

**Data Center Solution  
V100R001C00  
Network Design Guide**

**Issue**      01  
**Date**        2011-08-31

**Copyright © Huawei Technologies Co., Ltd. 2011. All rights reserved.**

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

## **Trademarks and Permissions**



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

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

## **Notice**

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

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

## **Huawei Technologies Co., Ltd.**

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

Website: <http://www.huawei.com>

Email: [support@huawei.com](mailto:support@huawei.com)

---

# Contents

---

<b>1 Overview.....</b>	<b>1</b>
1.1 Objectives.....	1
1.2 Applicability.....	1
<b>2 Service System Design.....</b>	<b>2</b>
2.1 Overview.....	2
2.2 Data Services.....	2
2.2.1 Overview.....	2
2.2.2 Network Design.....	3
2.2.3 System Reliability.....	3
2.2.4 System Security.....	3
2.3 Web Services.....	4
2.3.1 Overview.....	4
2.3.2 Network Design.....	5
2.3.3 System Reliability.....	5
2.3.4 System Security.....	5
2.4 Computing Service Design.....	5
2.4.1 Overview.....	5
2.4.2 Network Design.....	6
2.4.3 System Reliability.....	6
<b>3 Physical Network Design.....</b>	<b>7</b>
3.1 Overview.....	7
3.2 Topology Design.....	8
3.2.1 Overview.....	8
3.2.2 Considerations.....	9
3.2.3 Design Principles.....	14
3.3 Node Devices.....	15
3.4 Link Selection and Design.....	17
3.4.1 Internal Link Design.....	17
3.4.2 Egress and Ingress Link Design.....	17
3.5 Planning of Device and Interface Names.....	18
3.5.1 Device Naming.....	18
3.5.2 Naming of Links and Interfaces.....	18

<b>4 Design of Logical End-to-End Features .....</b>	<b>19</b>
4.1 Overview .....	19
4.2 VLAN Design .....	19
4.2.1 Overview .....	19
4.2.2 Considerations .....	19
4.2.3 Design Principles .....	20
4.2.4 Design Elements .....	20
4.2.5 L2/L3 Allocation Design .....	22
4.3 IP Service and Application Design .....	25
4.3.1 Overview .....	25
4.3.2 IP Address Planning and Design .....	25
4.3.3 DNS Design .....	27
4.4 Route Design .....	32
4.4.1 Overview .....	32
4.4.2 OSPF Design .....	32
4.4.3 BGP Design .....	34
4.5 VPN Design .....	37
4.5.1 Overview .....	37
4.5.2 VPN deployment .....	37
4.6 Reliability Design .....	38
4.6.1 Overview .....	38
4.6.2 Device Reliability .....	39
4.6.3 Network Reliability .....	40
4.6.4 Service Reliability .....	41
4.7 Load Balancing Design .....	42
4.7.1 Overview .....	42
4.7.2 Design Principles .....	42
4.7.3 LB Deployment Modes .....	43
4.8 QoS Design .....	44
4.8.1 Overview .....	44
4.8.2 Service QoS .....	44
<b>5 Network Management Design .....</b>	<b>46</b>
5.1 Overview .....	46
5.1.1 NMS .....	46
5.1.2 Network Scale .....	46
5.1.3 NMS Design .....	47
5.2 eSight System Design .....	47
5.2.1 Overview .....	47
5.2.2 Considerations .....	48
5.2.3 Design Principles .....	50
5.2.4 Design Elements .....	50



# 1 Overview

---

## 1.1 Objectives

This document provides a guide for marketing personnel to implement high-level design (HLD) on data center projects and specifies requirements for and key points of HLD document writing.

This document provides a reference for a network solution design and focuses on HLD implementation based on the HLD template.

## 1.2 Applicability

This document is applicable to data center solution design. You can refer to this document to implement HLD of a project.

# 2 Service System Design

## 2.1 Overview

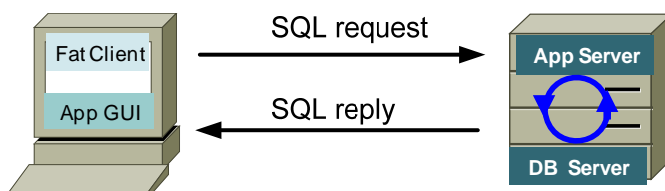
An enterprise-class data center involves three types of services: data, Web, and computing. These services may be independent of each other, or integrated into a large service system. You must design data center networks to meet the requirements of enterprise-class customers.

## 2.2 Data Services

### 2.2.1 Overview

Data services, such as file storage, mail system, and ERP, are basic for a data center. The basic service mode is client/server (C/S). For details, see [Figure 2-1](#).

**Figure 2-1** C/S service mode



The C/S mode consists of the following two parts:

- A client (usually a PC) is deployed on the foreground, that is, on the campus network of an enterprise or the branch network.
- A server is deployed on the background, that is, in a data center. Independent storage devices are used for storage space expansion of the server.

Data services have the following requirements on networks:

- Network bandwidth

Data service traffic is generated by data requests and responses between the client and server. Traffic volume changes frequently and may reach a peak value on a special date or within a time segment, such as a checkout day. High network bandwidth is required to prevent data service traffic from becoming congested or losing data.

- Network reliability  
When a data center link or network device is faulty, data service reliability (mainly data reliability) can ensure a timely recovery of data services.
- Network security  
Data center security is critical in an IT system because the data center processes services and data. Critical services in an enterprise, such as finance, are the data services. In addition to physical isolation, a network must protect service security.

## 2.2.2 Network Design

As data services are implemented in C/S mode, the network must provide high bandwidth. End-to-end bandwidth requirements must be considered in network design to avoid a traffic bottleneck on the network.

Network bandwidth can be improved by adding physical links or replacing existing boards with large-capacity boards, for example, replace a GE board with a 10 GE board, or replace a 10 GE board with a 40 GE board. In addition, you need to ensure that when a link fails, the other links can bear all service traffic.

## 2.2.3 System Reliability

Internal and external networks of an enterprise have different requirements for data service reliability as follows:

- Compared with external service systems, the internal service system of an enterprise has low requirements for network reliability. Failures that occur in a data center need to be rectified within 20 to 30 minutes. When the entire data center fails, services need to be recovered using the standby data center within 4 to 8 hours.
- The service system that provides services to external systems requires high reliability. Failures that occur in a data center can be rectified by automatic switchover or manually rectified within 10 minutes. When the entire data center fails, services need to be recovered using the disaster recovery center within two hours.

At the network layer, the key points in reliability design are as follows:

- Device reliability: Redundant device power supply units and main control boards
- Link reliability: Multi-link and load sharing technologies
- Network reliability: Bidirectional forwarding detection (BFD), Virtual Router Redundancy Protocol (VRRP), and ring protection technologies.
- Service reliability: Backup of data service systems and servers.

## 2.2.4 System Security

System security is designed at the following aspects:

- Service isolation: Implemented by using VLAN and virtual private network (VPN) technologies.
- Attack defense: Deploy network security devices to prevent network attacks from affecting data services.
- Antivirus: Deploy network security devices to prevent viruses from spreading in the data center.



