionStorage provides the snapshot mechanism, which allows the system to capture the status of the data written into a logical volume at a particular point in time. The data snapshot can then be exported and used for restoring the volume data when required.

FusionStorage uses the redirect-on-write (ROW) technology when storing snapshot data. Snapshot creation does not deteriorate performance of the original volume.

The following figure illustrates the FusionStorage snapshot mechanism.



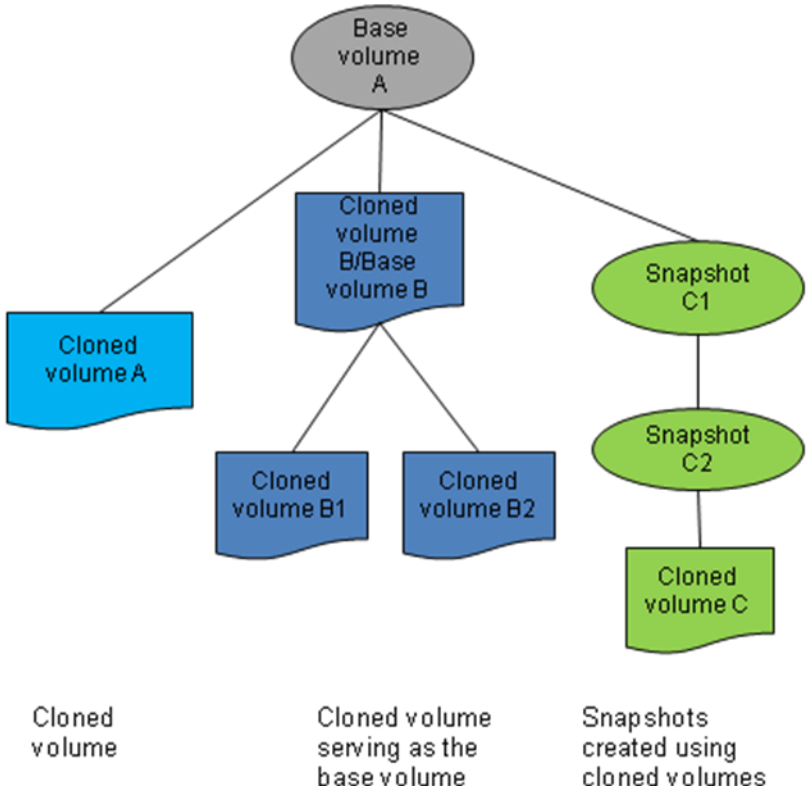
2.4.4.3 Linked Cloning

FusionStorage provides the linked cloning mechanism so that multiple cloned volumes can be created for a volume snapshot. The data in the cloned volumes is the same as that in the snapshot. Subsequent modifications to a cloned volume do not affect the snapshot or other cloned volumes.

FusionStorage supports a linked cloning ratio of 1:256, which can significantly improve storage space utilization.

A cloned volume has all the functions of a common volume. You can create snapshots for a cloned volume, use the snapshot to restore the data in the cloned volume, and clone the data in the cloned volume.

The following figure illustrates the linked cloning mechanism.



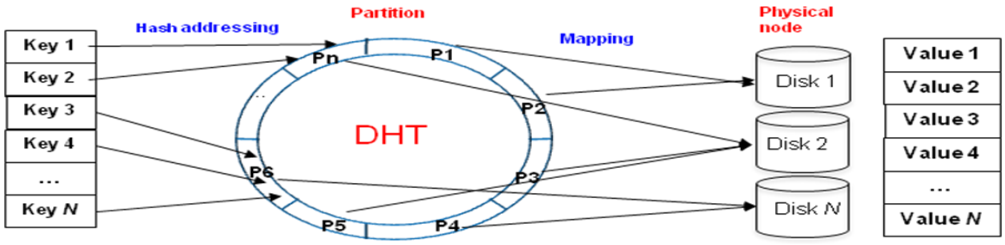
2.4.4.4 Elastic Expansion

The distributed architecture of FusionStorage supports elastic capacity expansion without any performance deterioration.

1. DHT Routing Algorithm

   FusionStorage employs the DHT mechanism. Using the DHT algorithm, each storage node stores and routes a part of data, and all storage nodes store and route data in the entire system.

   Traditional storage devices typically employ the centralized metadata management mechanism, which allows metadata to record the hard disk distribution of the logical unit number (LUN) data with different offsets. For example, the first 4 KB of data in LUN1+LBA1 is distributed on LBA2 of the thirty-second hard disk. On a traditional storage device, each I/O operation initiates a query request to the metadata service. As the system scale grows, the metadata size also increases. However, the concurrent operation capability of the system is subject to the capability of the server accommodating the metadata service. In this case, the metadata service may become a performance bottleneck of the system. Different from traditional storage devices, FusionStorage employs the DHT algorithm for data addressing. The following figure illustrates the mechanism of the DHT algorithm on FusionStorage.



FusionStorage sets the hash space to 2^{32} and divides the hash space into N equal parts. Each part is a partition, and all these partitions are evenly allocated to hard disks in the system. For example, the system has 3600 partitions by default. If the system is equipped with 36 hard disks, each hard disk is allocated 100 partitions. The "partition-hard disk" mapping has been configured during system initialization and will be flexibly adjusted with the changes of hard disks in the system. The mapping table requires only small space, and FusionStorage nodes store the mapping table in the memory for rapid routing purposes. In this regard, the routing mechanism of FusionStorage is different from that of any storage arrays. FusionStorage does not employ the metadata management mechanism and therefore will never confront with the performance bottleneck of the metadata service.

An example is provided as follows: When an application needs to access the first 4 KB of data in LUN1+LBA1, FusionStorage first constructs "Key = LUN1 + LBA1/1M", calculates the hash value for this key, performs modulo operation for the N value, gets the partition numbers, and then obtains the hard disk of the data based on the "partition-hard disk" mapping.

The DHT algorithm has the following characteristics:

- Balance: Data is distributed to all nodes as evenly as possible, thereby balancing load among all nodes.
- Monotonicity: When new nodes are added to the system, system data is evenly distributed to all nodes again. However, data migration is implemented only on new nodes, and the locations of data on the existing nodes are not adjusted.

2. Smooth Expansion

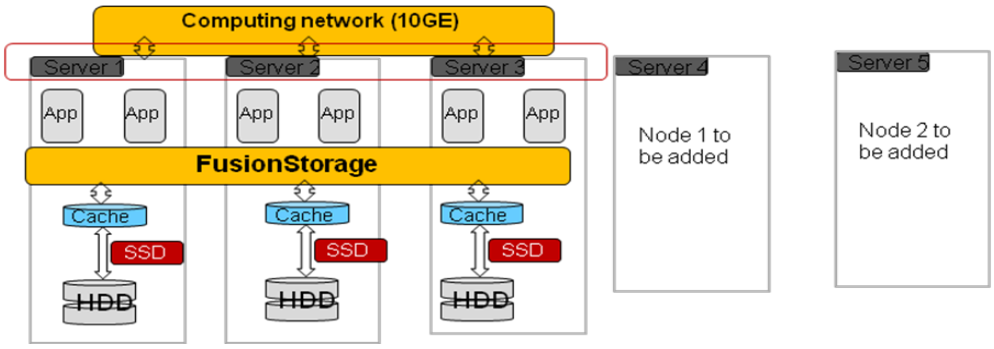
FusionStorage uses the distributed architecture to support easy capacity expansion and ultra-large storage capacity.

FusionStorage ensures rapid load balancing after capacity expansion and avoids migration of a large amount of data.

FusionStorage supports flexible capacity expansion and allows both concurrent and separate expansion of compute nodes, hard disks, and storage nodes. When compute nodes are added, the storage capacity is also added. After capacity expansion, computing and storage resources are still integrated.

FusionStorage evenly distributes engines, caches, and bandwidths to each server, ensuring that the system IOPS, throughput, and cache linearly increase with the increase of nodes.

The following figure illustrates the mechanism of smooth expansion on FusionStorage.



3. High Performance

   FusionStorage uses its distributed architecture to organize the dispersedly distributed, low-efficiency SATA or SAS disks into an efficient storage pool that provides similar functions as a SAN device but provides higher I/O throughput to maximize the storage performance.

   FusionStorage uses SSDs to replace HDDs as high-speed storage devices and the InfiniBand network to replace GE or 10GE networks to provide higher bandwidth. Therefore, FusionStorage can be used for processing massive data in real time and meet the performance-demanding requirements.

4. Distributed Engine

   FusionStorage uses distributed stateless engines. These engines are deployed on each server that requires access to FusionStorage, thereby preventing performance bottlenecks that may be caused by traditional, centrally deployed engines. Moreover, these distributed engines deployed on standalone servers consume much fewer CPU resources but provide much higher IOPS and throughput than centrally deployed engines do.

   An example is provided as follows: A system contains 20 servers that need to access the storage resources provided by FusionStorage, and the bandwidth that each server provides for the storage plane is 2 x 10 Gbit/s. One VBS node is deployed on each server (that is, each server has one storage engine), so that the total throughput can reach 400 Gbit/s (20 x 2 x 10 Gbit/s). With the growth of the cluster scale, storage engines can be linearly added, thereby eliminating the performance bottlenecks that may be caused by centralized engines in traditional dual-controller or multi-controller storage systems.

5. Distributed Cache

   FusionStorage integrates computing and storage resources and evenly distributes caches and bandwidths to each server.

   Each disk on FusionStorage servers uses independent I/O bandwidths, preventing a large number of disks competing limited bandwidths between computing devices and storage devices in an independent storage system.

   FusionStorage can use certain server memory as the read cache and NVDIMMs or SSDs as the write cache. Caches are evenly distributed to all nodes. The total cache size on all servers is far greater than that provided by external storage devices. Even when using large-capacity, low-cost SATA disks, FusionStorage can still provide one to three times higher I/O performance.

   FusionStorage can use SSDs for caching data. In addition to providing high capacity and the write cache function, the SSDs can collect statistics on and cache hotspot data, further improving system performance.

6. Global Load Balancing

The DHT mechanism of FusionStorage ensures that the I/O operations performed by upper-layer applications are evenly distributed on the hard disks of various servers and therefore load is globally balanced.

The system automatically distributes data blocks on the hard disks of various servers. Data that is frequently or seldom used is evenly distributed on the servers, thereby preventing hotspots in the system.

FusionStorage employs the data fragment distribution algorithm to ensure that primary and secondary copies are evenly distributed to different hard disks of the servers. In this way, each hard disk contains the same number of primary and secondary copies.

When a node is added or deleted due to a failure, FusionStorage employs the data rebuilding algorithm to balance load among all nodes after system rebuilding.

7. Distributed SSD Storage

   Huawei FusionStorage supports a distributed SSD storage system for high-performance applications and provides higher I/O performance than traditional hard disks, such as SATA or SAS disks.

   FusionStorage abstracts the PCIe SSD cards configured on storage nodes into a virtual storage resource pool to provide high-performance read and write for applications.

   FusionStorage supports Huawei-developed SSD cards and mainstream PCIe SSD cards developed by other vendors.

8. High-Speed InfiniBand Network

   FusionStorage supports the InfiniBand network designed for high-bandwidth, low-latency applications. The InfiniBand network has the following characteristics:

   - Provides a data rate of 56 Gbit/s and allows high-speed connection setup.
   - Uses standard multi-layer fat-tree networking and allows smooth capacity expansion.
   - Provides a communication network where congestion hardly occurs and avoids data switching bottlenecks.
   - Provides minor communication delays within nanoseconds and transmits computing and storage information promptly.
   - Provides lossless network QoS and ensures data integrity during transmission.
   - Allows multi-path communication for active and standby ports and ensures communication path redundancy.

2.4.4.5 High Reliability

1. Cluster Management

   FusionStorage manages the system in clusters. If a server or hard disk becomes faulty, it can be automatically isolated from the cluster and therefore has no adverse impact on system services. Details are as follows:

   ZK: The ZK node provides the arbitration service for electing the active MDC node. It also stores the metadata that was generated during the storage system initialization, for example, data routing information including the "partition-hard disk" mapping. An odd number of ZK nodes are required in the system. At least three ZK nodes must be deployed. The ZK nodes can function only when more than half of the deployed ZK processes are active. The number of ZK nodes cannot be added once the system is deployed.

   MDC: MDC nodes are deployed as an MDC cluster. The system initially has three MDC nodes deployed. When a resource pool is added, the system automatically starts or specifies an MDC node for this pool. The ZK process helps elect an active MDC node among multiple MDC nodes. The active MDC node monitors the status of other MDC

nodes. If detecting that an MDC node fails, the active MDC node will restart the failed one or specify another properly running MDC node for hosting the resource pool. If the active MDC node fails, the ZK cluster then elects another active MDC node from all proper MDC nodes.

FusionStorage Manager: Two FusionStorage Manager nodes are deployed working in active/standby mode.

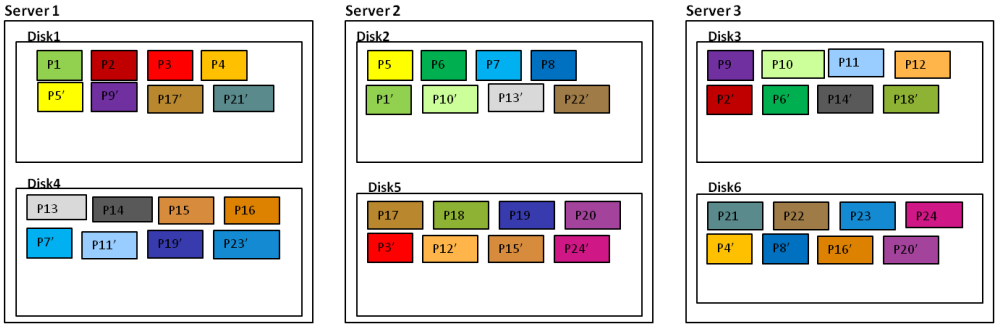
OSD: OSD nodes are deployed working in active/standby mode. The MDC node monitors the status of the OSD nodes in real time. If the primary OSD node of the host where the specified partition resides becomes faulty, storage services will be automatically switched to the secondary OSD node, thereby ensuring service continuity.

2. Multiple Data Copies

To ensure data reliability, FusionStorage stores two or three identical data copies for one piece of data. Before storing data, FusionStorage fragments the data to be stored on each volume in the system at the granule of 1 MB and then stores the data fragments on servers in the cluster based on the DHT algorithm.

The following figure shows the multiple data copies that FusionStorage stores. As shown in this figure, for data block P1 on disk 1 of server 1, its data copy is P1' on disk 2 of server 2. P1 and P1' are two data copies of the same data block. If disk 1 becomes faulty, P1' can take the place of P1 to provide storage services.

Figure 2-13 Multiple data copies stored by FusionStorage



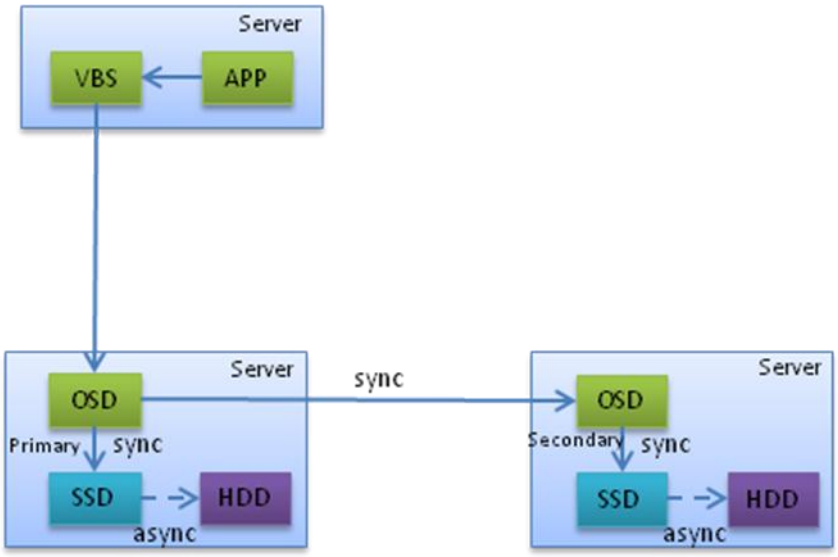
3. Data Consistency

When a user successfully writes application data into the storage system, the system automatically duplicates the data into multiple copies and stores them on different disks. Then the user can read the data from any of the disks.

FusionStorage uses multiple methods to ensure data consistency between data copies:

(1) Synchronous write of data copies

When the VBS node sends a write request to the specified primary OSD node, the OSD node synchronizes this write request to the secondary OSD node while writing it to the server hard disk. This synchronization process is implemented based on the I/O number, thereby ensuring that the sequence of the I/O operations received by the primary OSD node is the same as that synchronized to the secondary OSD node. After the write request is processed on both primary and secondary OSD nodes, a success message is returned to the application. The following figure shows the process of synchronous data copy write.



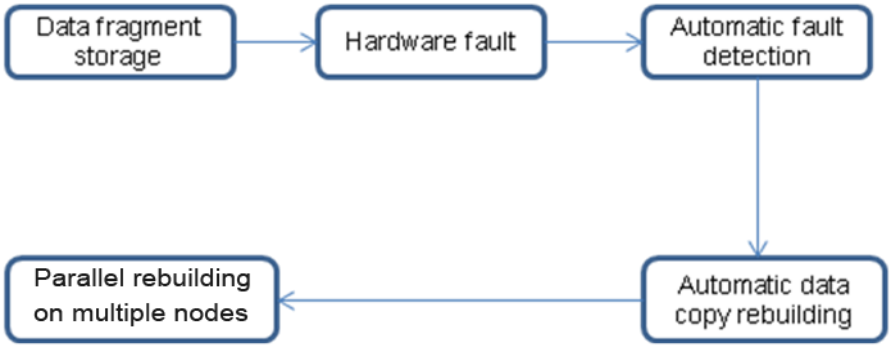
(2) Read repair

FusionStorage supports the read repair mechanism. If failing to read data, FusionStorage automatically identifies the failure location. If the data cannot be read from a disk sector, FusionStorage retrieves the data from other copies of the data on another node and writes the data back into the original disk sector. This mechanism ensures that the total number of data copies does not decrease and data among data copies is consistent.