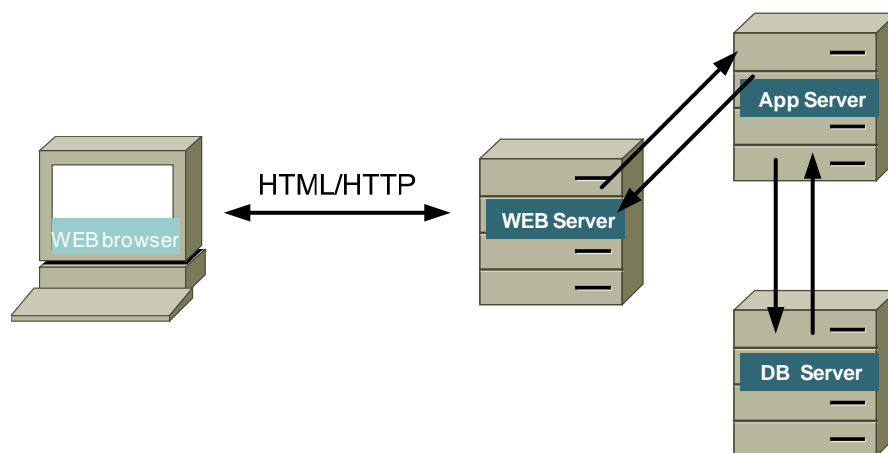
- Terminal protection: Enable an authorized terminal to access a server by means of security authentication. Use permission control technologies to control terminal permissions. This ensures that a service terminal accesses only a specified server.

## 2.3 Web Services

### 2.3.1 Overview

As Internet technologies develop, Web services occupy a growing proportion of an enterprise service system. This enables customers to access enterprise information through the Internet and implement e-business transactions. In addition, Web services can be used to outweigh the disadvantages of the C/S model, such as the heavy maintenance required by the client software.

**Figure 2-2** Web service mode



Compared to the data service mode, a Web server and an application (APP) server are added in Web service mode and a three-layer architecture is available to the Web service mode. The service handling involves page configurations (implemented by the Web server), service handling (implemented by the APP server), and database deployment (implemented by the DB server and the storage system).

Compared to the data service model, the Web service model has the following features: Servers are deployed in the data center. Data is transmitted between the Web server and APP server and between the APP server and the DB server.

Web services have the following requirements on networks:

- Network bandwidth  
Traffic can be generated by the data requests and responses between a client and a server or servers, and can also be generated by the data requests and responses between servers. Web service traffic volume changes frequently, and increases due to traffic volume generated between servers. Traffic interaction between servers must be considered in a network design to avoid a traffic bottleneck on the network.

- Network reliability  
Web service model consists of three layers: Web, APP, and DB server layers. As traffic interaction increases, a network is required to be ever more reliable. The network recovery time, however, has not increases, making the reliability requirement of Web services very similar to that of data services.
- Network security  
There are hop-by-hop network channels between the Web, APP, and DB servers. This may open the network to hop-by-hop attacks. Firewalls must be deployed in the network channels to block illegal access from attackers.

## 2.3.2 Network Design

Large data centers are deployed in hierarchical mode. Small or medium data centers are deployed in flat mode. Traffic volume on a hierarchical network is planned based on the traffic volume in each layer. Traffic volume on a flat network is overlaid on one server and planned based on the total traffic volume.

Traffic volume between a client and a data center is much less than that in the data center. When designing a hierarchical network, the bandwidth must be properly calculated based on the traffic models to avoid node congestion. When designing a flat network, in addition to the traffic volume in each layer, the server must provide sufficient bandwidth where traffic volume is overlaid to meet the Web application requirements and avoid traffic congestion on the server.

## 2.3.3 System Reliability

An increased network reliability is required due to increasing traffic interaction; however, the requirement for overall recovery time is not increased. For details, see section [2.2.3 "System Reliability."](#)

## 2.3.4 System Security

System security is designed to implement:

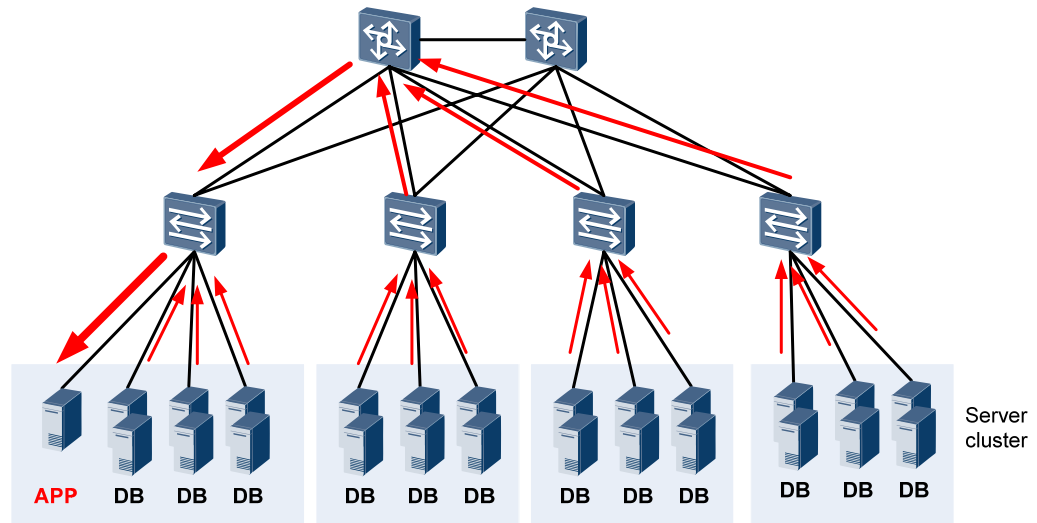
- Data security: Deploy a Web and an APP server to isolate the client from the DB server.
- Attack defense: Deploy network security devices to avoid hop-by-hop attacks from attackers.

# 2.4 Computing Service Design

## 2.4.1 Overview

Computing services, such as 3D shading, medical research, gene analysis, and Web search services, call for high-performance systems. In a typical computing service model, a lot of servers work with each other as a cluster to complete a computing task. In a computing service, the interaction among servers generates high traffic volume, as shown in [Figure 2-3](#).

Figure 2-3 Computing service traffic volume



Computing services require the following:

- High network bandwidth  
Services are implemented between any two servers. Therefore, high network bandwidth must be ensured to avoid traffic congestion when you plan the network.
- High network reliability  
A good scheduling mechanism is used by the APP server to distribute services. If a scheduling mechanism is not in place, all the handling results from the DB server are transmitted to the APP server in a short time frame, causing the data traffic to exceed the network interface bandwidth of the APP server. Therefore, network devices need to provide a large-capacity cache to prevent packet loss.

## 2.4.2 Network Design

In a computing service, most traffic is generated by the interaction among servers. The network device that connects to a server must provide high-speed forwarding performance and capacity to avoid traffic bottleneck on the network device. Select device based on the service traffic model. 10 Gbit/s ports must be available to meet requirements of servers with 10 Gbit/s ports.

## 2.4.3 System Reliability

If all the handling results are transmitted from the DB server to the APP server in a short period of time, the APP server will not be able to process all services and packets will be lost. To avoid this, the APP server frequently interacts services with the DB server, which prolongs the duration of service handling. The network device that connects to a server must provide a large-capacity cache and evaluate cached traffic based on service traffic model to ensure that the traffic cached on the board meets service requirements. This prevents packet loss.