4. Rapid Data Rebuilding

Each hard disk in the FusionStorage system stores multiple data blocks (partitions), whose data copies are scattered on other nodes in the system based on certain distribution rules. If detecting a hard disk or server fault, FusionStorage automatically repairs data in the background. The repair mechanism allows FusionStorage to simultaneously restore a minimal amount of data on different nodes because the data copies are stored on different storage nodes. This mechanism prevents performance deterioration caused by restoration of a large amount of data on a single node, and therefore minimizes adverse impacts on upper-layer services. The following figure shows the automatic data rebuilding process.

Figure 2-14 Data rebuilding process



FusionStorage supports parallel and rapid troubleshooting and data rebuilding:

- Data blocks (partitions) and their copies are scattered in the resource pool. If a hard disk is faulty, its data can be automatically rebuilt in the resource pool efficiently.

- Data is distributed to different servers so that data can be obtained or rebuilt even if a server is faulty.
- Load can be automatically balanced between existing nodes in the event of node failures or capacity expansion. You do not need to adjust application configuration to obtain larger capacity and higher performance.

5. Power Failure Protection

   A server power failure may occur while the system is running. In this case, the metadata and cached data stored in the memory may be lost due to the power failure. To prevent data loss in such cases, FusionStorage uses NVDIMMs or SSDs to store and restore metadata and cached data.

   Each server running the FusionStorage software must be equipped with the NVDIMM or SSD cache, so that the server can write the metadata and cached data into the flash of the NVDIMM or SSD cache upon a power failure. After the power supply is resumed, FusionStorage restores the data stored in the flash back to the memory.

   The following figure shows the NVDIMM, PCIe SSD card, and SSD used for power failure protection.

6. Hard Disk Reliability

   FusionStorage supports hard disk S.M.A.R.T detection, slowly rotating/fast rotating disk detection, hard disk SCSI fault handling, and hard disk scan. It allows upper-layer services to conduct read repair, remove a faulty disk, rebuild data, mark a bad block, scan valid data in disks, handle errors caused by S.M.A.R.T threshold exceeding, and handle slowly rotating disks.

   - Hard disk scan for valid data

     FusionStorage periodically scans valid data on hard disks to prevent silent data corruption. If detecting a bad sector, FusionStorage immediately repairs the sector.

   - Bad Sector Tag (BST)

     If a bad sector exists when the system scans disks or reads data, an access error is reported. Then the FusionStorage system attempts to perform read repair first. If all copies of the data are unavailable, FusionStorage marks the bad sector in the BST, generates an alarm, and restores the data in the application layer.

   - Hard disk health check

     The FusionStorage system monitors the S.M.A.R.T information and I/O processing of hard disks, identifies the hard disk in the suboptimal state, automatically rebuilds data originally stored in this hard disk, and removes the hard disk from the cluster.

   - Hard disk error detection

During I/O processing, FusionStorage proactively identifies hard disk errors, such as WP, ABRT, and DF errors. If detecting such an error, FusionStorage automatically rebuilds data and removes the faulty hard disk from the cluster.

2.4.4.6 Multiple Resource Pools

FusionStorage V100R003C30 supports multiple resource pools to facilitate use of different storage media and fault isolation. A pair of FusionStorage Manager nodes can manage multiple resource pools. Multiple resource pools share the ZK cluster and active MDC node.

Each resource pool has its own MDC node. When a resource pool is created, the system automatically starts an MDC process for the pool. The system supports up to 128 resource pools and 96 MDC nodes. If more than 96 resource pools are created, the system will not start new MDC processes but specify existing MDC processes for the resource pools.

One MDC node can manage at most two resource pools. The MDC node of a resource pool initializes this resource pool, including dividing partitions and storing the partition and OSD view data into the ZK disk. If this MDC node fails, the active MDC node will specify another properly running MDC node for hosting this resource pool.

FusionStorage V100R003C30 supports offline volume migration among multiple resource pools.

The support for multiple resource pools complies with the following rules:

- If two identical data copies are stored for a piece of data, one resource pool can contain a maximum of 96 disks. If three identical data copies are stored for a piece of data, one resource pool can contain a maximum of 2048 disks. If the number of disks exceeds the upper limit, create a new resource pool.
- The hard disks in one resource pool must be of the same type and the same size. Different types of hard disks must be added to different resource pools. If the disk sizes in a resource pool are different, the resource pool capacity is determined by the smaller disks.
- The cache media in one resource pool must be of the same type. Different types of cache media must be added to different resource pools.
- When a resource pool is created, each storage node must have the same number of hard disks. The gap between hard disk numbers on different servers cannot be greater than 2, and the proportion of the gap to the maximum number of hard disks on a server cannot be greater than 33%.
- One server may have different types of hard disks, but these hard disks belong to different resource pools.

2.4.4.7 VMware VAAI

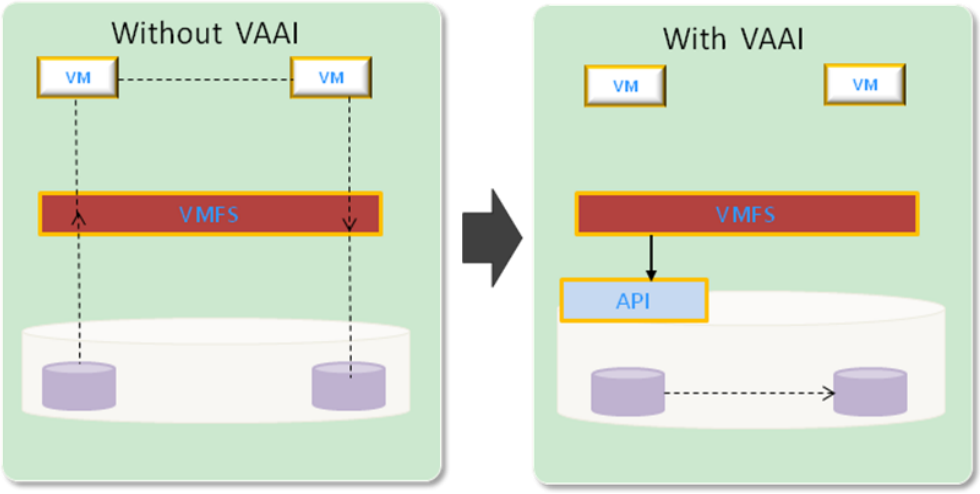
VMware proposes vStorage APIs for Array Integration (VAAI) that offload storage-related operations onto storage devices to improve VM performance. Storage devices must comply with certain standards to implement these APIs.

FusionStorage implements the APIs for the following features:

1. Copy Offload

   Copy Offload is also known as Full Copy or Clone Blocks. When a VM is cloned or when a VM is created using the template, a large number of data blocks need to be copied. Copy Offload enables data to be copied on the storage devices, rather than on ESXi servers, thereby significantly reducing resource overheads on the ESXi servers. This feature also applies to storage vMotion scenarios.

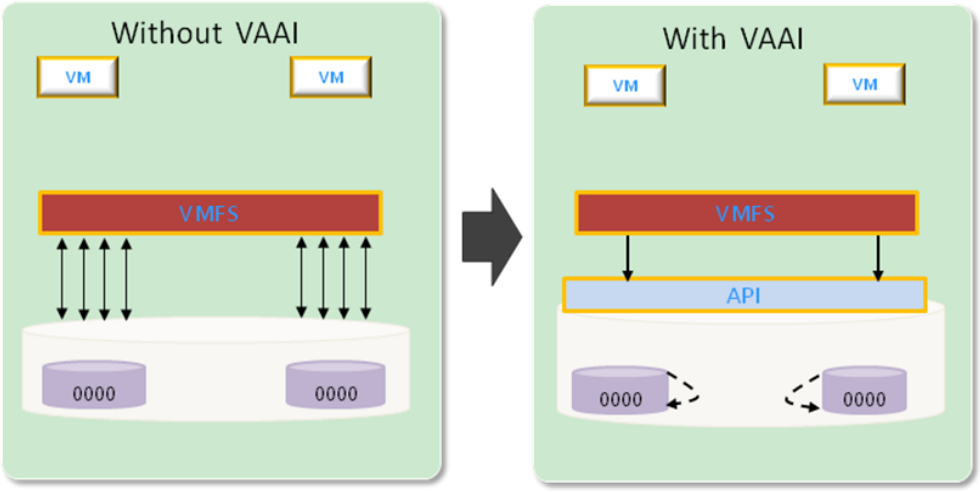
Figure 2-15 Copy Offload mechanism



2. Block Zeroing

   Block Zeroing allows users to zero out virtual disks immediately when creating a VM, which ensures data security and high storage performance and significantly reduces the exchanges between ESXi hosts and storage devices.

Figure 2-16 Block Zeroing mechanism



3. Automatic Test and Set (ATS)

   The virtual machine file system (VMFS) allows multiple hosts to concurrently access one shared logical volume, which is indispensable for running vMotion. The VMFS has a built-in security mechanism that prevents a VM from being run or modified concurrently by more than one host.

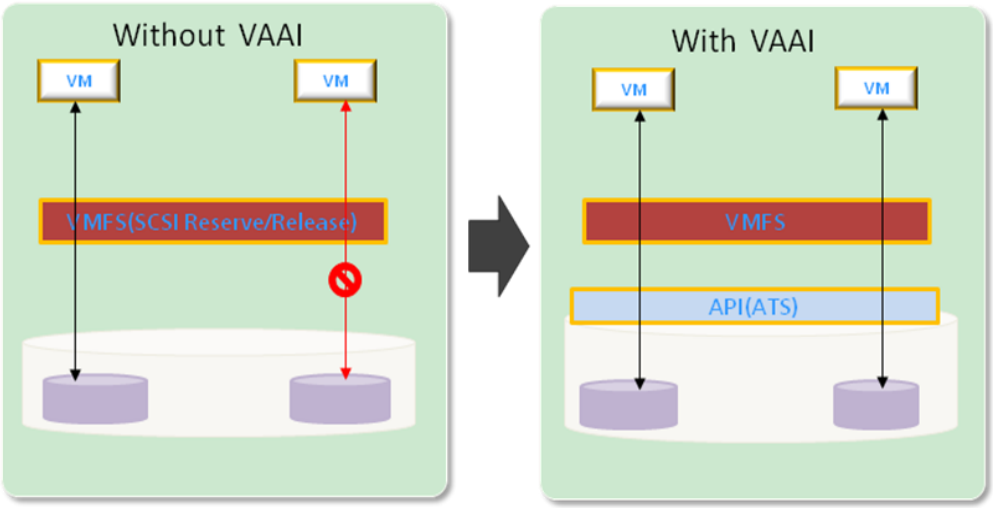
   vSphere uses SCSI reservations as the locking mechanism for traditional files. SCSI reservations use the RESERVE SCSI command to lock the entire logical volumes during the execution of storage-related commands, for example, the increase or occurrence of incremental snapshots.

   The SCSI reservations help prevent conflicts but postpone the storage work, because hosts can proceed to write data only upon receiving the RELEASE SCSI command

issued by the logical volumes. The Atomic Test and Set (ATS) feature is a hardware-assistant locking mechanism. It supports locking on storage arrays offline.

Therefore, this feature can be used to lock certain data blocks instead of the entire logical volume. The rest logical volumes are still accessible to hosts, which prevents storage performance deterioration. In addition, this feature uses the VMFS datastores to allow more hosts to be deployed in one cluster and more VMs to be stored in one logical volume.

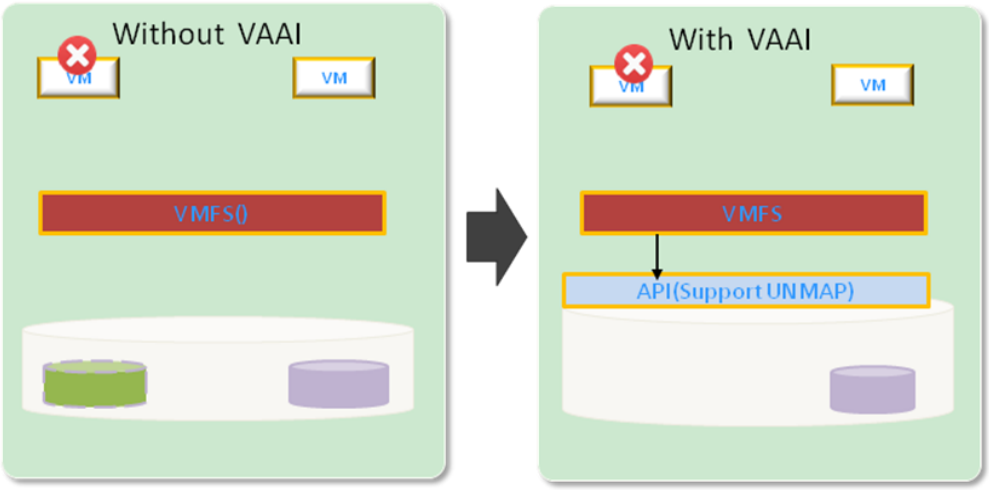
Figure 2-17 ATS mechanism



4. UNMAP/Reclaim command

The VMFS is connected to block storage in its downlink. When a VM is deleted or migrated, the available space displayed on the VMFS increases, but actually the space is not reclaimed on block storage. After the storage layer provides a reclaim API for the VMFS, the ESXi server can call this API to inform the lower-layer block storage of reclaiming space in a timely manner. Therefore, the space on the disk array can be rapidly reclaimed after a VM is migrated or deleted from a datastore.

Figure 2-18 UNMAP mechanism



2.4.5 Compatibility

2.4.5.1 Hardware Platform

Huawei FusionStorage Server SAN solution supports both Huawei-developed servers and servers that are provided by other vendors but have been certified by Huawei. Requirements for servers are as follows:

FusionStorage must be installed on general-purpose x86 servers.

At least three servers must be configured.

The hard disks in one resource pool must be of the same type, the same number, and the same size. If the disk sizes in a resource pool are different, the resource pool capacity is determined by the smaller disks. The gap between hard disk numbers on different servers cannot be greater than 2, and the proportion of the gap to the maximum number of hard disks on a server cannot be greater than 33%.

If servers use SAS or SATA disks as the main storage, the PCIe SSD cards or SSDs must be used as cache devices.

Each server must have sufficient memory reserved for running FusionStorage. The reserved memory is calculated as follows: 3 GB x Number of VBS processes + 3.5 GB x Number of OSD processes + 4 GB x Number of MDC processes. The space required by VBS, OSD, and MDC processes is small and can be omitted.

When deployed on storage nodes, FusionStorage Manager nodes can use only local hard disks on servers. The memory and storage resources to be reserved for FusionStorage Manager nodes are determined by the number of hosts in the system:

If the system contains 3 to 64 hosts, 16 GB of memory, 100 GB of hard disks, and 80 IOPS are required.

If the system contains 65 to 4096 hosts, 32 GB of memory, 400 GB of hard disks, and 200 IOPS are required.

The local hard disk on the server running the ZK process must have greater than or equal to 55 GB of space. It is best practice to deploy the ZK process on an independent hard disk. In this case, reserve 5 GB of memory for the ZK process.

Each server requires 4 Gbit/s bandwidth for FusionStorage communication. 10GE networks are recommended. If FusionStorage is deployed in a cloud resource pool and contains more than 16 nodes, the service plane and the storage plane must use dedicated network interface cards (NICs).

2.4.5.2 Virtualization Software

Huawei FusionStorage solution supports mainstream virtualization platforms, including Huawei FusionSphere, VMware vSphere, and KVM.

2.4.5.3 OS

In addition to providing storage services for virtualization platforms, Huawei FusionStorage provides storage services for the physical servers with VBS nodes deployed in the OSs. The VBS nodes can be deployed in mainstream Linux OSs.

2.4.5.4 Compatibility List

For details about FusionStorage compatibility, see the *Compatibility List for FusionStorage V100R003C30*.

2.5 Big Data Service

2.5.1 Feature Overview

The big data platform of a customer needs to provide data big service for each of its departments and each ISV's tenants. The number of tenants may be large and each tenant requires a different resource quota and a different data permission when applying for big data service. Therefore, manual management for all those may be complicated and inefficient. Huawei uses an automated big data service provisioning process and closed-loop management functions to improve data provisioning efficiencies and customer satisfaction.

2.5.2 Application Scenarios

Big data analysis and processing applications such as extract, transform, and load (ETL), bill querying, and data mining are typical application scenarios in a distributed cloud data center. Such users in those application scenarios are generally service units of the enterprise and tenants of ISVs. For security propose, computing and data resources of different application scenarios must be isolated from each other. In addition, to ensure an effective use of enterprise cloud data center resources, quotas are assigned to users of big data service to control resource usage and enable real-time O&M monitoring.

2.5.3 Key Roles

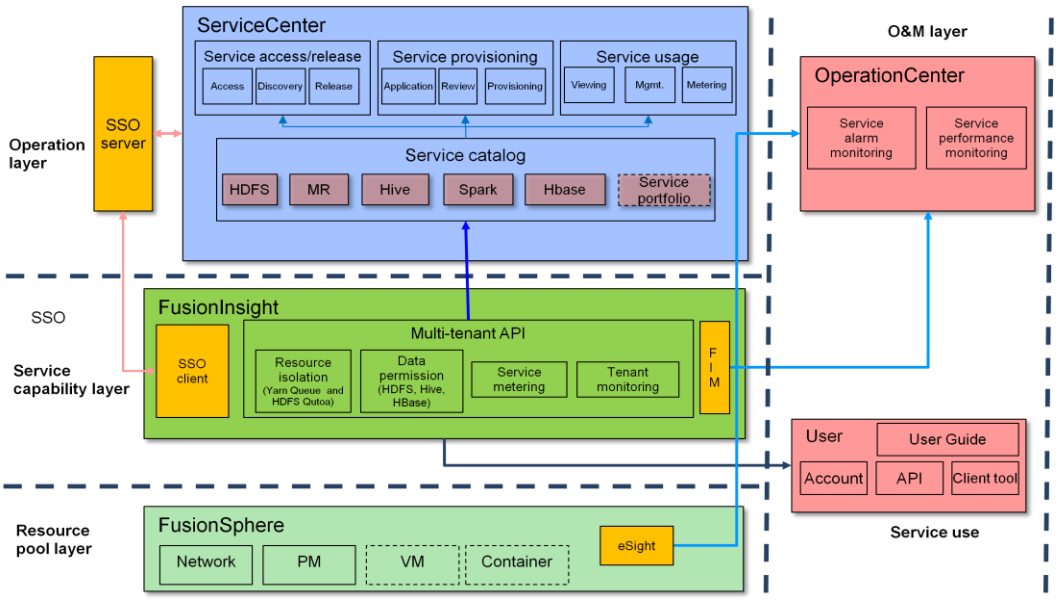
The following table lists the roles and their capabilities in big data service.