# 3 Physical Network Design

---

## 3.1 Overview

The physical network design in a data center must meet the following requirements:

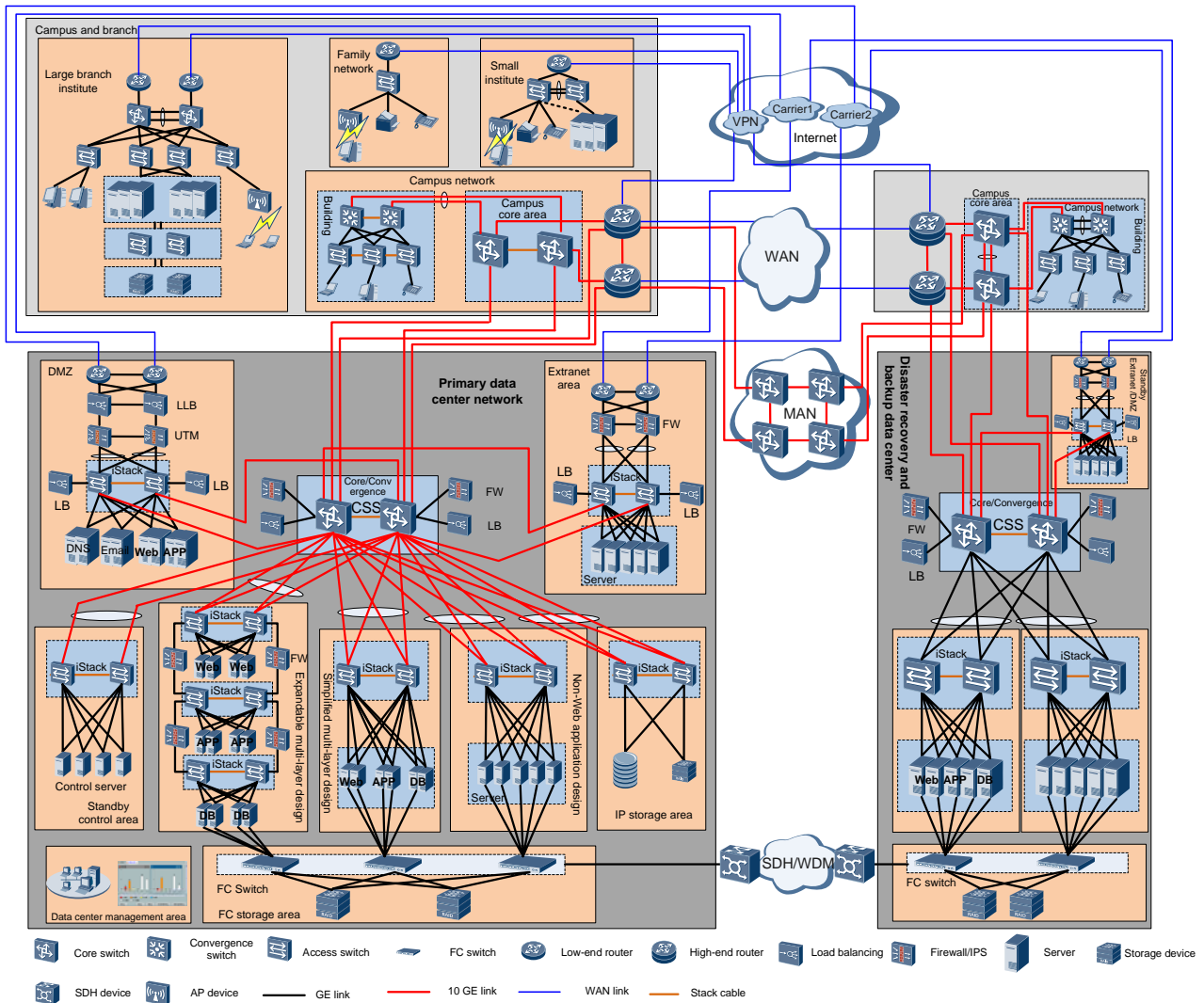
- Modularization
- High reliability
- Secure isolation
- Manageability
- Maintainability

The data center is allocated into the following areas:

- Core network area
- Server area
- Internet area
- Network management area
- Storage area

Figure 3-1 shows the typical networking modes of data centers.

Figure 3-1 Typical networking of data centers



## 3.2 Topology Design

### 3.2.1 Overview

The typical networking diagrams of data centers show that a data center adopts a layered structure, in which each functional module has its own responsibility. Such network structure clearly shows the end-to-end service process and facilitates network maintenance for data center.

## 3.2.2 Considerations

### Architecture Design at the Core Layer

In most cases, a data center is deployed in two- or three-layer architecture. The two-layer architecture consists of the access layer and aggregation or core layer. A two-layer network is deployed in flat mode. The three-layer architecture consists of the access layer, aggregation layer, and core layer.

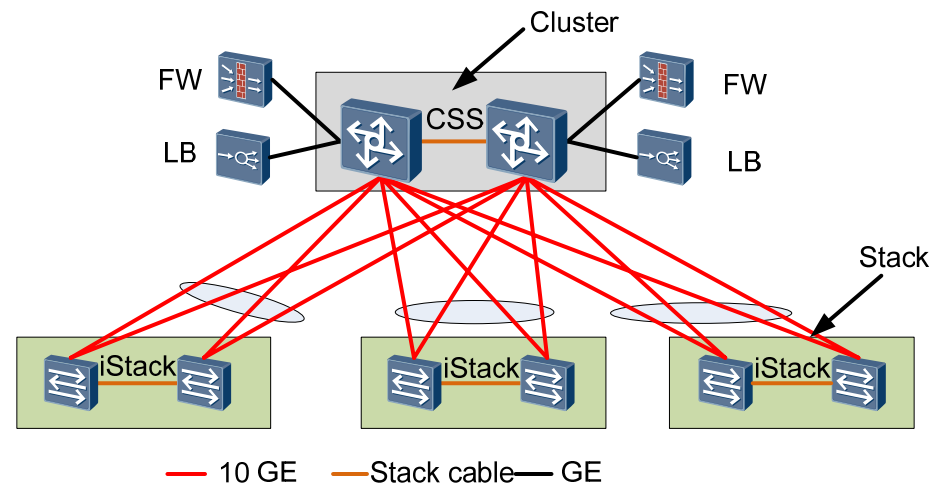
In the two-layer mode, network complexity is reduced, network topology is simplified, and forwarding efficiency is improved. Therefore, Huawei recommends a two-layer architecture for the network. In this architecture, the cluster+stack networking solution can be used to implement link redundancy:

- Access switches are deployed into a VLAN to ensure L2 forwarding.
- A core or aggregation switch is deployed with a load balancer (LB) and firewall (FW) in bypass mode.
- Configure the server gateway that provides load balancing services on the LB and configure the server gateway that does not provide the services on the FW.

Alternatively, you can adopt the L3-to-edge solution to prevent loops. In this solution, routing protocols are configured at the access layer and aggregation/core layer, and services are forwarded based on IP routes.

#### a. Cluster+stack networking solution

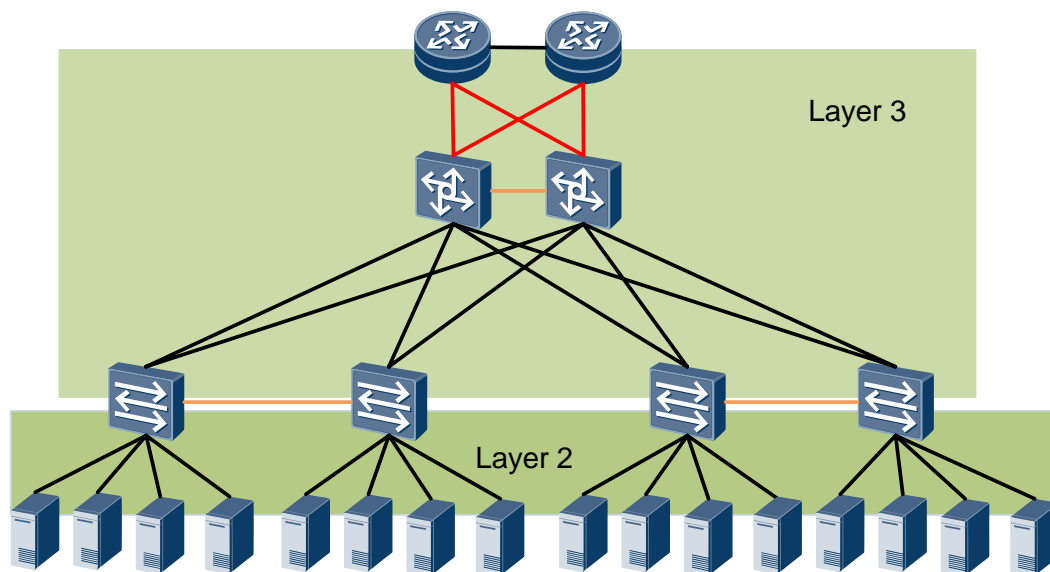
**Figure 3-2** Cluster+stack networking solution