Role	Capability
Domain Service Manager	Accesses a big data cluster. Manages VDCs, including creating VDCs and allocating VDCs to the VDC Service Manager. Manages domain application templates. Manages domain service catalogs. Manages software packages, scripts, and VM templates in the domain. Reviews VDC service applications. Queries resources provisioned in the domain. Monitors domain capacity. 9. Manages domain users.

Role	Capability
VDC Service Manager	Applies for, changes, extends, and releases a VDC. Queries resources that have been applied for in the VDC. Queries all application orders in the VDC. Manages end users in the VDC. Reviews service end users' applications. Creates resources in the VDC and allocates resources to end users. Manages VDC application templates, including orchestrating and commissioning application templates and publishing the templates to VDC service catalogs. Manages VDC service catalogs. Manages the VDC software library. Monitors VDC capacity.
VDC Service User	1. Applies for, changes, extends, and releases services. 2. Queries and maintains resources that have been applied for. 3. Queries application orders.

2.5.4 Key Features

The following figure illustrates the big data service provisioning solution architecture.



The Huawei big data service provisioning has the following characteristics:

1. Single sign-on (SSO)

On ServiceCenter, when users are connecting to, releasing, provisioning, and using big data service, if the FusionInsight component is required, ServiceCenter automatically calls the FusionInsight interface to implement related functions. The calling process is transparent to users and users do not need to log in to the FusionInsight GUI to perform any operations.

2. Automatic provisioning

   Huawei offers an automatic management capacity for big data service provisioning. After managers approve the applications for big data service, ServiceCenter automatically completes the big data service (and other associated services) creation, installation, and configuration procedures according to user-defined settings.

3. On-demand resource quota application

   When users are applying for big data service, they can have desired settings for the total number vCPUs, the exclusive number of vCPUs, and storage capacity size. The quota of CPUs and that of memory must be at the same percentage in a FusionInsight cluster. When users apply for big data service on ServiceCenter, they only need to specify the number of vCPUs, and the memory quota will be automatically calculated accordingly.

4. Self-help O&M

   The Huawei cloud data center O&M software OperationCenter provides comprehensive big data O&M monitoring capabilities. The P layer is monitored by FusionInsight and the I layer is monitored by eSight. Users can learn the big data service status in OperationCenter whenever needed. After users' applications for big data service are approved, ServiceCenter provides the portal address, client tool, user guide, and other information for users to use big data service with ease.

5. Security isolation

   ServiceCenter supports the multi-tenant management. Computing and data resources can be isolated between tenants to ensure the data of each user.

6. Accurate metering

   ServiceCenter can provide detailed metering information on the use of FusionInsight resources from aspects of CPU, memory, and storage space. The metering statistics can be used to support capacity planning and charging planning for big data clusters.

2.6 SDN

2.6.1 Application Scenarios

SDN is new network architecture that provides programmable networks. The SDN architecture provides maximum network control flexibility. With the development of the mobile Internet and big data technologies, more and more IT services are migrated to data centers. Network as a service (NaaS) is a basic IT service in the cloud data center. Tenants can flexibly apply for virtual network resources to meet IT service requirements.

The SD-DC^2 solution, SDN applies to the following scenarios:

- Network Automation

  Northbound APIs are provided for upper-layer management software, so that upper-layer management software can invoke APIs to implement network automation and provide real-time network services. This ensures quick service rollout.

  Example 1: A large media asset enterprise has over 40 departments and subsidiaries. The headquarters are required to centrally manage all basic IT resources. Departments and

subsidiaries apply for or release the IT resources (including computing, network, and storage resources) based on service requirements.

Example 2: A development and test environment has ever-changing requirements on network resources, which requires IT systems to implement automatic network resource provisioning.

Example 3: A government and enterprise cloud requires rapid network resource provisioning and on-demand access control list (ACL) configuration based on subsidiaries' requirements (for accessing external networks). The required configuration duration is even shorter than one day.

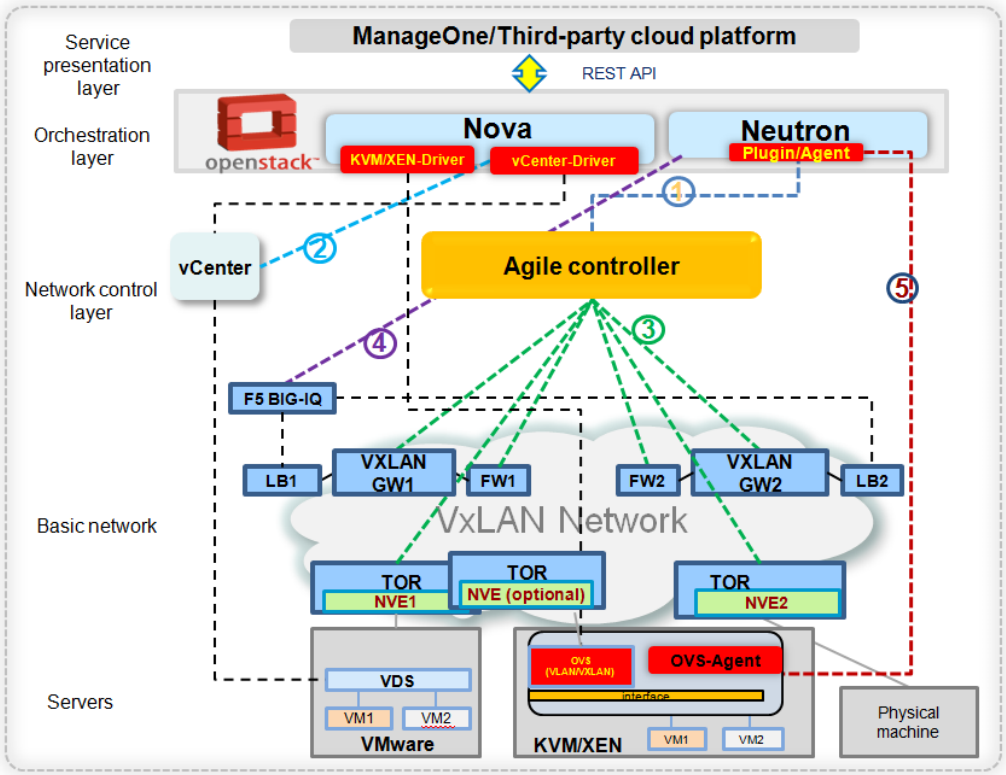
- Elastic Resource Allocation

  According to resource allocation flexibility and elasticity requirements, computing resources requires dynamic network configuration and interconnectivity between Layer 2 and Layer 3. Including Layer 2 interconnectivity between single data center and single zone (POD zone), Layer 2 interconnectivity among single data center across zones, and automatic Layer 3 interconnectivity between data centers.

  If a data center does not have network automation and resource elasticity requirements, the SDN solution is unnecessary.

2.6.2 Deployment Architecture

The SD-DC^2 network subsystems adopt the network design of the SDN architecture, as shown in the following figure:



1. Service presentation and O&M layer (ManageOne)

   Provides user-oriented service interfaces.

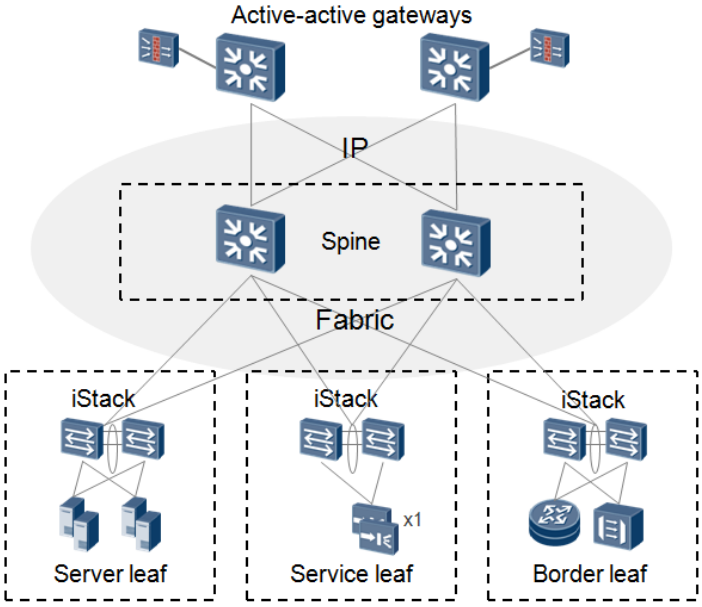
2. Service orchestration layer (FusionSphere OpenStack)

Orchestrates storage, computing, and network resources, adopts the standard open OpenStack architecture, and supports multiple vendors.

3. Network control layer (Agile Controller)
   - Implements service policy orchestration, network modeling, and network instantiation.
   - Supports northbound open APIs and connects to a cloud platform or applications, thereby achieving quick service customization and automatic provisioning.
   - Supports southbound OpenFlow/NETCONF/BGP/OVSDB interfaces and centrally manages physical and virtual networks.
4. Basic network layer

   Physical networks: Fabrics based on VXLAN or VLAN
5. Virtual network layer
   - The overlay solution manages virtual networks in a unified manner.
   - This layer houses software virtual network components such as vSwitches, vFWs, vRouters, and vLBs.
6. Key design points
   - This solution employs an open OpenStack+SDN architecture that allows upper-layer services to focus on service innovation.
   - In a homogeneous resource pool (KVM/VRM), Open vSwitches (OVS) or Top of Rack (ToR) switches are used for Network Virtualization Edges (NVEs) encapsulation. Both all-hardware overlay solution (NVEs are all on TOR switches) and hybrid overlay solution (homogeneous resource NVEs are all on OVS) supported.
   - This solution uses Top of Rack (ToR) switches as NVEs to connect third-party resource pools (VMware resource pools and bare metal resource pools).
   - This solution allows Neutron plug-ins of third-party devices to directly manage devices. (Neutron plug-ins of F5 directly manage F5).
7. Physical networking in a zone

   (1) All-active gateways (recommended)



- Both active-passive and all-active gateway solutions are acceptable. However, the active-passive solution leads to low device utilization. The all-active gateway solution is recommended.
- Configure the same gateway address and tunnel endpoint address for the multiple gateway devices. VMs then do not sense the actual location of each gateway device. On an Underlay network, symmetric link routes share loads, enhancing the gateway reliability.
- All-active gateways in a group share loads using IP ECMP to improve the forwarding capability.

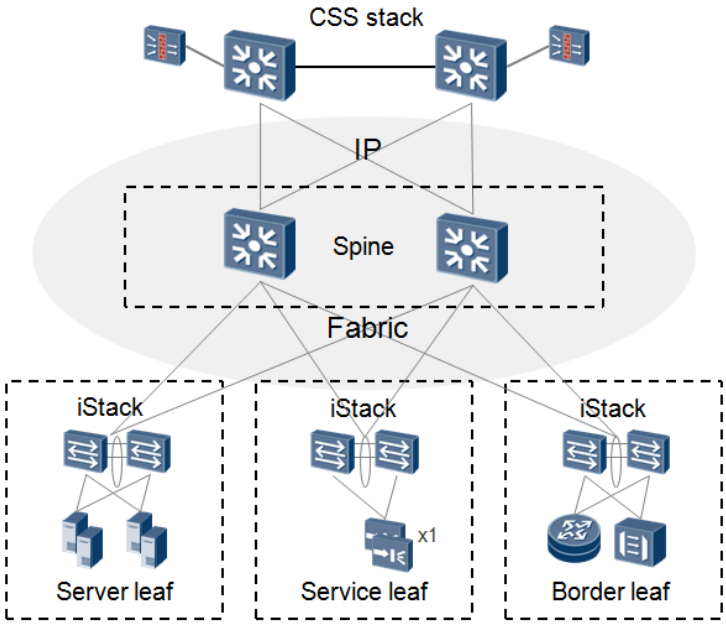
All-active gateway solution design

- A group of BDIF IP address/MAC address and virtual tunnel end point (VTEP) IP address is configured for the multiple gateway devices. Cross-subnet traffic that VMs need to send to gateway devices encapsulates the destination VTEP IP address on the VXLAN L2 gateway. The traffic is then forwarded to one member gateway device through IP ECMP load sharing links.
- DFS groups and peer links are configured between all-active gateway devices to synchronize ARP and MAC forwarding information tables to ensure traffic consistency.
- Active-active gateway devices are supported (firewalls and load balancers are connected in bypass or service leaf mode).
- Active-active gateway devices are supported (firewalls and load balancers are connected in service leaf mode).

Characteristics of all-active gateway solution

- No enhancements on the specifications of VRF/Subnet/RIB/FIB/ARP/MAC resources on gateway devices
- Applicable to VPCs that require high reliability
- High reliability and scalability
- Support for the active-active gateway solution
- Connecting load balancers to upstream TOR switches or connected to aggregation switches in bypass mode (load balancer automation is not for commercial use and is selected by project requirements)

(2) Stacked gateways



Requirements

- CSS stacks have already been deployed in the data center.
- CSS stacks are maintained to support gateway devices.

Solution design

- A group of BDIF IP address/MAC address and VTEP IP address is configured for the stack gateway device. Cross-subnet traffic that VMs need to send to gateways encapsulates the destination VTEP IP address on the VXLAN L2 gateway. The traffic is then forwarded to the CSS through IP ECMP load sharing links.
- Firewalls and load balancers are connected to two gateways in dual-homing mode and connected to the gateways in bypass mode.

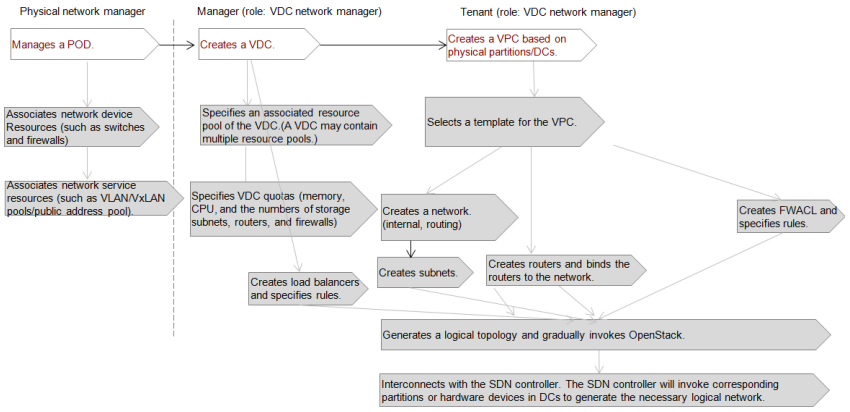
Application scenarios

- VLAN evolves to VXLAN.
- As an independent logical device, CSS is easy to configure and manage, with clear O&M interfaces.
- Gateway stacking delivers higher reliability than standalone gateway deployment.
- Generally a stack contains two gateway devices.

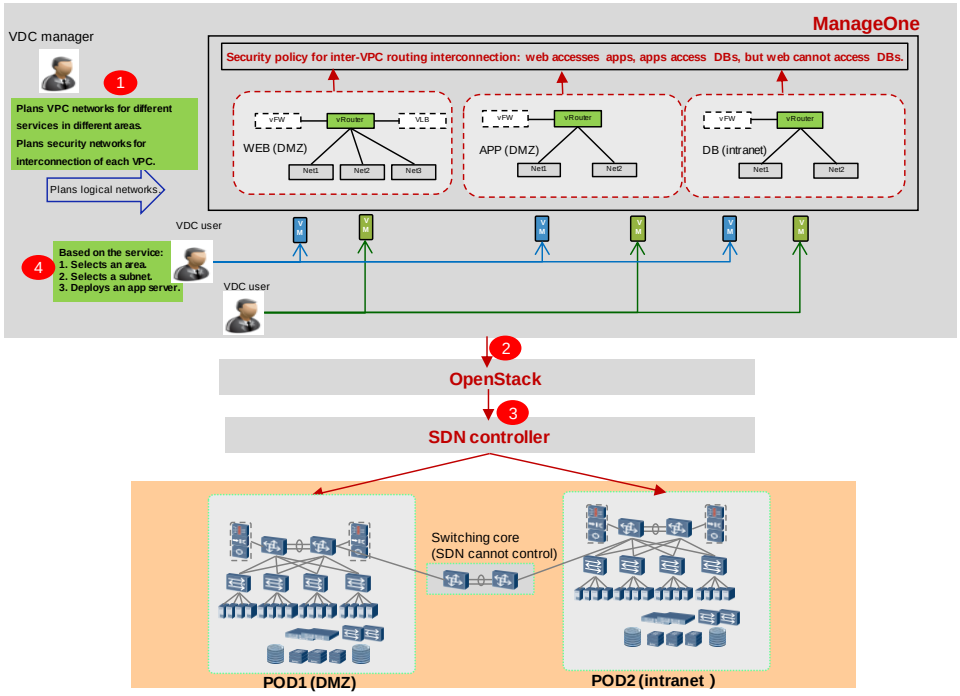
  Connect load balancers to upstream TOR switches or connected to aggregation switches in bypass mode. (Load balancer automation is not for commercial use and is selected by project requirements.)

8. Service provisioning process design

   (1) VDC/VPC provisioning



(2) VDC/VPC provisioning



a. VDC managers plan service networks. For example, in area DMZ1, a VDC manager plans a VPC to provide network resources for web services; in the intranet area, the VDC manager plans two VPCs to provide network resources for application and database services. Each VPC has its own firewall, LB, router, and network segments. Then, the VDC manager configures interconnectivity among and security policies for the three VPCs.