- S93 devices are deployed at the aggregation layer. Stack cables are used to connect two S93 devices so that the two devices are virtualized as one device.
- S57 devices are deployed at the access layer. Stack cables are used to connect two S57 devices so that the two devices are virtualized as one device.
- The four inter-subrack links between the access switch and the aggregation/core switch are bound into an Eth-Trunk group. The network architecture is displayed in tree mode. This prevents loops on networks.

b. L3-to-edge solution

**Figure 3-3** L3-to-edge networking



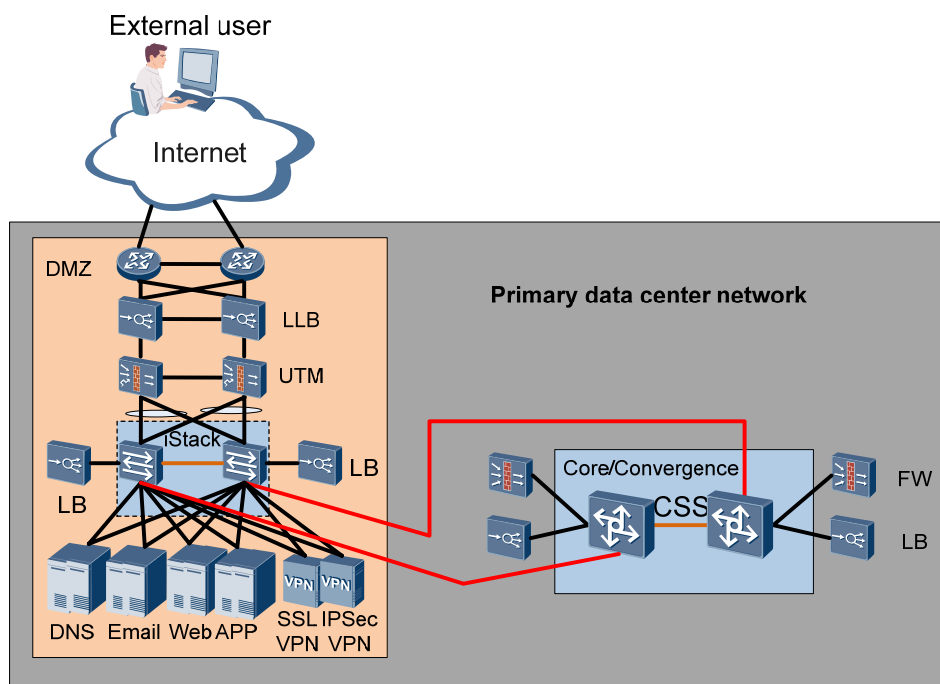
In this solution, the server gateways are moved from the core/aggregation layer to the access layer. L3 routes are deployed at the layers above the access layer. This avoids the creation of a two-layer loop between the access layer and aggregation layer. Gateways must be configured in the network segment where the servers are deployed and IP-based forwarding control must also be configured for each access switch. The configuration workload is heavy.

When servers that support the same service system connect to different switches, the IP addresses of the servers are nonconsecutive. This complicates server expansion.

## Architecture Design for the Data Center Accessed by External Users

External users can access only the demilitarized zone (DMZ) of a data center to ensure the security of the data center. Huawei recommends the following network.

**Figure 3-4** The data center network available to external users



Unified threat management (UTM) devices with the intrusion protection system (IPS) and firewall must be configured on the egress of the Internet.

- The IPS prevents network attacks and malicious behaviors, such as viruses, worms, Trojan horse, spyware, and DDoS, in real time by means of L7 analysis and detection. In addition, the IPS can effectively manage non-critical services, such as point-to-point (P2P) and instant message (IM) services, on the network. In this way, the IPS provides comprehensive protection for network applications, infrastructures, and performance.
- The firewall constrains communications between two or more networks by using access security policies.

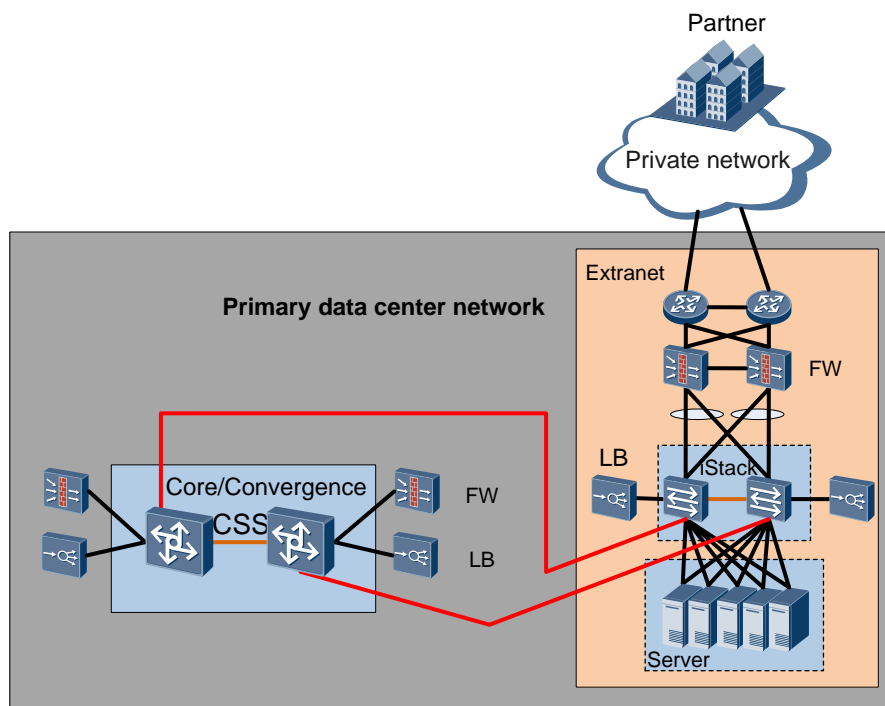
The UTM is a server gateway in the DMZ. Internet users can access only the DMZ of the data center because the access security policies are deployed on the firewall. The firewall deployed in the core zone of the data center connects to the core switch in dual-homing mode. Security policies are configured on the firewall so that internal users of the data center can access the DMZ.

## Architecture Design for the Data Center Accessed by Partners

Partners can access the data center only by using private lines to ensure the security of the data center. Huawei recommends the following network architecture.