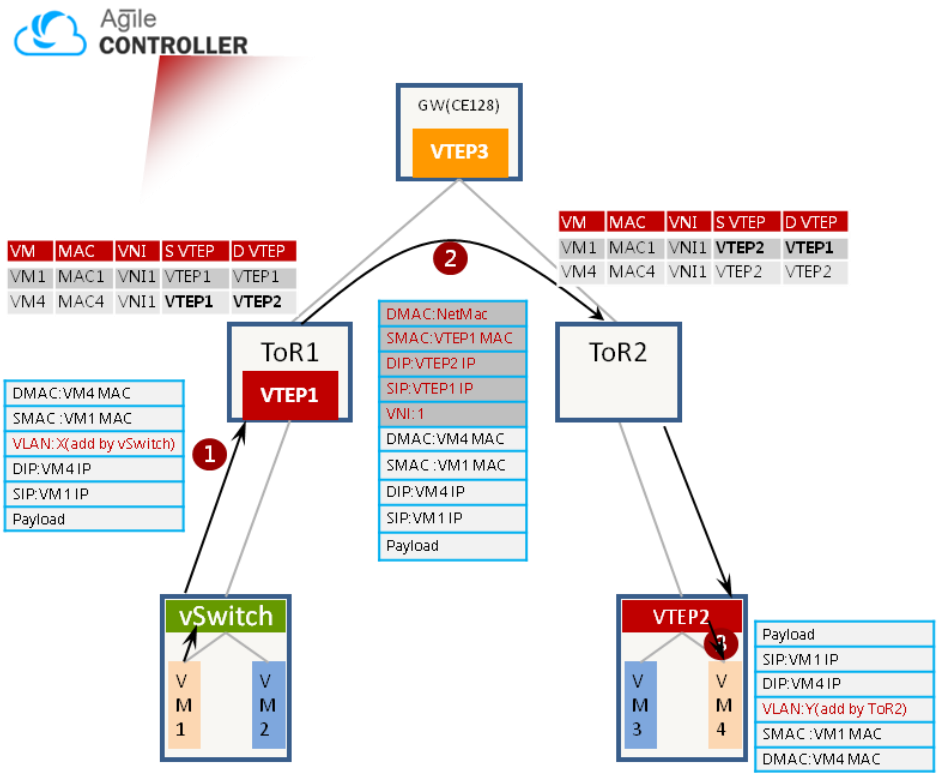
b. ManageOne uses standard RESTful APIs to invoke OpenStack.

c. SDN controllers are connected to OpenStack as plug-ins, and separately invoke network devices in their own areas. Based on devices' northbound interfaces, networks required by services are generated automatically. Interconnectivity among different service areas is automatically enabled. Users do not need configure any network devices. VPCs across PODs use three-segment VXLAN for interconnectivity, which is transparent to core switching devices.

d. Applications select specific VPCs to create VMs and deploy related services.

9. IP packet forwarding process

(1) Intra-VPC L2 forwarding



Prerequisites

- The dynamic routing protocol has been enabled for underlay networks. It is a good practice to configure IP route reachability between OSPF and VTEP1, 2, and 3.
- AC has completed network service provisioning. vRouters and networks have been created on the gateway side. Networks have been created on ToR switches. vSwitches have completed VLAN pushing.
- VM1 and VM4 in BD1 correspond to VNI1, and VM2 and VM3 in BD2 correspond to VNI2. Ingress replication tables are sent to ToR1 and ToR2, and VM1 to VM4 are connected to the VXLAN.
- ARP packets are sent between VM1 and VM4 so that the VMs learn each other's ARP table entries. At the same time, ToR1 and ToR learn MAC address table entries of VM1 and VM4 and associate the learned information with the underlay tunnel (VTEP1 <-> VTEP2).

Conversion Process

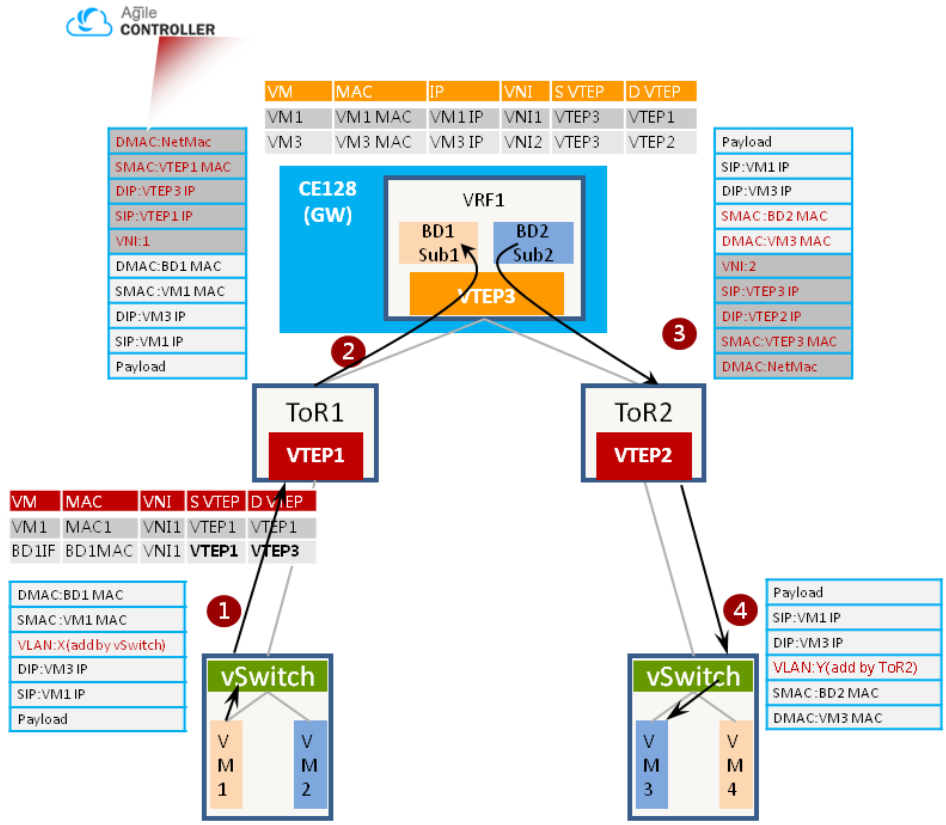
- VM1 sends data packets to VM4, which is in the same network segment. The source MAC address/source IP address (SMAC/SIP) is the MAC or IP address of VM1, the VLAN is *X* (local VLAN under ToR1 added by a vSwitch), the destination MAC address (DMAC) is the MAC address of VM4, and the destination IP address (DIP) is the IP address of VM4.
- The packets are sent to ToR1. Based on the access port and VLAN, VNI1 is queried. VNI1+VM4 DMAC are used to query MAC table entries in the MAC table. The egress is found, which is the VTEP1 <-> VTEP2 tunnel.
- The VTEP1 end point encapsulates VXLAN packets based on tunnel information. The outer MAC information is the MAC address of the next hop (NetMac). VXLAN packets are sent to peer-end end point VTEP2 in the IP Fabric hop by hop.

- The VTEP2 end point decapsulates VXLAN packets. Based on inner DMAC:VM4 of VNI1, the egress is queried in the MAC address table. Then VLAN Y is added and then forwarded to VM4.
- The vSwitch deletes VLAN:Y and then sends packets to VM4.

Description

L2 traffic is not dependent on AC or gateway devices for ARP learning. The forwarding plane learns the MAC address and correlates to the underlay tunnel.

(2) Intra-VPC L3 forwarding



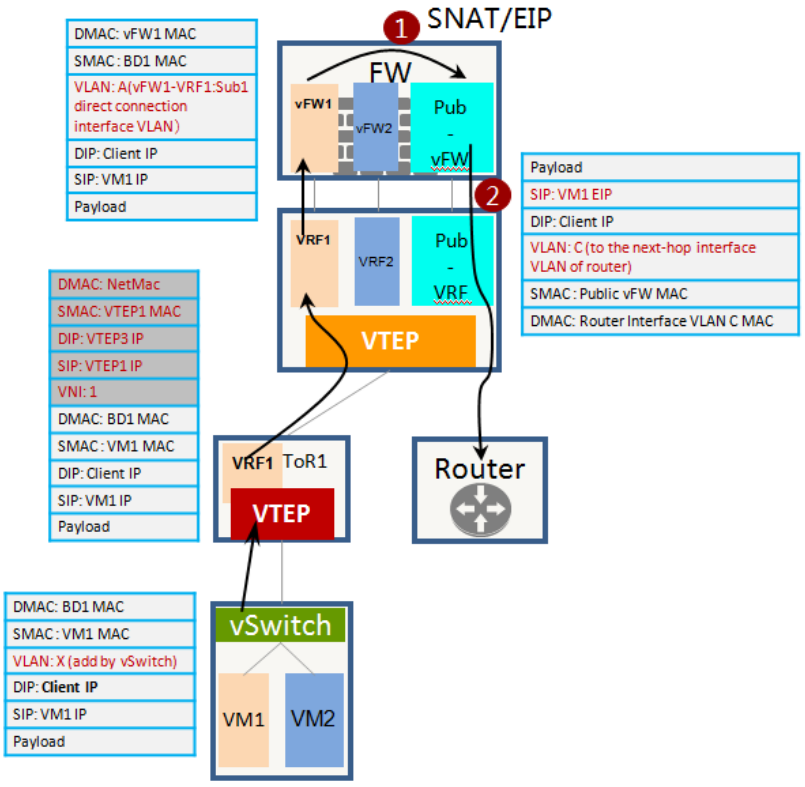
Prerequisites

- The dynamic routing protocol has been enabled for underlay networks. It is a good practice to configure IP route reacheability between OSPF and VTEP1, 2, and 3.
- AC has completed network service provisioning. vRouters, networks, and subnets have been created on the gateway side. Networks have been created on ToR switches. vSwitches have completed VLAN pushing.
- VM1 and VM4 in BD1 correspond to VNI1, and VM2 and VM3 in BD2 correspond to VNI2. Ingress replication tables are sent to ToR1 and ToR2, and VM1 to VM4 are connected to the VXLAN.
- When VM1 accesses VM3, because the VMs reside in different network segments, ARP requests of gateway addresses are sent. AC (above the control plane) and gateways (below the control plane) learn the ARP information and send it to the forwarding plane to guide the data forwarding. As shown in the preceding figure, L2/L3 table entries have been created.

  Conversion Process

- VM1 sends data packets to VM3 that resides in a different network segment. VLAN is X (local VLAN under ToR1 added by a vSwitch).
- The packets are sent to ToR1. Based on the access port and VLAN, VNI1 is queried. VNI1+BD1 DMAC are used to query MAC table entries in the MAC table. The egress is found, which is the VTEP1 -> VTEP3 tunnel.
- The VTEP1 end point encapsulates VXLAN packets based on tunnel information.
- The gateway device decapsulates packets received in step 2. The inner packet DMAC is the BD1 MAC address. Therefore, the IP address of VM3 is used for L3 forwarding. After a matching with learned ARP table entries, the egress is found, which is the VTEP3 -> VTEP2 tunnel.
- The VTEP3 end point encapsulates VXLAN packets based on tunnel information. VNI is 2, and the inner packet SMAC is BD2 MAC.
- The VTEP2 end point decapsulates VXLAN packets. Based on VNI 2+VM3 MAC, the egress is queried in the MAC address table. Then VLAN Y is added and then forwarded to VM3.
- The vSwitches delete VLAN:Y and then send packets to VM3.

(3) Intra-VPC L3 to L7 forwarding



Prerequisites

- VM1 in VRF1 has the east-west access capability, and relevant MAC and ARP entries are ready.
- AC sends default routes to VRF1 of the gateway device, with the next hop to vFW1 interface.
- AC sends the NAT and forwarding policies to firewalls.

Conversion Process

- After being forwarded to vFW1, a route table matching is performed and packets are forwarded to the public-vFW tunnel.
- After traffic is sent through the public-vFW tunnel, each vFW uses an independent VLAN interface to send packets to routers.

Description

Public-VRF of the gateway device forwards traffic of the router on an L2 network, thereby reducing the needed number of table entries of gateway device routes.

2.6.3 Compatibility List

The following table describes the version compatibility and mapping.

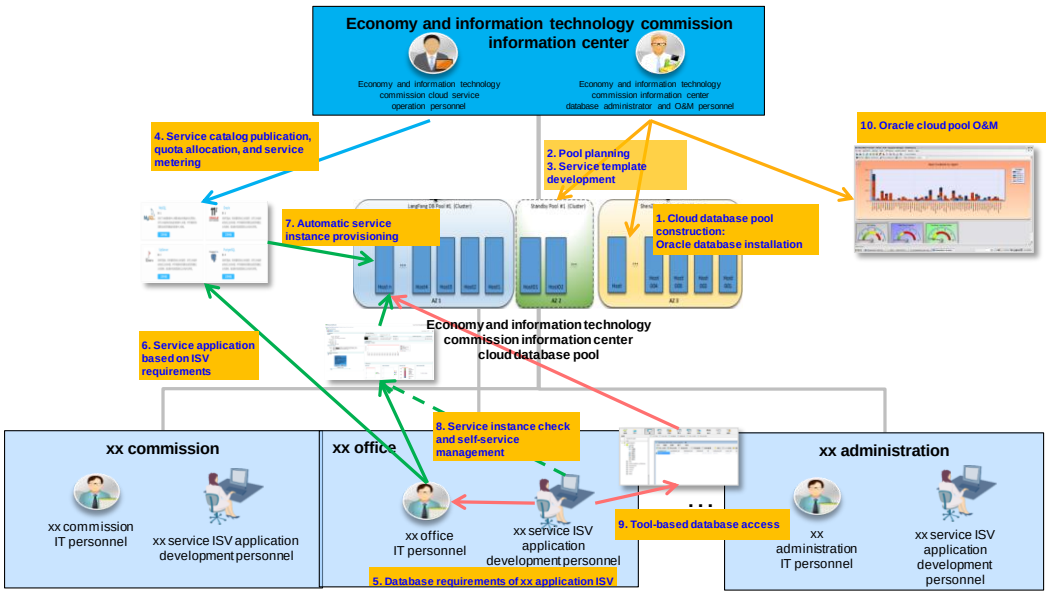
Item	Component Model	Component Version
Operations/O&M	ManageOne	ManageOne V1R3C00
Cloud platform	FusionSphere	FusionSphere V1R6C00
Intra-DC network controller	AC 2.0	V2R1C00
VXLAN/VLAN gateway	CE12800/CE12800S CE6851HI CE6850HI CE6855 CE7850 CE8860	V100R6C00
ToR (NVE)	CE6851HI CE6850HI CE6855 CE7850 CE8860	V100R6C00
Firewall	NGFW (hardware firewall) series (1) Enterprise market: USG9520/USG9560/USG9580 (2) Carrier market: Eudemon8000E-X3/8/16	NGFW V5R1C00SPC
LB (integrating F5)	F5 BIG IQ: F5-BIG-LTM-2000S-AC F5-BIG-LTM-2000S-DC F5-BIG-LTM-2200S-AC F5-BIG-LTM-2200S-DC F5-BIG-LTM-4000S-AC F5-BIG-LTM-4000S-DC F5-BIG-LTM-4200S-AC	Neutron plug-in version: 1.0.10 Not for commercial use; driven by projects

Item	Component Model	Component Version
	F5-BIG-LTM-4200S-DC	

2.7 RDS

2.7.1 Scenario Description

The RDS is applied to e-government scenarios, as shown in the following figure:



- Constructor: economy and information technology commission information center

  As the resource and service provider, the economy and information technology commission information center performs general planning and O&M for pools. Detailed operations are as follows:

  - Build RDS pools, including planning for and deploying Oracle PM pools.
  - Develop RDSs, including defining service templates and planning for database types and pools.
  - Provide service metering and SLA assurance.
  - Perform O&M, monitoring, and management for database pools.
- Users: managers of commissions, offices, and administrations

  As service users and operation managers, the commissions, offices, and administrations perform the following operations:

  - Collect database requirements from independent software vendors (ISVs) and allocate related quotas to the organizations or virtual data centers (VDCs).
  - Manage RDS service publication and operation within the organization and approve service applications of ISV development personnel.
- Users: ISV development personnel

As end users of the services, ISV development personnel perform the following operations:

- Apply for the RDS on the RDS platform in a self-service manner.
- Manage the lifecycle of RDS instances in a self-service manner, for example, modifying the specifications or releasing the RDS instances.
- Manage RDS instances in a self-service manner, for example, creating databases, creating users, analyzing logs, monitoring performance, and performing backup or recovery.
- Develop application and database access instances.

2.7.2 Cloud Database Pool Planning

Planning, deployment, and configuration must be performed for cloud database pools. During planning, the database type, database version, and pool size are specified. During deployment and configuration, the Oracle database and OEM Cloud Control are deployed, Oracle hosts are managed, and zones, pools, and database templates are created.

This section describes the planning and basic configurations for database pools.

2.7.2.1 Resource Pool Planning

Oracle pool planning contains the following content:

- Zone planning: Create RDS zones based on the data center distribution plan. For example, if the Beijing and Shanghai data centers provide the RDS as planned, you can create two zones: Zone_Beijing and Zone_Shanghai.
- Oracle version and service type planning: For example, apply Oracle 12.1.0.2.0 and Oracle 11.2.0.4.0, and provide the service types of DBaaS and pluggable database as a service (PDBaaS). PDBaaS is specific to Oracle 12c.
- Pool planning:
  - Create Oracle pools based on the Oracle version and service types. For example, Pool 1 is configured with DBaaS and Oracle 11.2.0.4.0 as planned, and Pool 2 is configured with PDBaaS and Oracle 12.1.0.2.0 as planned.
  - To maximize the resource utilization, it is recommended that you apply the same configurations to PMs in the same pool.
- Service template planning: Create one or more service templates for each pool. Different templates specify different database instance specifications. DBaaS templates specify user-defined specifications such as system global area (SGA) size, program global area (PGA) size, and CPU quantity. PDBaaS templates specify user-defined specifications such as CPU quantity, memory size, and storage size.

The following table describes the specifications specified in service templates.

Server Specifications	Description
Low-end host	CPU quantity: 0.5, memory size: 2.5 GB, SGA size: 2 GB, PGA size: 0.5 GB
Low-end host	CPU quantity: 1, memory size: 4 GB, SGA size: 3 GB, PGA size: 1 GB
Mid-range host	CPU quantity: 2, memory size: 6 GB, SGA size: 4 GB, PGA size: 2 GB

Server Specifications	Description
High-end host	CPU quantity: 4, memory size: 12 GB, SGA size: 8 GB, PGA size: 4 GB
Maximum scale	CPU quantity: 8, memory size: 24 GB, SGA size: 16 GB, PGA size: 8 GB
Specification definition	CPU quantity: > 8, memory size: > 24 GB, SGA size: > 16 GB, PGA size: > 8 GB

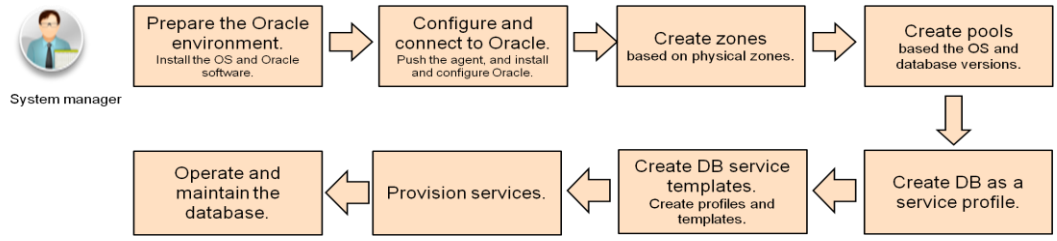
The above table provides reference for template planning. In specific projects, it is recommended that you determine the specifications in the templates based on actual customer services.

You can apply small specifications to databases in the testing environment, to create more instances.

You can apply high specifications to databases in the production environment. For services with high input/output (I/O) requirements, configure a large memory size and storage size to ensure better service performance.

2.7.2.2 Closing the Process

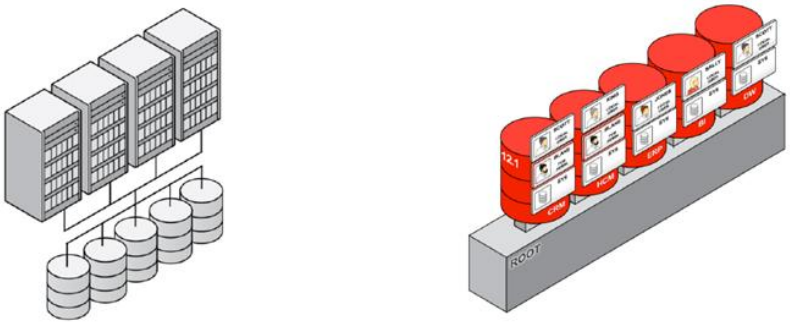
Basic configurations related to zones, pools, and templates are performed in the OEM system as planned. The following figure shows the key process.



1. Manually deploy the Oracle single-instance environment or Oracle RAC environment as planned, including installing the operating system and configuring Oracle software.
2. Identify a deployed Oracle host by auto scanning in the OEM system and push an agent to the host, to manage the Oracle host in the OEM system.
3. Create a zone and incorporate all managed Oracle hosts in the zone.
4. Create a pool and specify the maximum number of instances that can be created by each host in the pool.
5. Create data profiles (required in DBaaS templates) and specify an existing database instance as the reference.
6. Create a template, specify the zone and pool where the template belongs, and specify the storage location, manager account, and specifications of the database instances created based on the template.
7. Publish the database template as a service. The service operating system cooperates in this step.
8. Manage O&M of database services as the system manager, such as troubleshooting and performance monitoring.

2.7.3 Database Type

At present, the RDS for Oracle provides two types of services: DBaaS and PDBaaS. The following figure compares the two service types.



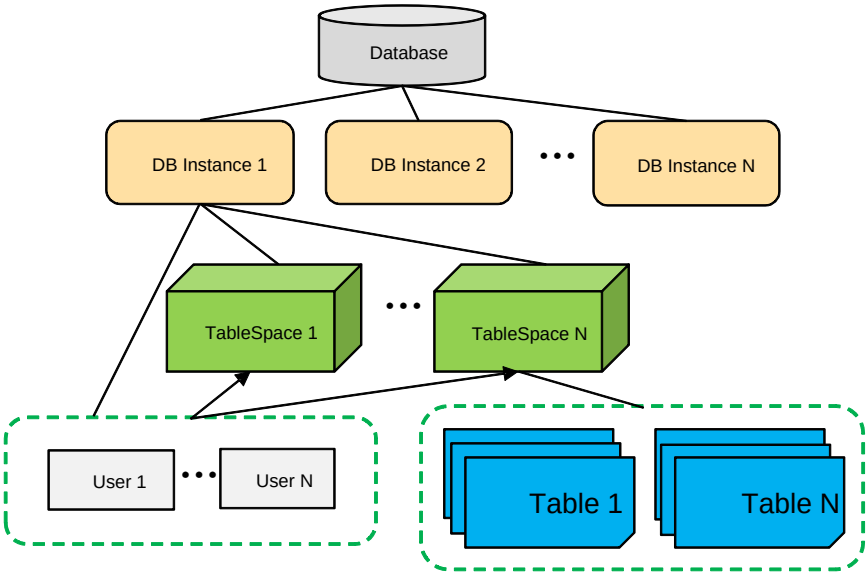
As shown in the above figure, the two service types have the following differences:

- DBaaS: Tenants share the PMs and operating system. Tenants are isolated from each other by using database instances.
- PDBaaS: Tenants share the PMs, operating system, and container database (CDB). Tenants are isolated from each other by using pluggable databases (PDBs).

PDBaaS is a new service type developed for the multi-tenant database mode in cloud environment based on Oracle 12c. Compared with DBaaS, PDBaaS provides higher consolidation density, lower total cost of operation (TCO), and simpler O&M and provisioning.

2.7.3.1 DBaaS

DBaaS contains multiple database instances in a database. The following figure shows the concepts related to Oracle database instances and relationships between the concepts.



- Database

  An Oracle database is physical storage space of data, containing ORA or DBF data files, control files, online logs, and parameter files. One operating system is configured only with one Oracle database.

- Database instance

  One Oracle database instance consists of a series of background processes and memory structures. One database may contain multiple instances. All database operations are performed based on the instances.

- User

  Users are created based on database instances. Users in different instances may have the same user name. Multiple users can be created in one instance.

- Table space

  The table space is a logical concept for managing data storage. The table space is related only to data files, such as ORA or DBF files. Data files are physical concepts. One table space may contain multiple data files, but one data file can be stored only in one table space.

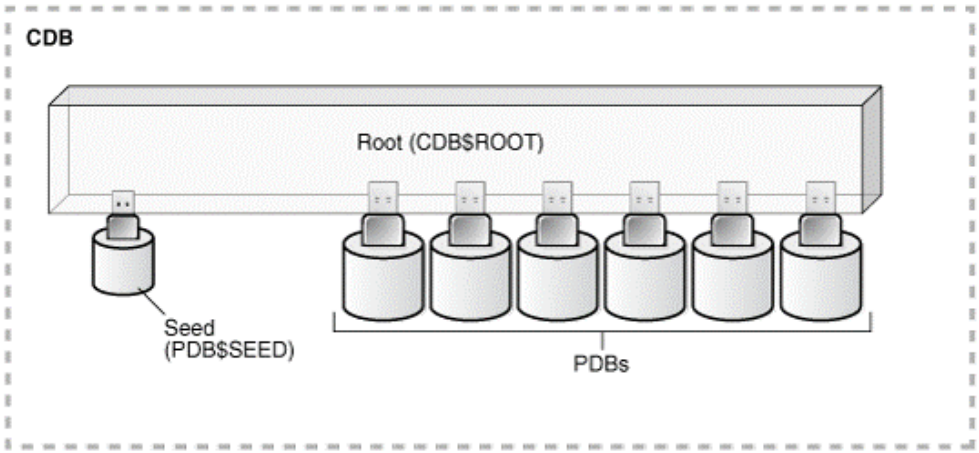
- Data file

Data files are physical storage units in a database. Data in a database is stored in table spaces and is stored in one or more data files. One table space contains one or more data files, but one data file can be stored only in one table space.

2.7.3.2 PDBaaS

Oracle 12c incorporates new features of the CDB and PDB. In the multi-tenant environment incorporated in Oracle 12c, one CDB may bear multiple PDBs. In versions earlier than Oracle 12c, instances and databases are in a one-to-one or many-to-one relationship. That is, one instance is associated only with one database, and one database can be loaded by multiple instances. Instances and databases cannot be in a one-to-many relationship. However, in Oracle 12c, a one-to-many relationship is enabled between instances and databases.

The following figure shows the relationship between one CDB and multiple PDBs.



As shown in the above figure, a CDB consists of the following components:

- Root

The root component is also referred to as CDB$ROOT. It stores the metadata and common users provided by Oracle. One example of the metadata is source code of a procedural language (PL)/structured query language (SQL) packet provided by Oracle. Common users refer to users in each container.

- Seed

  The seed component is also referred to as PDB$SEED. It is a PDB template. You cannot add or delete objects in the seed component. One CDB contains only one seed.

- PDB

  One CDB contains one or more PDBs. The PDBs provide backward compatibility. PDB operations are similar to operations in earlier-version databases.

2.7.4 Database Instances

2.7.4.1 Database Instance Provisioning

The following parameters are required for creating a common database service instance:

- Zone: indicating the zone that requires service instance provisioning
- DBaaS template: indicating the template required for creating the service instance and defining the SAG, PGA, and CPU specifications
- Service instance name
- Database ID and password: used for accessing the database after the service instance provisioning is successful
- Service instance description

After the service instance provisioning is successful, the service instance is enabled. You can check detailed information about the service instance, including:

- IP address (and also a scan IP address for an RAC instance)
- **SID** and **ServiceName**
- Connection string, used for connecting to the database based on related tools, such as PL/SQL

The provisioning of a common database service instance requires 40 minutes.

2.7.4.2 PDB Instance Provisioning

The following parameters are required for creating a PDB service instance:

- Zone: indicating the zone that requires service instance provisioning
- PDBaaS template: indicating the template required for creating the service instance and defining the CPU, memory, and storage specifications
- Service instance name
- Database ID and password: used for accessing the database after the service instance provisioning is successful
- Service instance description

After the service instance provisioning is successful, the service instance is enabled. You can check detailed information about the service instance, including:

- IP address (and also a scan IP address for an RAC instance)
- **SID**, **ServiceName**, and **PDBName**

- Connection string, used for connecting to the database based on related tools, such as PL/SQL

The provisioning of a PDB service instance requires 10 minutes.

2.7.5 Self-Service Management on Database Instances

The following table describes self-service management operations applicable to database instances.

Service Type	Self-Service Operation	Description
DBaaS	Application, enabling, disabling, deletion, and self-service monitoring	Self-service monitoring in the ManageOne is available only for the VDC manager.
PDBaaS	Application, enabling, disabling, deletion, and self-service monitoring	Self-service monitoring in the ManageOne is available only for the VDC manager.

2.7.6 Database Resource Quota and Metering

The RDS for Oracle enables quota management on database service instances. The following table describes the quota items.

Quota Item	Unit
Memory capacity	GB
Storage capacity	GB
Number of database instances	Piece
Number of PDB instances	Piece

The RDS for Oracle enables metering management on database service instances. The following table describes the metering items.

Service Instance Type	Metering Item	Description
Database instance	SGA memory size	No support for CPU metering, storage metering, or precise metering upon startup and shutdown
PDB instance	Memory size and storage size	No support for CPU metering or precise metering upon startup and shutdown because no quota is specified for the CPU

2.8 Security Management

2.8.1 Application Scenarios

With the IT development, Web 2.0, service oriented architecture (SOA), and cloud computing technologies are emerging. Mobile devices, remote access devices, browsers, plug-ins of various applications, intelligent terminals, and cloud hosts come into being. Information security faces new challenges. Attacks from the intranet and extranet and system vulnerabilities are major threats to information security. The most valuable information assets are frequently attacked. As the core of information, data centers bear the brunt.

Based on cloud computing and distributed deployment of data centers, data center elements embrace some changes, such as virtualization and boundary extension. Therefore, a systematic distributed cloud data center security solution should cover all elements, and security elements should support logical isolation. Security of all elements cannot be ensured by only traditional technologies and physical boundaries.

The security subsystem of the SD-DC^2 is designed based on the best practice in the industry and Huawei's expertise and experience. Objectives of the security subsystem architecture are as follows:

- Modularization

  The security subsystem is designed based on eight modules: physical security, network security, host security, application security, virtualization security, user security, security management, and security services. Security architecture can be quickly formed based on customer requirements to provide a customized security system.

- End-to-end security

  The security subsystem provides end-to-end protection from user access, use, and exit. Technologies such as two-factor authentication, rights control technology for privileged users, VPN, application protection technology, and event auditing technology are used to control user access to IT resources, ensure data communication security and secure application access, and audit operations.

- Low coupling

  The security subsystem must provide protection for multiple layers, such as the data layer, network layer, and application layer. Therefore, the security subsystem involves various security technologies, products, and security management policies. The security subsystem features low coupling. That is, various security technologies are not tightly associated, security products provided by different vendors can be used and are not limited to specific models, and security management policy formulation does not depend on specific security products.

- Logical isolation

  Network security technologies, such as the firewall, Anti-DDoS, IDS, IPS, network antivirus, and Web security gateway, support the one-to-N virtualization mode, and can build logical boundaries for distributed cloud data centers (which do not have clear physical boundaries), to ensure VDC security.

- Flexible scalability

  The security subsystem is a guiding framework. Users can implement security construction based on the guiding framework and security requirements, which protects investments while meeting security requirements.

- Standards compliance

  The security subsystem of the SD-DC^2 is designed from aspects of physical security, network security, host security, application security, data security, user management, and

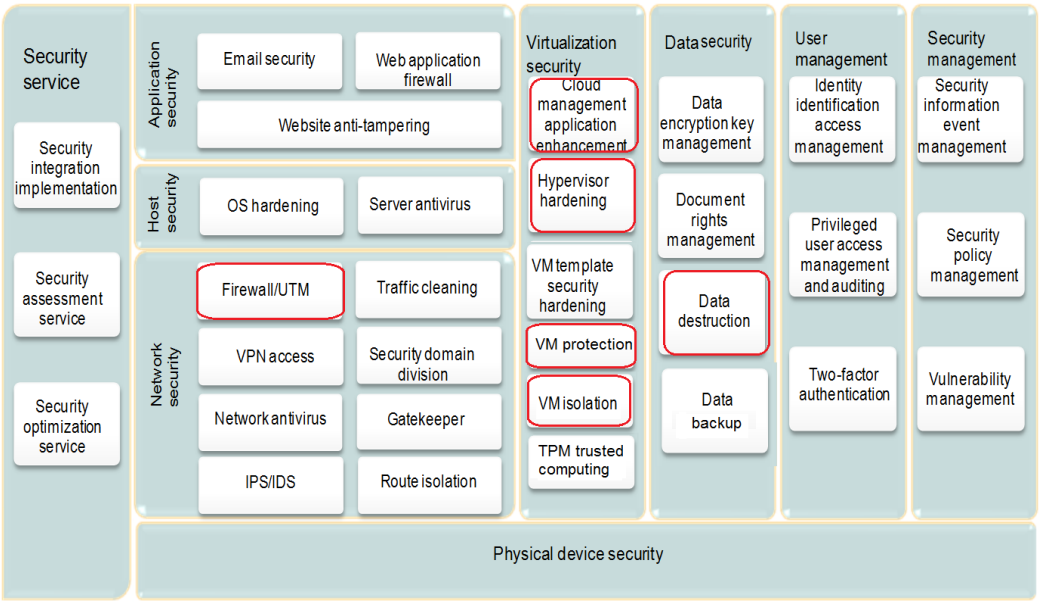
security management to meet level 3 security protection requirements of e-government. The security subsystem is one of the best guiding frameworks for constructing e-government data centers. Virtualization security is also ensured by the security subsystem based on cloud computing characteristics.

The SD-DC^2 security solution meets the requirements for e-government security protection (level 3).

2.8.2 Deployment Architecture

According to the ideas of layered and in-depth defense, the security subsystem is divided into physical device security, network security, host security, application security, virtualization security, data security, user management, and security management layers. The security subsystem meets different security requirements. The following figure shows the security subsystem architecture, in which the modules in red boxes are the basic security modules of the DC^2.

Figure 2-19 Security subsystem architecture



This architecture provides the following security capabilities:

- Physical device security: uses the access control system, video surveillance system, and environment monitoring system to control physical access and ensure the security of data center environments and facilities.
- Network security: uses the firewall, IPS, SSL VPN, Anti-DDoS, IDS/IPS, and network isolator technologies to ensure the isolation and security of VDC boarder, VDC internal system, data, and communication. These technologies prevent data from being damaged, changed, or disclosed accidentally or intentionally. With these technologies, the system is reliable, secure, and able to run continuously without service interruption.
- Host security: protects host OSs. Hosts are protected against attacks by security hardening, antivirus software, host IPS, and host patch management.
- Virtualization security: implements virtualization layer hardening, cloud management application hardening, and VM isolation to ensure virtualization security.

- Application security: uses protection technologies, such as the email protection technology and Web application protection technology, to protect the data on the application layer. These technologies prevent application data from being damaged, changed, or disclosed accidentally or intentionally.
- Data security: uses data encryption, residual data protection, data backup, and other technologies to ensure data security.
- User management: audits access requests from privileged users.
- Security management: adopts security information and event management technologies.
- Security service: covers security integration, security assessment, security optimization, and phase-specific professional services, and constructs a secure IT system for users.

2.8.3 Key Features

1. Network security

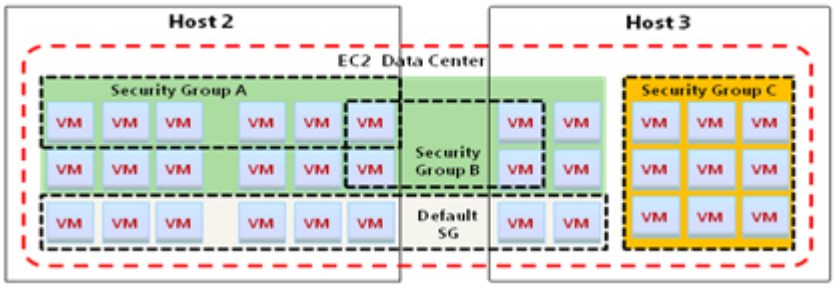
   The firewall, IDS, IPS, SSL VPN, Anti-DDoS, network antivirus gateway, and data ferry technologies are used to protect systems and communication data. These technologies prevent data from being damaged, changed, or disclosed accidentally or intentionally. With these technologies, the system is reliable, secure, and able to run continuously without service interruption.

   Traditional physical boundaries cannot meet security requirements for scenarios where VDCs are used as the main body of distributed cloud data centers. To meet the requirements of cloud technology development, network security products evolve to support virtualization as well as one-to-N device virtualization, and provide logical network security isolation. The vFW technology is the most widely used cloud technology. In addition, cloud technology-based software boundary firewalls and security groups provide comprehensive security protection.

   – Security group

     Users can create security groups based on VM security requirements. Each security group provides a set of access rules. VMs that are added to a security group are protected by the access rules of the security group. Users can add VMs to security groups when creating VMs.