**Figure 3-5** Networking for the data center accessed by partners

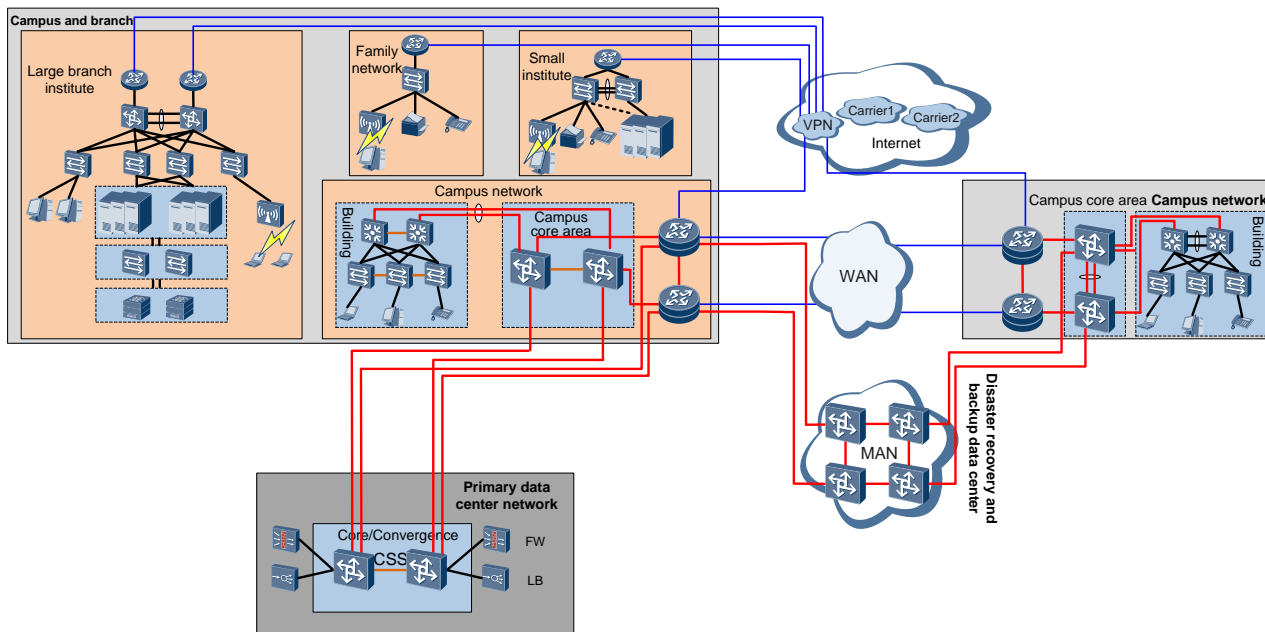


A firewall device connects to a switch at the core layer of the data center in dual-homing mode. The firewall is used to control access permissions of partners by deploying access security policies. Partners can access specified servers in the Extranet area in the data center.

## Architecture Design for the Data Center Accessed by the Headquarters and Branches

Users from the headquarters and branches access the data center using the MPLS VPN technology, which isolates the services on the carrier network from other services, ensuring high reliability of services. In addition, the MPLS VPN technology can flexibly control the access of VPN services between users to facilitate service interconnection and isolation.

**Figure 3-6** Networking for the data center accessed by the headquarters and branches

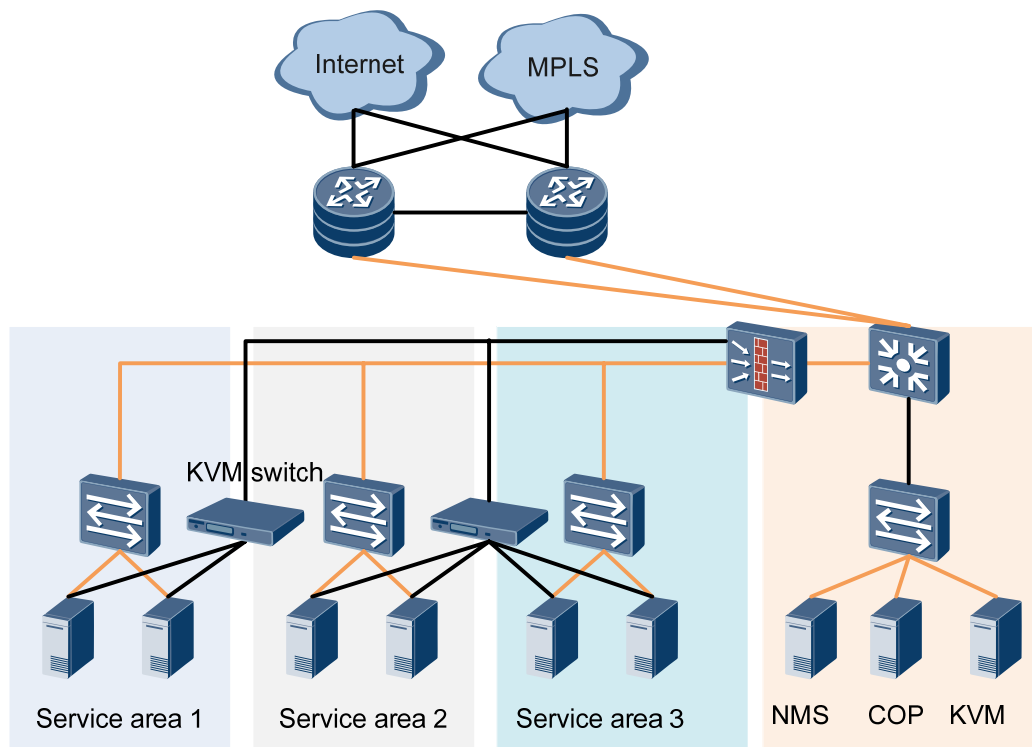


The firewall deployed above the server layer controls the security access of VPN users and monitors the running status of the network and system using the intrusion detection system (IDS) function. In this case, attack risks, behaviors, or results are detected to ensure privacy, integrity, and availability of network system resources.

### Architecture Design for the Data Center Accessed by the O&M Network

An operation and maintenance (O&M) network implements routine maintenance and collects device and service information on servers, storage devices, and network devices. An O&M network must be connected to the core switch of the data center. Huawei recommends the following network architecture.

**Figure 3-7** Networking for the data center accessed by the O&M network



The administration zone connects to the egress router of the data center in dual-homing mode. In this networking mode, this zone can access servers, storage devices, and network devices, and the routes are reachable.

### 3.2.3 Design Principles

#### Principles for the Design of a Reliable Topology in the Data Center

- **Link reliability**  
Link reliability is improved by dual-uplink or link bundling. Dual-uplink links connect to different network devices. This prevents the interruption of data center services due to the failure of a single network device. Link bundling is used in the network environment where network devices are deployed in the same hierarchy and a dual-uplink network cannot be deployed. With this technology, multiple links are bundled to improve the reliability of service forwarding.
- **Logical device reliability**  
With the cluster and stack technologies, multiple physical devices can be virtualized to one device so that a loop network is simplified to a tree network. In this case, complex ring network protocols need not be deployed, and physical device failures do not affect service forwarding. This significantly improves network reliability. The number of logical devices to be managed is reduced and the network topology is simplified. Therefore, the network is more easily managed and maintained.

## Principles for the Design of Topology Scalability in the Data Center

As data center services are added and the number of servers increase, network devices must be increased. The scalability of the logical topology of the data center must be considered to ensure that data center services are not affected in smooth network upgrade.

- **Link scalability**  
In the design phase, certain physical ports must be reserved to ensure that the network can be upgraded by increasing physical links when the link traffic incurs a bottleneck. In this case, the link bandwidth is increased without affecting running data center services. Certain slots are reserved on the S9300 switch to facilitate future expansion by increasing service cards. In addition, the type of links interconnecting devices can be upgraded, such as from GE to 10 GE, or from 10 GE to 40 GE or 100 GE, to increase link bandwidth.
- **Device scalability**  
The stack technology can be used to stack up to nine physical devices. When the reserved physical ports cannot meet the requirements for data center development, the network can be upgraded by increasing physical devices. With the stack technology, the network topology is not affected by new physical devices. This ensures the normal running of data center services.