     Figure 2-20 Security groups

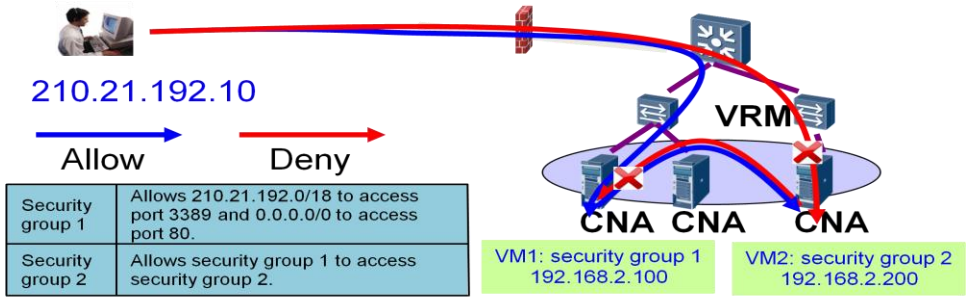


     Security groups work on vNICs. As shown in the following figure, one VM has two NICs, one of which belonging to security group A and the other belonging to security group B. VMs in the same security group can be distributed on different physical servers. The VMs in a security group can communicate with each other, while the VMs in different security groups are not allowed to communicate with each other. However, the VMs in different security groups, when configured, can also communicate with each other. Security groups allow all outbound packets, and deny

all inbound packets by default. Users can set inbound rules (similar to whitelists) to manage sources of data packets.

Users can configure security group rules as follows:

- Inter-group authorization: specifies which security groups can access a specific security group.
- VM authorization: specifies which network devices can access a specific VM.



As shown in the previous figure, VM 1 belongs to security group 1 and VM 2 to security group 2. After the security group rule is applied, VM 1 has access to VM 2 but VM 2 cannot access VM 1. 210.21.192.0/18 has access to port 3389 of VM 1 and all IP addresses have access to port 80 of VM 1.

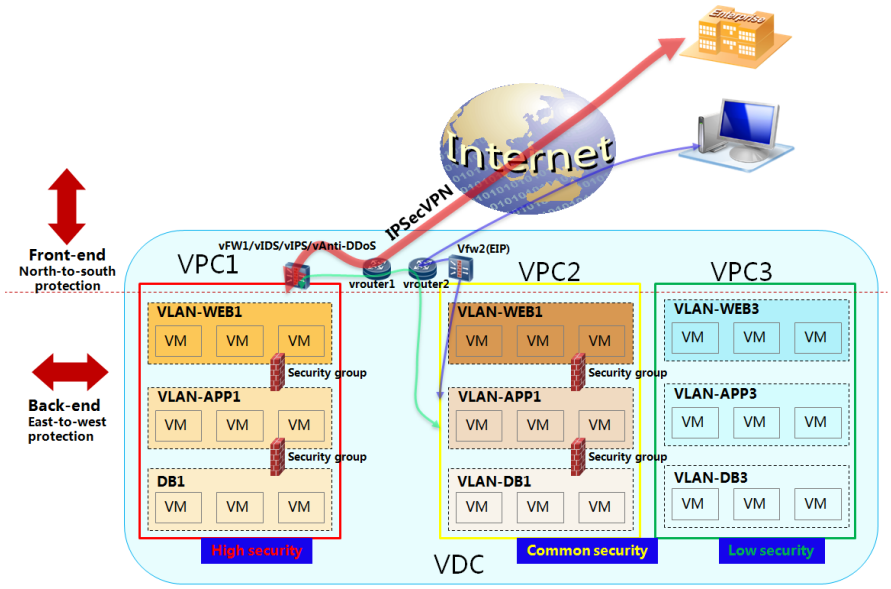
- Next-generation firewall protection

  Huawei next-generation firewall (NGFW) supports the firewall, VPN, IDS, IPS, Anti-DDoS, antivirus gateway, anti-spam protection, and Web protection technologies. These security protection technologies can also be virtualized.

- VDC network security protection framework

  On VDC boundaries, vFWs (hardware firewall in one-to-N virtualization mode or software VSA) and network security technologies such as vIDS, vIPS, and vAnti-DDoS are deployed to protect the north-to-south traffic of VDC. In a VDC, VPC boundaries protect the east-to-west traffic between VPCs using vFWs. In a VPC, the east-to-west traffic between applications is protected by security groups.

Figure 2-21 VDC security protection framework



- Anti-counterfeit of IP addresses and MAC addresses

  Binding an IP address to a MAC address prevents users from initiating IP address or MAC address spoofing attacks after changing the IP address or MAC address of a virtual NIC. Therefore, enhance the network security of user VMs. With this policy enabled, an IP address is bound to an MAC address using DHCP snooping feature, and then the packets from untrusted sources are filtered through IP Source Guard and dynamic ARP inspection (DAI).

- DHCP quarantine

  DHCP quarantine of VMs is supported. DHCP quarantine disallows users from unintentionally or maliciously enabling the DHCP server service for a VM, ensuring common VM IP address assignment.

- Broadcast packet suppression

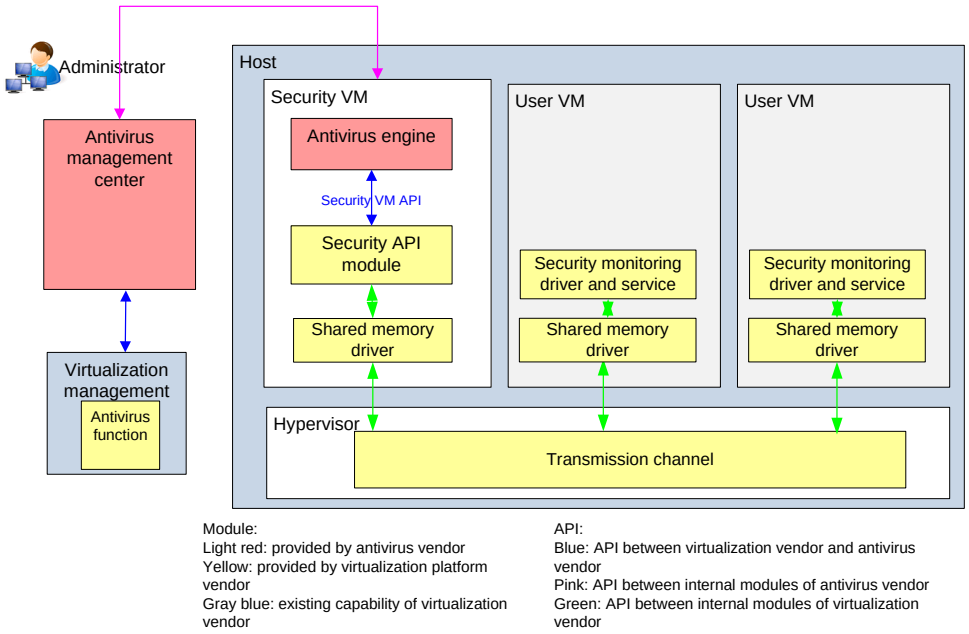
  In server consolidation and desktop cloud scenarios, if broadcast packet attacks occur due to network attacks or virus attacks, the network communication may be abnormal. In this case, the broadcast packet suppression can be enabled for vSwitches.

  The ARP broadcast packet suppression, IP broadcast packet suppression, and suppression bandwidth threshold for VM outbound traffic can be configured on vSwitches. Layer 2 network bandwidth consumption can be reduced by enabling the broadcast packet suppression and setting thresholds on the port group where the VM NIC is located.

  On the system portal, the manager can enable the ARP and IP broadcast packet suppression and set thresholds on a port group basis.

2. Virtualization agentless antivirus

Figure 2-22 Virtualization antivirus architecture



The Huawei virtualization platform provides APIs for antivirus vendors to perform secondary development and generate virtualization antivirus solutions that allow users to remove viruses by deploying an antivirus engine in a specific security VM and installing a lightweight driver on a local VM. The virtualization agentless antivirus supports the integrated verification with Rising, Trend Micro, and Kaspersky virtualization antivirus software, and supports the agentless antivirus on Windows cloud hosts.

The agentless antivirus has two advantages:

Advantages on virus library management, that is, only security VM management is required, instead of management of virus library installation and update on each VM.

Virus scan results are shared on all VMs of the host, which improves the virus scanning efficiency.

2.9 DR Service

2.9.1 Solution Overview

An increasing number of governments, telecom carriers, and large enterprises have demand for their own cloud DC DR systems. To help them resolve the DR issue, Huawei provides the FusionCloud BC&DR Solution, which consists of two sub-solutions: cloud backup solution and cloud active-passive solution. In an OpenStack-compliant cloud environment, DR service resources can be provisioned based on tenants' service level agreement (SLA) requirements. Then tenants in a cloud DC can possess virtual machine (VM) data protection and business continuity capabilities in self-help way.

The FusionCloud DR Solution (for Private Cloud) has three technical highlights: unified DR service operation and management (O&M) interface, service orchestration and scheduling, and ensured system reliability.

- Unified DR service O&M interface

  The FusionCloud BC&DR solution is integrated with Huawei cloud O&M platform ManageOne ServiceCenter (SC), on which tenants apply for DR service resources and operate backup services. ManageOne ServiceCenter enables multi-tenant backup management and backup service process management. ManageOne ServiceCenter also integrates Elastic Cloud Server (ECS) and Elastic Volume Service (EVS) O&M, simplifying cloud DC management.

- Service orchestration and scheduling

  The FusionCloud BC&DR solution has OceanStor DJ inside to provide data protection service catalogs for ManageOne ServiceCenter, and BCManager inside to provide full backup and permanent incremental backup capabilities for tenant VMs.

- Ensured system reliability

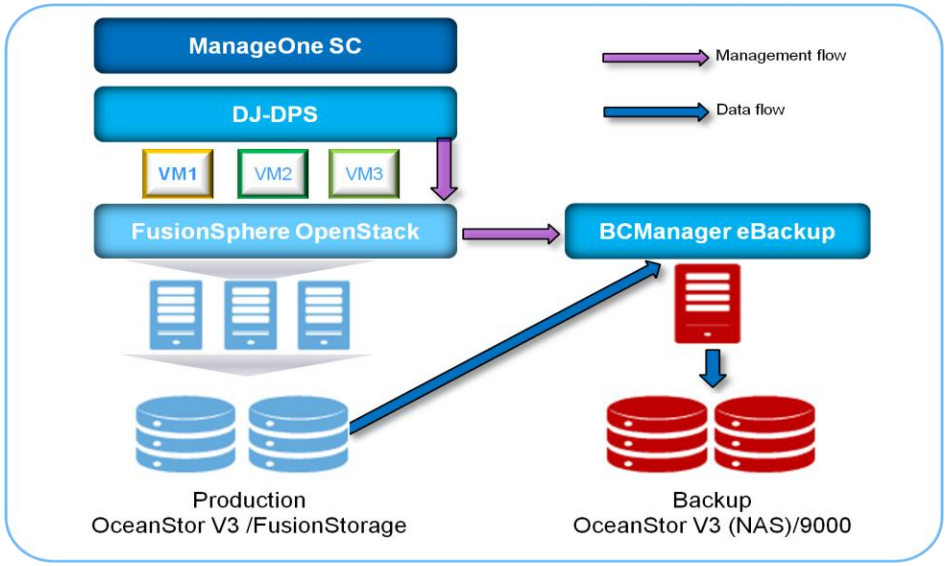
  The FusionCloud BC&DR solution offers comprehensive DR technologies, including VM volume data backup, application-specific multi-VM cross-site active-passive DR, and high availability (HA) with the DR system itself.

2.9.2 Logical Architecture of the Backup Solution

2.9.2.1 Logical Architecture of the Backup Solution

This chapter introduces the logical architecture of the backup solution.

Figure 2-23 Logical architecture of the backup solution



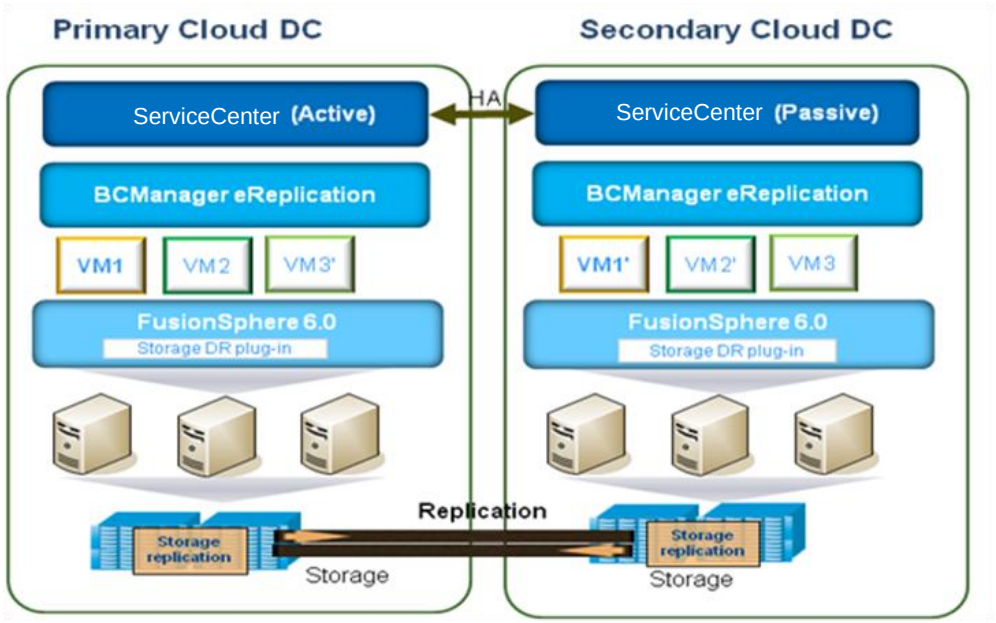
As shown in Figure 2-23, the cloud backup solution has five layers: cloud management platform ManageOne ServiceCenter, data protection service platform OceanStor DJ-DPS, virtualization cloud platform FusionSphere OpenStack, backup management software BCManager eBackup, and backup storage devices OceanStor V3 NAS/OceanStor 9000. The solution architecture is as follows:

- Backup service management flow: ManageOne ServiceCenter issues the backup service when tenants request. OceanStor DJ-DPS orchestrates and schedules the backup service as instructed. OceanStor DJ-DPS invokes the storage drivers integrated in FusionSphere

OpenStack to trigger the production storage to create snapshot volumes, and invokes backup drivers to trigger eBackup to back up snapshot volumes.

- Backup data flow: eBackup reads snapshot volume data from production storage devices over SAN and backs up the data, and then uses backup servers to write backup data to back up storage devices.

2.9.2.2 Logical Architecture of the Active-Passive DR Solution



Huawei FusionCloud BC&DR solution is underlain by Huawei cloud platform FusionSphere 6.0 and leverages storage replication technologies. It is the first in the industry to deliver OpenStack cross-site DR capability. It requires a FusionSphere cloud platform at each DR site and those sites share the same certification and authentication system Keystone. Logically, resources of each site are centralized in a unified space. When a site passes Keystone certification and authentication, it has access to all other sites. In this way, resources and DR services are provisioned and centrally managed. In the FusionCloud BC&DR solution, ManageOne ServiceCenter provisions DR services, BCManager eReplication configures and manages DR services, and storage DR plug-ins offer connectivity to OpenStack Cinder API for cross-site DR.

ManageOne ServiceCenter provisions VM computing resources, LUN storage resources, and virtual private cloud (VPC) network resources. On ManageOne ServiceCenter, tenants can apply for and maintain DR resources. ManageOne ServiceCenter is deployed in active-passive mode. The active node is deployed in the primary cloud DC and the passive node in the secondary cloud DC. After the active node fails, its services can be manually switched to the passive node.

BCManager processes DR management logic. It interworks with OpenStack, invoking the computing interface (Nova API), storage interface (Cinder API), and DR management interface (DRExtend API) of OpenStack to perform functions including service protection configuration, fault switchover, and DR drilling. BCManager can be deployed in DR sites.

Storage arrays provide the replication function, which efficiently replicates VM data from the production site to DR sites without no impact on host services. If a DR site has part of

production services, the production site can function as a DR site for that DR site, making the production site and the DR site manually back up for each other.

2.9.3 FusionCloud BC&DR Solution Deployment

2.9.3.1 Backup Solution Deployment

The backup system is deployed based on the FusionSphere OpenStack production environment. Deploy one set of eBackup backup software for each availability zone (AZ), one set of OceanStor DJ-DPS data protection service platform for each region, and one set of cloud management platform ManageOne ServiceCenter globally. See Figure 2-24

Figure 2-24 Physical network of the backup solution

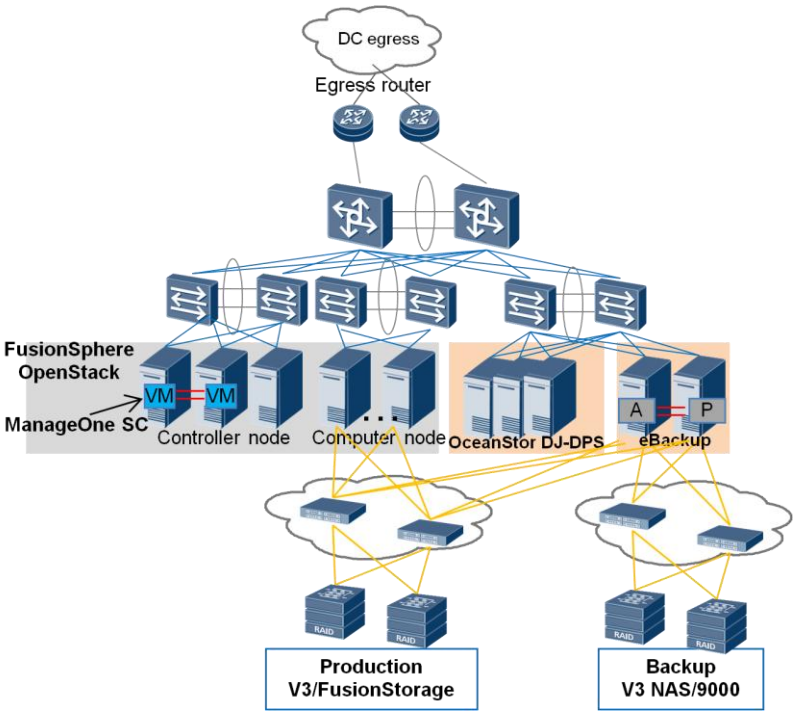


Table 2-3 lists the deployment of solution modules.

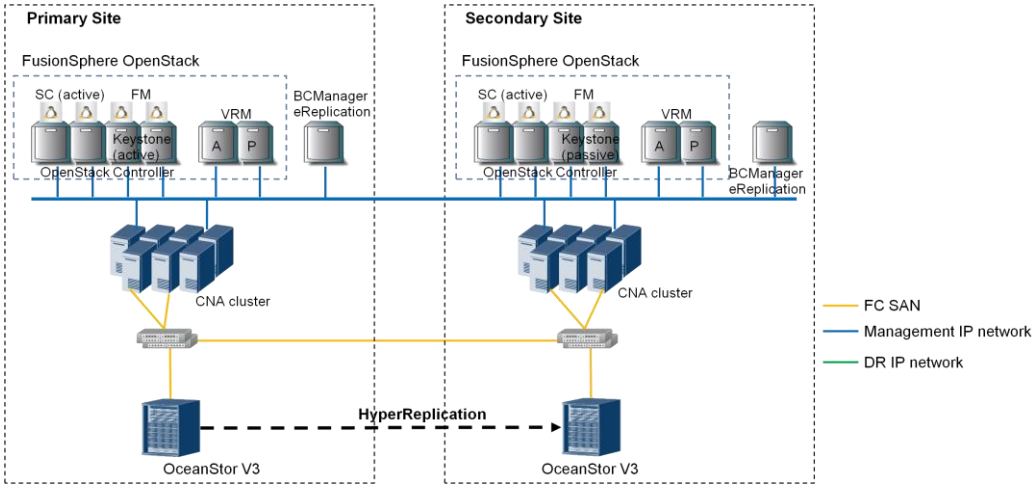
Table 2-3 Typical deployment Principle:

Component	Deployment Principle
FusionSphere OpenStack (VRM/KVM)	FusionSphere OpenStack is deployed on three physical controllers. FusionCompute Virtual Resource Manager (VRM) is deployed in VMs or physical machines (PMs) in active-passive mode. FusionCompute Computing Node Agent (CNA) is deployed on compute nodes. Cinder Driver and Backup Driver are deployed on control nodes.
ManageOne ServiceCenter	ManageOne ServiceCenter is deployed in active-passive (recommended) or standalone mode on VMs in the FusionSphere OpenStack virtualization environment.

Component	Deployment Principle
OceanStor DJ-DPS	OceanStor DJ-DPS is deployed on one or (recommended) three independent PM nodes. It can be deployed in a virtualized way.
eBackup	eBackup is deployed on two PM backup management servers, which form an active-passive HA deployment. eBackup backup proxy nodes and backup management servers can share server resources. One or more backup proxy nodes can be deployed based on the number of VM volumes.

2.9.3.2 Active-Passive DR Solution Deployment

Figure 2-25 Physical networking of active-passive DR:



The FusionCloud BC&DR solution has four network planes: cloud platform management plane, front-end host VM data plane, back-end storage access plane, and DR data plane. The second to fourth planes are similar to those in general DR solutions. The following introduces only the cloud platform management plane.

Based on Huawei cloud solutions, the FusionCloud BC&DR solution has three additional components, ManageOne ServiceCenter, Keystone active-passive deployment, and BCManager eReplication. eReplication can be deployed on PMs or VMs. The following introduces ManageOne ServiceCenter and Keystone active-passive deployment.

Keystone is the cloud certification and authentication component. All access requests for cloud resources and interfaces must be authenticated by Keystone. To ensure its reliability, Keystone must possess HA and DR capabilities. In the FusionCloud BC&DR solution, Keystone is deployed in active-passive mode. The active Keystone is deployed in the active site and the passive Keystone in the passive site. The Keystone active-passive deployment requires one additional compute node besides three OpenStack controllers to form a four-controller structure. After the active DC fails, manually switch the Keystone service to the passive DC.

ManageOne ServiceCenter issues DR services. It must possess HA and DR capabilities. In the local ManageOne ServiceCenter, dual-host deployment ensures HA. In the passive DC, two

nodes are deployed to receive data synchronized from the active DC. After the active DC fails, manually switch the ManageOne ServiceCenter service to the passive DC, and then services will be provisioned from the passive DC. ManageOne ServiceCenter can be deployed on VMs.

2.9.4 Unified Portal

The FusionCloud BC&DR solution uses a unified portal for service operation. The portal involves two roles, managers and tenants.

After you log in as a manager, you can release or cancel DR services, add DR services to the service catalog, approve or reject tenants' application for DR services, apply for or cancel DR services for tenants, and view all tenants' service metering information.

After you log in as a tenant, you can apply for and confirm DR services as well as query all the requests you have made. Tenants apply for and use backup services online. After tenants apply for backup services on ManageOne ServiceCenter, backup services are provisioned online. Then ManageOne ServiceCenter performs backup and recovery jobs online according to the backup scheduling policy. Alternatively, tenants can manually start backup and recovery jobs when needed on ManageOne ServiceCenter.

Tenants apply for and use DR services online. After tenants apply for DR services and VMs in ManageOne ServiceCenter, background DR managers provision DR services and performs related operations including DR creation, drilling, and switchover.

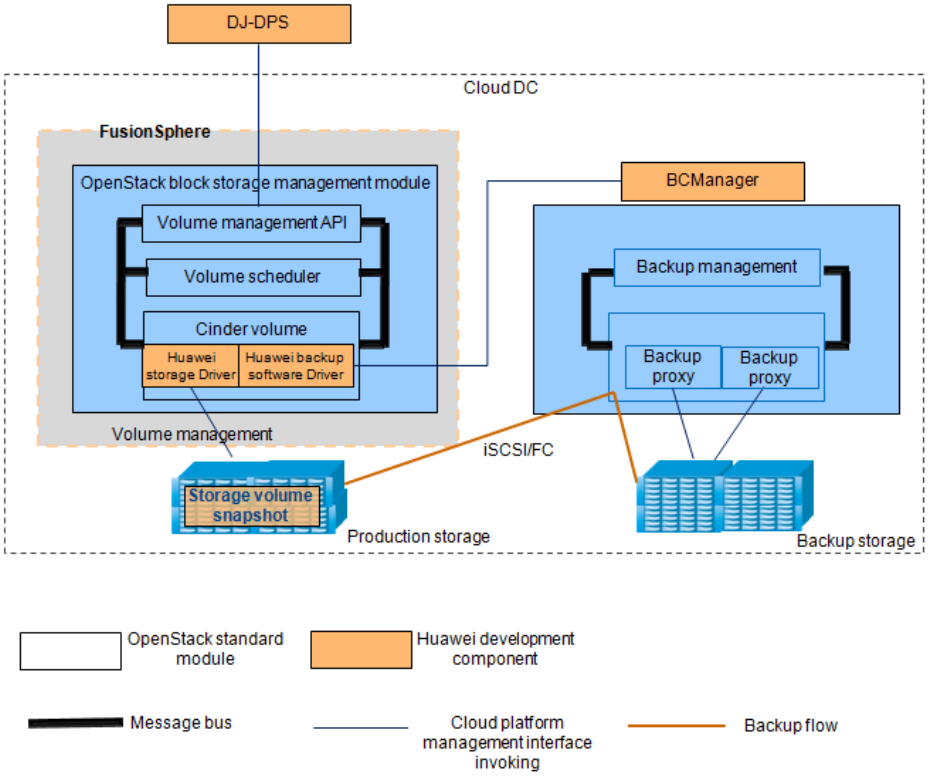
Table 2-4 Roles who operate DR services

Role	Responsibility	Task
Manager	Operates and manages DR services.	Creates, releases, and cancels DR services. Creates tenants. Approves or rejects service requests. Views the number of DR services and the backup capacity metering statistics of DR services.
Tenant	Uses DR services.	Applies for, modifies, and cancels DR services. Views the tenant's DR service list, service details, backup capacity metering information, and backup and recovery job status. Manually backs up and recovers volumes as well as deletes backup copies.

2.9.5 Key Backup Technologies

2.9.5.1 Cinder BackupInterface Extension Architecture

Figure 2-26 Extension architecture



During the construction of a backup system, impact minimization on the production system must be considered. To ensure that online backup, incremental backup, and other backup services do not affect production nodes, Huawei provides extensions for OpenStack backup capability. Figure 2-26 shows the two extensions: Cinder Driver for storage and eBackup Driver for backup software. The two drivers combine with BCManager eBackup backup software to deliver two backup features:

1. Online and incremental backup

   Cinder Driver is designed for Huawei storage. It offers online backup and recovery capabilities for Huawei OceanStor V3 and FusionStorage volumes. It can compare between snapshots, specifically, identifies the differences between two consecutive snapshots and then performs an incremental backup.

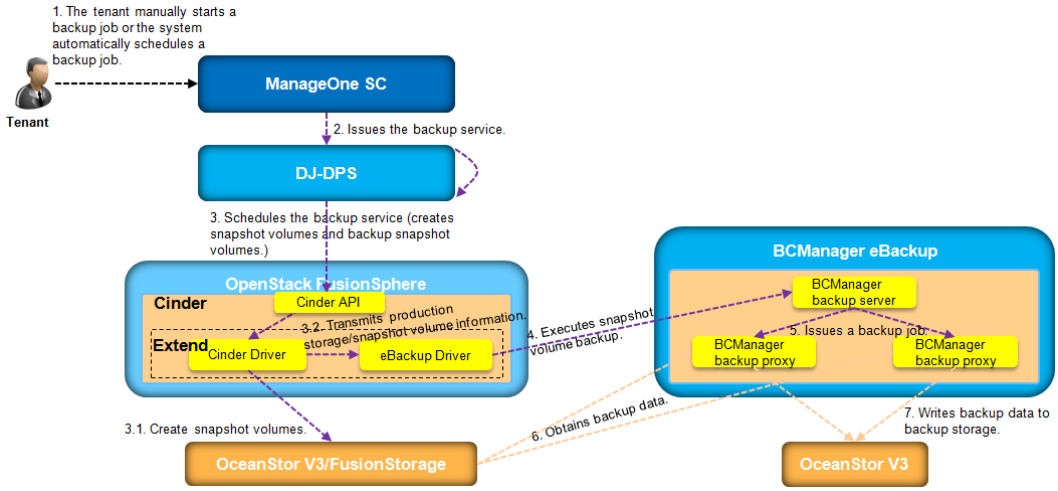
2. Backup data processing without passing through a Cinder management node

   An issue with an extended interface is how to trigger backup software to perform a backup job and how to direct backup data flow. To enable eBackup to execute backup and recovery, Huawei develops eBackup Driver. With eBackup Driver, OpenStack obtains the capability of interacting with BCManager eBackup to back up and recover volumes. During backup and recovery, Cinder Volume Driver does not need to mount volumes. Instead, eBackup Driver transmits volume information to the eBackup server, and then the eBackup server mounts FusionStorage or OceanStor V3 volumes, occupying no host resources of the management node that runs Cinder Volume Driver.

2.9.5.1.1 Backup Automation

In addition to manual backup, tenants can also select automatic scheduling in the backup policy. Then backup jobs are automated according to the scheduling policy. Figure 2-27 shows the automated backup process.

Figure 2-27 Backup automation



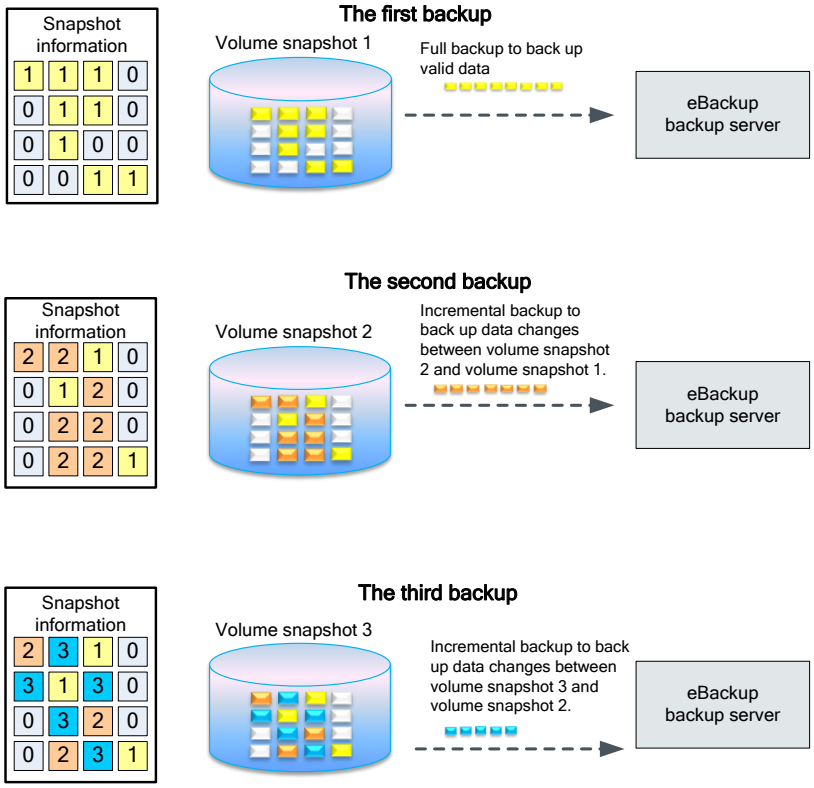
A backup process is automated with the following steps:

1. The tenant selects manual backup, or enables the automatic scheduling policy to trigger an automatic backup job.
2. ManageOne ServiceCenter issues the backup service to OceanStor DJ-DPS according to the backup scheduling policy.
3. As the backup scheduling engine, OceanStor DJ-DPS schedules the backup service according to the information sent by ManageOne ServiceCenter.
   (1) Sends a message of creating volume snapshots to FusionSphere OpenStack Cinder, for Cinder Driver to invoke the production storage to create volume snapshots.
   (2) Sends a message of creating backups for backup snapshot volumes to FusionSphere OpenStack Cinder, for Cinder Driver to send production storage and volume snapshot information to Backup Driver.
4. eBackup Driver invokes the backup software BCManager eBackup to perform volume snapshots.
5. The eBackup management node issues a backup job to the backup proxy node.
6. The backup proxy node obtains volume snapshot data from the production storage and backs up the volume snapshot data.
7. The backup proxy node writes the backup data to the backup storage.

After the backup job is complete, temporary snapshot volumes created by the storage system for the backup purpose are automatically deleted.

2.9.5.1.2 Permanent Incremental Backup

Working principle: Based on storage snapshot and CBT technologies, data blocks are obtained from production storage and no backup proxy is required in protected VMs, thereby reducing the impact on user services and simplifying backup solution deployment and maintenance.

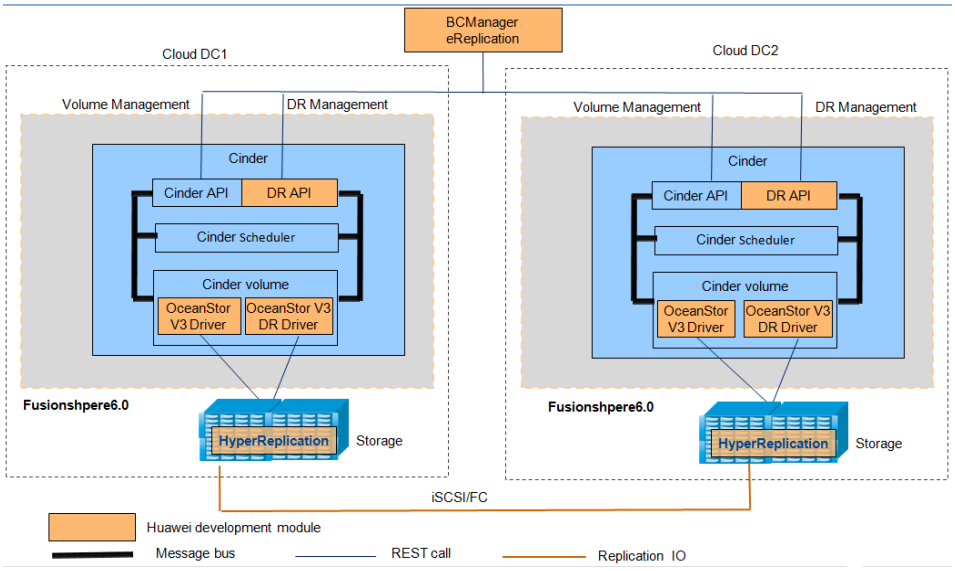


For each backup job, the storage system creates a snapshot for the backup object VM. eBackup performs a full backup initially using snapshot data, and performs incremental backups (backs up only data changes) afterwards by comparing the current data with the last piece of snapshot data. After the backup job is complete, eBackup retains only the latest snapshot and deletes all other VM volume snapshots. The last snapshot will be used to determine data changes since the last backup job.

2.9.5.2 Active/Passive DR service

2.9.5.2.1 Cinder Cross-Site DR API Extension

Architecture of Cinder Cross-Site DR API Extension:



To prevent service interruption against natural disasters such as earthquakes and floods, the production and DR DCs must be deployed across sites. The native OpenStack Cinder API does not support replication across Cinder (sites) but only intra-Cinder volume replication and switchover. The DR relationship established in the primary cloud DC cannot be queried or managed in the secondary cloud DC. When the primary cloud DC fails, cross-site switchover is unavailable. Therefore, the OpenStack Cinder DR API must be extended. The active-passive DR solution designed by Huawei supports DR replication cross OpenStack sites and VM-level DR management. API extension must resolve two issues:

- The API must maintain normal compatibility with the OpenStack community after being extended.
- VM DR must be supported cross sites.

To maintain the normal compatibility with the OpenStack community, Huawei does not modify the Cinder API directly but use the DRExtend API instead. In this way, the DRExtend API is not affected even after the Cinder API has updates. Third-party storage vendors that use the DRExtend API can interwork with OpenStack for unified management.