### 3.3 Node Devices

An increasing number of services and data are collected in the data center to meet enterprises' requirements. In this case, the following high performance and reliability for network devices must be met:

**High performance:** high-capacity, high-density, and modularized L2 to L4 line rate forwarding performance.

**High reliability:** perfect QoS assurance, effective security management mechanisms, and carrier-grade high reliability. The following lists the solutions from Huawei for networks in different hierarchies:

- **Core/aggregation switches**  
Huawei recommends that Quidway® S9300 carrier-grade campus switches are used at the core/aggregation layers in the data center. Through the ports on S9300 switches, services can be forwarded at line rate, such as at ACL line rate. These services include those forwarded in IPv4, MPLS, or L2 forwarding mode. S9300 switches support 2 Tbit/s switching capacity and multiple types of high-density cards to meet large-capacity and high-density requirements for devices at the core and aggregation layers. In this case, higher and higher bandwidth requirements of customers can be met to help customers reduce their investment and maximize return on investment.
- **Access switches**  
Huawei recommends that Quidway® S5700 GE switches and Quidway® S6700 10 GE switches are used at the access layer. Ports on an S5700 switch feature large capacity and high density and a 10 GE uplink rate. These switches can meet customers' service requirements. S6700 switches provide up to 24 or 48 full-line-rate 10 GE ports to ensure the high-density access of 10 GE servers. In addition, S5700 and S6700 switches support various service features, complete security control policies, and various QoS features to ensure the scalability, reliability, manageability, and security of the data center.
- **Egress area**

Huawei recommends that the NE40E router is used as the egress router. The NE40E uses a platform with 400 GB capacity to meet requirements in future 10 years. The NE40E provides end-to-end reliability solutions to prevent interruption of data center services.

Huawei recommends that an Eudemon device is used as a firewall device in the data center. The Eudemon device uses the advanced ATCA+multi-core+distributed architecture to ensure high scalability. With more than 100 GB capacity, the firewall has the highest performance and reliability to ensure the security of top application networks in a large data center.

- Maintenance area

Huawei recommends that the eSight NMS is used. eSight is a new management system for enterprise network management. It provides unified management and intelligent association functions for enterprise resources, services, and users. eSight provides the following functions:

- Unified management on IT&IP devices and third-party devices.
- Intelligent analysis of network traffic and access authentication roles.
- Automatic adjustment of network control policies to ensure network security.
- A flexible and open platform and a custom intelligent management system for the enterprise.

Table 3-1 describes the device selection list.

**Table 3-1** Device selection list

Network Layer	Device	Function	Requirement on Device (Such as Performance and Stack)	Adopted Device Model
Access layer	GE access switch	Connects to a GE server.	Services can be forwarded at line rate through all ports. The switches can be stacked.	S5700
	10 GE access switch	Connects to the server with the 10 GE capacity.	Services can be forwarded at line rate through all ports. The switches can be stacked.	S6700
Aggregation layer	Aggregation switch	Is converged with a device at the access layer.	Services can be forwarded at line rate through all ports. The switches can be clustered.	S9300
Core layer	Core switch	Is aggregated with a device at the aggregation layer.	Services can be forwarded at line rate through all ports. The switches can be clustered.	S9300
Egress area	Egress router	Provides the routing function.	Services can be forwarded at line rate through all ports.	NE40E
	Firewall	Provides the security function.	Security access and defense are implemented.	Eudemon
Maintenance area	NMS	Manages and maintains NEs.	Unified network management is implemented.	eSight

## 3.4 Link Selection and Design

### 3.4.1 Internal Link Design

The network architecture of a data center consists of the access, aggregation, and core layers. Aggregation layer and core layer can be integrated into one layer as required.

#### Links at the Access Layer

Select GE or 10 GE downlinks based on physical ports of servers. Uplinks connect to aggregation and core layers in dual-homing mode to improve link reliability. You can adopt 10 GE links or the link bundling technology to increase the link bandwidth.

#### Links at the Aggregation and Core Layers

The 10 GE links or link bundling technology can be used to interconnect downlinks to an access switch, increasing the bandwidth and ensuring high reliability. Use Ethernet trunk links to interconnect the aggregation layer with the core layer to ensure link reliability. If the data center has three layers, the aggregation layer must connect to the switches at the core layer in dual-homing mode and 10 GE links are preferred. This improves the reliability and increases the link bandwidth.

### 3.4.2 Egress and Ingress Link Design

Egress and ingress links of the data center are used to interconnect with:

- Enterprise headquarters and branches
- Partners and external customers
- Internet users
- Disaster recovery data center

The following sections describe the design in the preceding scenarios:

#### Links Between Headquarters and Branches

Over 10 GE links or with the link bundling technology, a core switch is interconnected with an egress router in dual-uplink mode, increasing the link bandwidth and improving the reliability. The egress router connects to the WAN over WAN links, or to the MAN over 10 GE links. The VPN technology is used to interconnect enterprise headquarters and branches with the data center.

#### Links Between Partners and External Customers

Over 10 GE links or with the link bundling technology, an Extranet area is interconnected with an aggregation or core switch in the data center. The egress router of the Extranet area connects to carrier networks over WAN links to ensure high reliability.

#### Links Among Internet Users

Over 10 GE links or with the link bundling technology, a DMZ is interconnected with the aggregation or core switch in the data center. The egress router of the DMZ connects to carrier networks over WAN links to ensure high reliability. When an Internet user accesses services in the DMZ, the intelligent domain name system (DNS) can be used to distinguish users and

resolve a domain name into an IP address of a carrier. If the user is from China Netcom, the policy resolution server in the DNS resolves the Netcom IP address corresponding to the domain name for the user.

## Links for Disaster Recovery Data Center

Leased private lines or VPN lines are used for remote cross-domain disaster recovery links. It is recommended that leased point-to-point lines are used for well-funded enterprises. The bandwidth of the leased links is determined based on the backup services. Fiber channel (FC)-based disaster recovery links in a city connect to dense wavelength division multiplexing (DWDM) transmission equipment to implement service data backup.

# 3.5 Planning of Device and Interface Names

## 3.5.1 Device Naming

Device naming criteria are critical to the application of the NMS when devices in the data center increase.

The following naming rule is recommended: service information + position information + device supplier + device model + number. The following describes detailed information:

- Service information: Services are classified by service type in the data center, such as data service, Web service, and computing service, so that the NMS collects data based on service information.
- Position information: specifies physical positions for network devices, such as floor number.
- Device supplier: provides information about device suppliers.
- Device model: specifies network device models, for example, the name of a router starts with R and the name of a switch starts with S.
- Number: uses A, B, C..., or 01, 02, 03....

Devices can be named as required during network planning.

## 3.5.2 Naming of Links and Interfaces

The name of a link or an interface needs to contain the NE device that the port belongs to, port type and bandwidth, peer NE device, and physical slot of the card that the peer port belongs to. Other information is optional. For example, To-[S9300-1]GE-1/0/0 is the name of the interface GE1/0/0 through which the access switch S5700 connects to the core switch S9300-1.

# 4 Design of Logical End-to-End Features

---

## 4.1 Overview

Logical network design ensures the connectivity of physical network protocols as well as QoS and network reliability. The components of logical network design are described as follows:

- VLAN planning: constraints, ID allocation, and VLAN layers
- IP address planning: service IP addresses, management IP addresses, and active and standby DNSs
- Route design: open shortest path first (OSPF) design, external BGP (EBGP), and AS allocation
- MPLS/VPN: allocates security areas and the sets routes on server and user VPNs
- Reliability design: reliability of the access layer, aggregation/core layer, and the service layer
- QoS design: burst traffic and size of the buffer

## 4.2 VLAN Design

### 4.2.1 Overview

Logically a local area network (LAN) is divided into multiple subnets and each subnet is a broadcast domain, that is, a virtual local area network (VLAN). Devices in a LAN are allocated into network segments in logical instead of physical mode and each network segment is a VLAN. In this way, broadcast domains are isolated in a LAN.

Interconnecting devices are divided into a VLAN, and those that do not interconnect are divided into different VLANs. In this case, broadcast domains are isolated to reduce broadcast storms and improve information security. With the VLAN technology, a network failure can be restricted to a local area, protecting the overall network. This keeps the network robust.

### 4.2.2 Considerations

#### Single-Layer or Double-Layer VLAN

VLAN is used to isolate services and users. L2 network covers the core layer, aggregation layer, and access layer of a data center bearer network. The core and aggregation layer can be



integrated into one layer. The core layer is configured on the L2 network as a server gateway. It exchanges routing information with the access routers by using IP or MPLS technology. Services provided by the access servers in the data center are allocated to a single-layer VLAN on the L2 network.

## Allocation of VLANs in Other Areas

Server areas must be allocated to VLANs based on service or server types. In addition, external access layers are allocated to VLANs. The external access layer implements internal network interconnection, interconnection with customers, Internet interconnection, interconnection with the disaster recovery center, and management and maintenance area. Huawei recommends that for L3 interconnection links, multiple physical ports are virtualized to a logical port, tags are sealed on a VLAN, and then IP packets are forwarded.

## Constraints

A maximum of 4094 different single-layer VLANs are allocated based on services and areas to ensure that certain VLANs can be reserved for future expansion.

### 4.2.3 Design Principles

- Distinguish service VLANs, management VLANs, and interconnection VLANs.
- Allocate VLANs based on service areas.
- In a service area, allocate VLANs based on service types, such as Web, APP, and DB services.
- Allocate VLANs consecutively to ensure the proper use of VLAN resources.
- Reserve certain VLANs for future expansion.

### 4.2.4 Design Elements

Allocate network segments for VLANs based on the following areas:

- Core area: 100–199 network segments.
- Server area: 200–999 network segments. 1000–1999 network segments are reserved.
- Access network: 2000–2999 network segments.
- Management network: 3000–3999 network segments.

Figure 4-1 shows the allocation schematic drawing.

Figure 4-1 VLAN planning

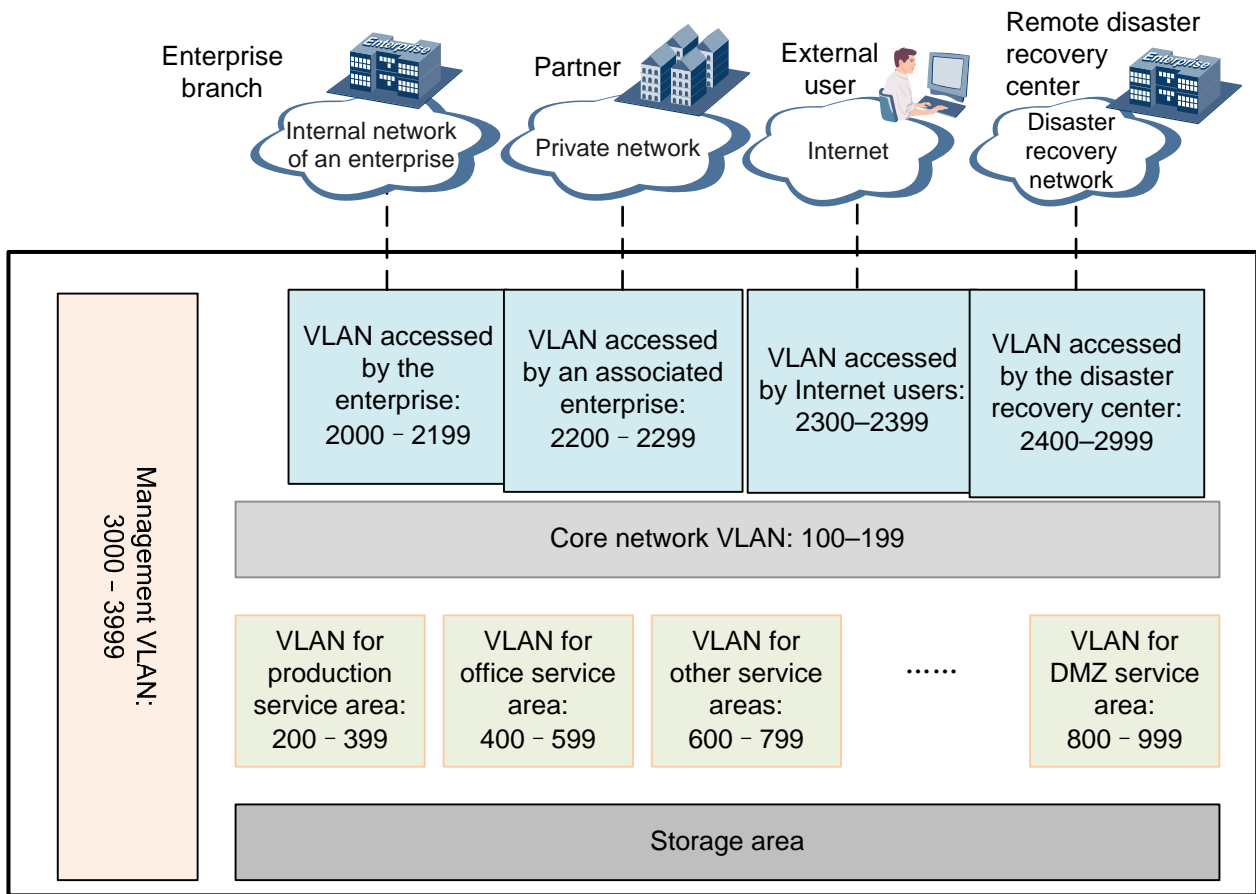


Figure 4-2 shows the architecture of the access-layer VLAN in the service area of the data center.

Figure 4-2 Architecture of the access-layer VLAN in the service area of the data center

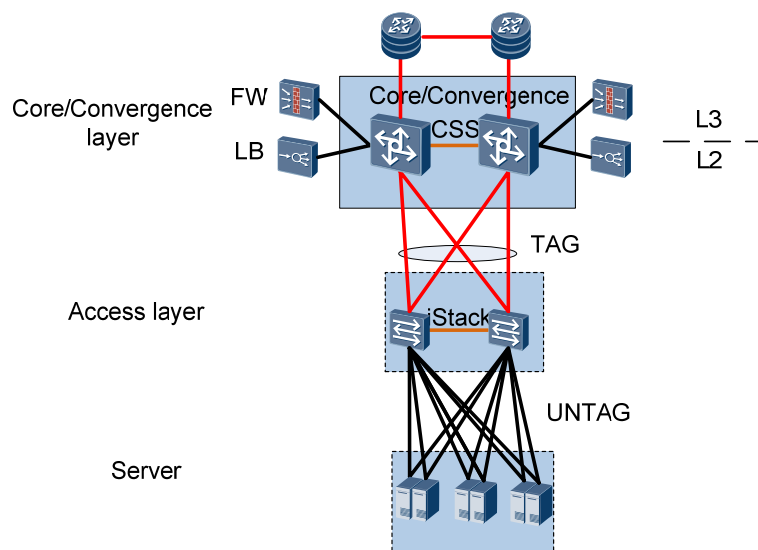


Table 4-1 describes the functions of nodes in the VLAN architecture.