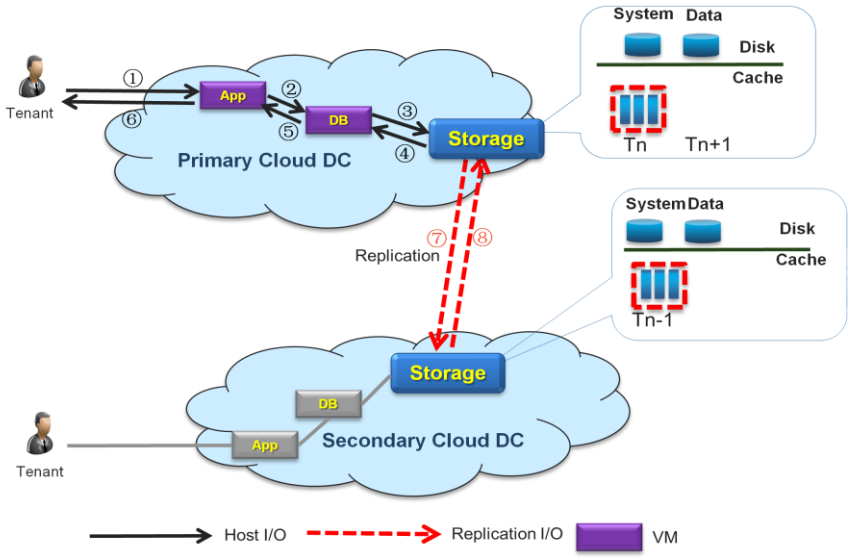
Another challenge is to synchronize the OpenStack DR configuration data of the production center to the OpenStack in the DR center. Huawei uses BCManager eReplication software to invoke the DR creation interface of the DRExtend API in both the production and DR centers. The DR creation interface creates DR configuration data in the production OpenStack and DR OpenStack. Then the transmission flag is used to determine whether to perform creation on storage devices. That prevents relationship creation failures in a scenario in which a LUN has a DR configuration on the production storage already and then a DR configuration is also created for that LUN on the DR storage. Replication consistency groups are created when creating a replication relationship. They can be used for VM-level replication. In addition, together with BCManager eReplication, they enable VM-level DR management activities, such as DR protection, fault switchover, and DR drilling.

2.9.5.2.2 Cross-Cinder Replication

On the data plane of the FusionCloud BC&DR solution, the Cinder DR API and the storage replication capability combine to implement cross-Cinder and cross-site data replication, including synchronous replication with an RPO of zero, asynchronous replication with an

RPO of seconds, and reverse incremental synchronization. As an example, Figure 2-28 illustrates the working principle of asynchronous replication.

Figure 2-28 Working principle of asynchronous replication

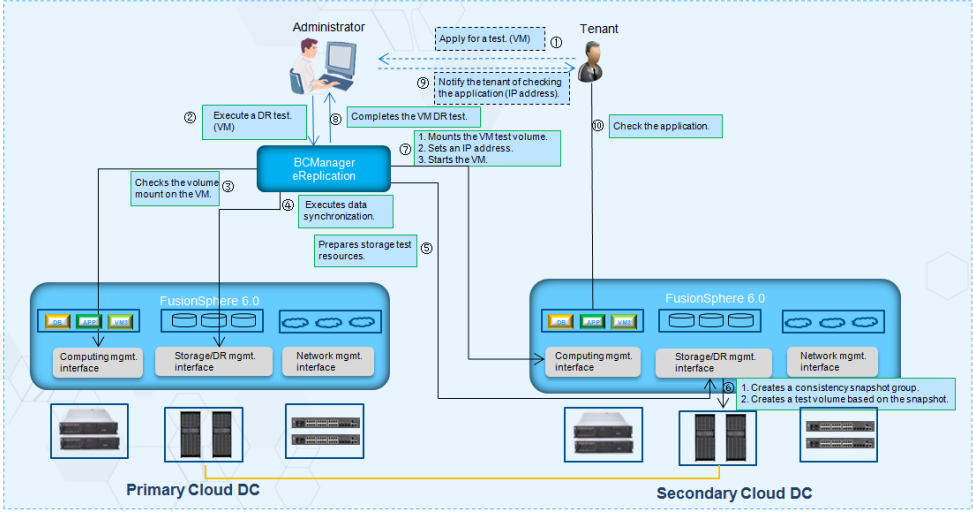


In a cloud DC, applications and databases are running on VMs of cloud servers. Tenants access applications. Data generated by applications is stored in databases. I/Os are stored on storage devices using those VMs. In a synchronous replication process, tenants use applications to deliver I/Os to databases. Databases deliver I/Os to storage devices and then return them to hosts instantly without waiting them to reach passive storage devices, saving the time used by hosts to return host I/Os.

Tags are added in the cache periodically to generate consistency points on the source storage device. The I/Os delivered before the next replication period comes are all stored in the cache tagged T_n. When the next replication period comes, a new cache is created and is tagged T_{n+1}. New I/Os will be stored in the cache and data in T_n will be replicated to the cache of the passive storage device till the replication process is complete.

2.9.5.2.3 DR Testing

DR testing has two purposes. On one hand, DR testing checks whether and how long data and VMs in the secondary cloud DC can be started. On the other hand, in a DR test, the data and VMs in the secondary cloud DC are used for query and analysis as well as application testing. DR testing does not require stopping VM services in the primary cloud DC.



A DR test process includes four steps. The DR manager prepares test resources and the tenant checks services. Except test starting and test result checking, all other operations are automatically performed by the system in the background.

1. The tenant applies to the manager for a DR test, and notifies the VM of the DR test.

   Currently, DR testing is not provisioned as a service to tenants. When tenants want to perform DR testing, they need to apply to the manager and the DR manager starts a DR test.

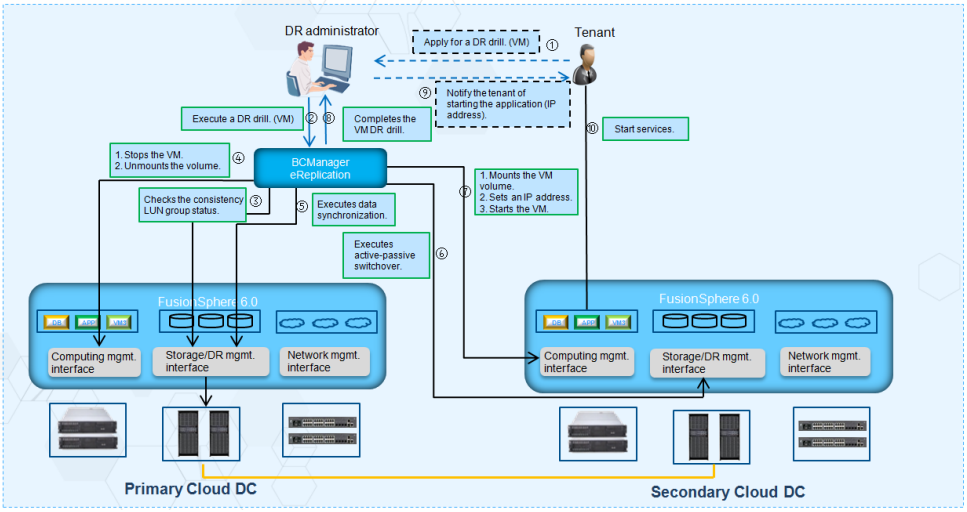
2. After receiving a DR test request, the DR manager logs in to BCManager eReplication and starts a VM DR test.
   (1) BCManager queries the VM on which the DR test is to be performed.
   (2) The manager performs the VM DR test.
   (3) BCManager eReplication checks whether the volumes mounted on the test object VM in the primary cloud DC are consistent with the volumes associated with the VMs in the BCManager eReplication protection group. If they are not consistent, the test will not be continued.
   (4) BCManager eReplication queries the replication status of the consistency LUN group mounted on the test object VM. If data synchronization is required, a data synchronization will be performed.
   (5) After the data synchronization is complete, BCManager eReplication creates test storage resources on the storage device of the secondary cloud DC. It invokes the storage interface to create a consistency snapshot group, and then creates snapshot copies accordingly for the VM to mount.
   (6) BCManager eReplication invokes the computing management interface to mount snapshot copies to the test object VM. Then it sets an IP address and start the VM.
   (7) BCManager eReplication notifies the DR manager of the VM test status.
3. After the VM DR test is complete, the DR manager notifies the tenant of the result, prompting the tenant to test the application.
4. The tenant logs in to the VM, and starts and checks the application.

2.9.5.2.4 DR Drilling

The purpose of DR drilling is to check whether the DR data and facilities are available in the secondary cloud DC, and whether services can be started in the secondary cloud DC when a disaster occurs in the primary one.

DR drilling has two differences from DR testing. DR drilling requires stopping the object VM in the primary cloud DC and migrating the object VM to the secondary cloud DC. After DR drilling is complete, the VM needs to be migrated back to the primary cloud DC. Another difference is that DR tests are performed in snapshot groups of the secondary cloud DC, whereas DR drilling is performed directly in LUN groups.

Figure 2-29 DR drilling interaction process



A DR drilling process includes the following steps:

1. The tenant applies to the manager for a DR drill, and notifies the VM of the DR drill.

   Currently, DR drilling is not provisioned as a service to tenants. When tenants want to perform DR drilling, they need to apply to the manager and the DR manager starts a DR drill.

2. After receiving a DR drill request, the DR manager logs in to BCManager eReplication and starts a VM DR drill.

   (1) BCManager queries the VM on which the DR drill is to be performed.

   (2) The manager performs the VM of DR drill.

   (3) BCManager eReplication queries the replication status of the consistency LUN mounted on the drill object VM in the primary cloud DC. If the replication status is abnormal, the drill will not be continued.

   (4) BCManager eReplication invokes the computing management interface to unmount the volume from the VM.

   (5) BCManager eReplication invokes the computing management interface to stop the VM.

   (6) BCManager eReplication invokes the DR management interface to perform data synchronization for the consistency LUN group.

   (7) After the data synchronization is complete, BCManager eReplication invokes the DR management interface to perform an active-passive switchover.

(8) BCManager eReplication invokes the computing management interface to mount the consistency LUN group of the secondary cloud DC to the drill object VM. Then it sets an IP address and start the VM.

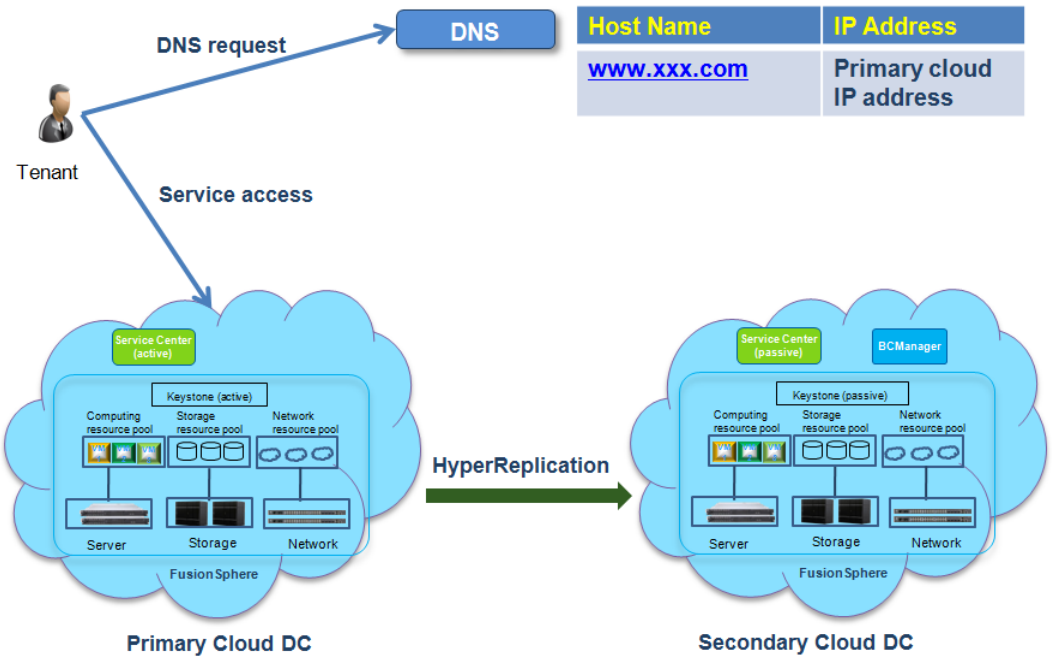
(9) BCManager eReplication notifies the DR manager of the VM drill status.

3. After the VM drill test is complete, the DR manager notifies the tenant of the result, prompting the tenant to run services in the secondary cloud DC.
4. The tenant logs in to the VM of the secondary cloud DC to start the application and check whether services are running properly.

2.9.5.2.5 One-Click Fault Switchover

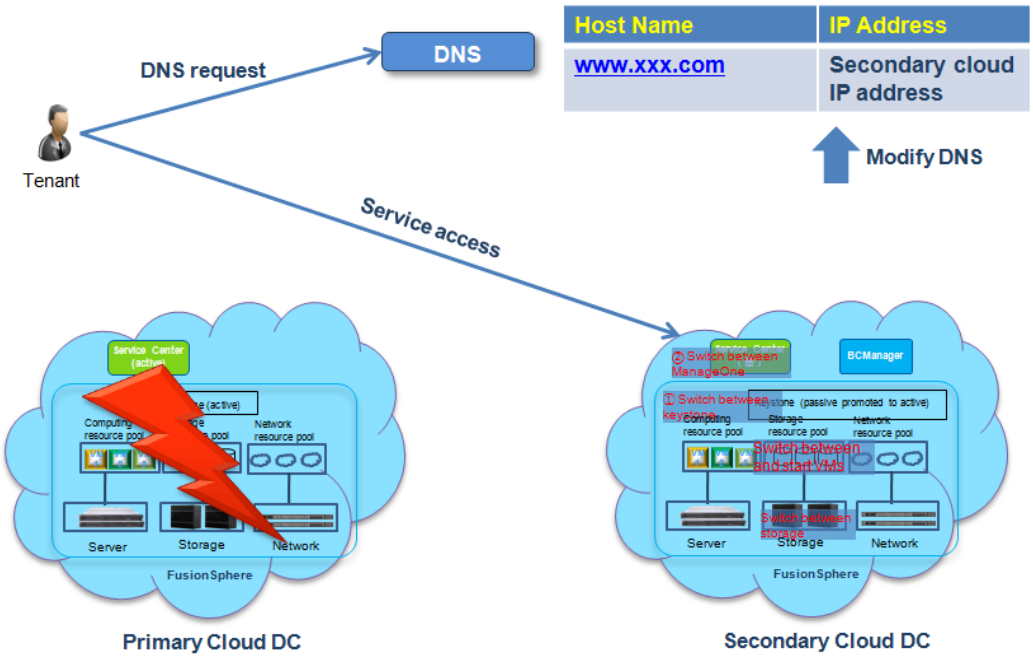
Fault switchover is different from DR testing and DR drilling. DR testing and DR drilling are conducted when the cloud platform and the DR management system are running properly, whereas fault switchover is conducted when the primary cloud DC breaks down. When the cloud platform, service provisioning system, and DR management system in the primary cloud DC all malfunction, the service provisioning system and DR management system need to be switched over to the secondary cloud DC, ensuring that cloud DC functions such as resource provisioning and operation can continue in the secondary cloud DC.

Figure 2-30 Fault switchover



In the active-passive DR solution, when the primary site fails, manually switch services over to the secondary site. Figure 2-31 illustrates the switchover process.

Figure 2-31 Switchover process



A fault switchover process includes switching between DR environments, between service systems, and between domain name systems (DNSs). When the DR environment is ready, the DR manager can switch between service systems by one click in BCManager eReplication.

Switching between DR environments:

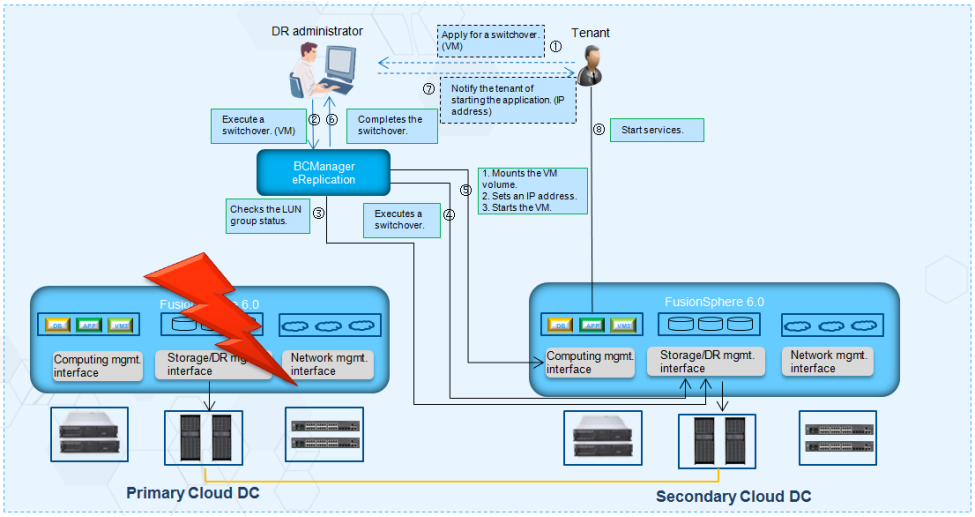
1. Switch between Keystone, the access authentication component.
2. Switch between ManageOne ServiceCenter, the service provisioning system.

Switching between service systems:

1. Switch between storage systems.
2. Switch between VMs.
3. Start up the VM and the service system.

   Figure 2-32 illustrates the interaction process during a BCManager eReplication fault switchover.

Figure 2-32 Interaction process



4. The tenant applies to the DR manager for fault switchover.

   Currently, fault switchover is not provisioned as a service to tenants. When tenants want to perform a fault switchover, they need to apply to the manager and the DR manager starts a fault switchover.

5. After receiving a fault switchover request from tenants, the DR managers logs in to BCManager eReplication and starts a fault switchover by one click. The internal execution is as follows:

   (1) BCManager queries the VM on which the fault switchover is to be performed.

   (2) The manager performs the VM of a fault switchover.

   (3) BCManager eReplication uses the storage query interface to check whether the replication status of the LUN group in the secondary cloud DC is faulty. If the replication status is not faulty, the fault switchover will not be continued.

   (4) BCManager eReplication invokes the DR management interface to perform a fault switchover.

   (5) BCManager eReplication invokes the computing management interface to mount the consistency LUN group of the secondary cloud DC to the fault switchover object VM. Then it sets an IP address and start the VM.

   (6) BCManager eReplication notifies the DR manager of the fault switchover status.

6. After the fault switchover is complete, the DR manager notifies the tenant of the result, prompting the tenant to run services in the secondary cloud DC.

7. The tenant logs in to the VM of the secondary cloud DC to start the application and direct services to the secondary cloud DC.