**Table 4-1** Function of nodes in the VLAN architecture

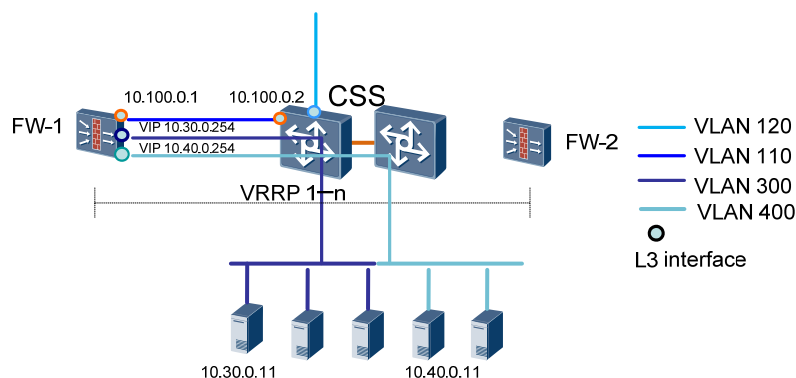
Node	VLAN Configuration	Description
Access switch	Multiple VLANs are configured. The downlink ports of servers are added to the VLANs in access mode. Uplink ports are added to the VLANs in trunk mode.	VLANs are allocated based on services and areas. Servers are connected to different VLANs to implement L2 broadcast isolation.
Core/aggregation switch	A VLAN is configured. Server traffic is forwarded to the FWs or the LBs based on the VLAN at layer 2.	Gateways are configured on the LBs or FWs to implement server load balancing.

## 4.2.5 L2/L3 Allocation Design

In the cluster+stack scenario, the stacked switches forward packets at layer 2. Server gateways are configured on the FWs or LBs. When the servers communicate with clients, the FWs implement traffic filter and protection. The communication between servers in the same network segment is not implemented through the FWs, and the communication between those in different network segments is implemented through the FWs.

### Server Gateways on the FWs

**Figure 4-3** Server gateways on the FWs



- **Deployment**  
The core/aggregation switch and FW-1 advertise routes to each other by configuring OSPF. VRRP is configured between FW-1 and FW-2 to ensure that high availability (HA) and load balancing are implemented between FW-1 and FW-2. Configure VIP 10.30.0.254 as the server gateway IP address.  
**Figure 4-3** shows the connection relationship between FW-1 and the core, aggregation, and access switches. FW-2 configurations are similar to FW-1 configurations.
- **Flow path**

Client-server (C-S)

1. Data flows are transmitted from the core/aggregation switch to the access switch through VLAN120.
2. The core/aggregation switch forwards the data to FW-1 through VLAN110 based on the destination IP address (10.30.0.11).
3. FW-1 queries the route and forwards packets to the server in VLAN300.

----End

Server-client (S-C)

1. A server returns data flows to the FW-1.
2. FW-1 queries the route and forwards packets at layer 3 to the core/aggregation switch through VLAN110.
3. The core/aggregation switch queries the route and forwards data flows through VLAN120.

Server-server (S-S) in the same network segment

4. The access switch forwards data flows directly at layer 2.

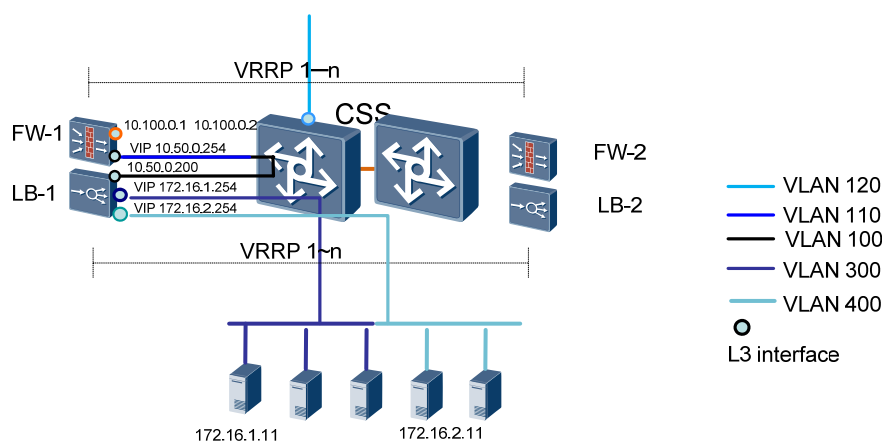
S-S crossing network segments

5. Data flows are transmitted from the access switch and core/aggregation switch to FW-1 through VLAN300.
6. FW-1 filters data flows and forwards filtered flows at layer 3 to the server through VLAN400.

----End

## Server Gateways on the LBs

Figure 4-4 Server gateways on the LBs



- Deployment

The core/aggregation switch and FW-1 advertise routes to each other by configuring OSPF. Configure VRRP between FW-1 and FW-2. Set the gateway IP address of the

LB-1 next hop to VIP 10.50.0.254. Configure a default route for LB-1, that is, the next hop of LB-1 points to the VIP 10.50.0.254 of FW-1. LB-1 uses the private address 172.16/16 for the servers and uses the server VIP address 10.50.0.11 for external users or devices. Configure VRRP between LB-1 and LB-2 and set the IP address of the server gateway to VIP 172.16.1.254.

Figure 4-4 shows the connection relationship between FW-1, LB-1 and the core and aggregation switches. FW-2 configurations are similar to FW-1 configurations and LB-2 configurations are similar to LB-1 configurations.

- Flow path

C-S

1. Packets are transmitted from the core/aggregation switch to the access switch through VLAN 120.
2. The core/aggregation switch forwards the packets to FW-1 through VLAN 110 according to the destination IP address (10.50.0.11).
3. FW-1 searches routes to forward the packets to LB-1 through VLAN 100.
4. LB-1 terminates the packets as an agent and replaces the source IP address (172.16.1.254) and the destination IP address (172.16.1.11). In addition, LB-1 allocates the packets to a server based on a load balancing algorithm.

----End

S-C

1. The server returns data flows to LB-1 (gateway).
2. LB-1 terminates packets as an agent. After replacing the source IP address (10.50.0.11) of the data flows with the destination IP address (the IP address of a client), LB-1 forwards data packets to FW-1 through VLAN 100.
3. FW-1 searches routes to forward the packets to the core/aggregation switch through VLAN 110.
4. The core/aggregation switch searches routes to forward the packets out through VLAN 120.

S-S in the same network segment

5. The access switch forwards data flows directly at layer 2.

S-S crossing network segments

6. Packets are transmitted from the access switch and core/aggregation switch to LB-1 through VLAN 300.
7. LB-1 terminates packets as an agent. After replacing the source IP address and destination IP address of the packets, LB-1 forwards packets to FW-1.
8. If FW-1 finds that the route to the destination server is directly connected, FW-1 forwards the packets to the destination server.
9. If FW-1 queries that the next hop of route points to LB-1, FW-1 forwards packets to LB-1. LB-1 terminates packets as an agent. After replacing the source and destination IP address of the packets, LB-1 forwards the packets to the destination servers.

----End

## 4.3 IP Service and Application Design

### 4.3.1 Overview

In most cases, an IP address (network ID+host ID) is used to specify a network device, such as an interface. The IP addresses of devices ensures network interconnection and applications; IP address planning is the base of implementing network functions.

Figure 4-5 shows the five types of IP addresses.

**Figure 4-5** Types of IP addresses

A	0	Network(7 bits)			Host(24 bits)			
B	1	0	Network(14 bits)			Host(16 bits)		
C	1	1	0	Network(21 bits)			Host(8 bits)	
D	1	1	1	0	Multicast address			
E	1	1	1	1	0	Reserved		

In addition to IP addresses, TCP/IP provides a special naming mechanism for hosts in character string mode, that is, DNS. As a hierarchical naming method, the DNS specifies a name for a network device and sets a domain name resolution server on the network to create the mapping between a domain name and an IP address. This enables users to use easy-to-remember and meaningful domain names instead of complex IP addresses.

### 4.3.2 IP Address Planning and Design

#### Considerations

While planning the IP addresses of network segments, network segment routes are aggregated on a core switch or router. In this case, the route quantity and maintenance cost are reduced when routes are distributed. For example, to allocate a class C address 192.168.1.0/24 to the data center, you can use the variable length subnet mask (VLSM) to allocate the address into four network segments: 192.168.1.0/26, 192.168.1.64/26, 192.168.1.128/26, and 192.168.1.192/26. Allocate the four network segments to the service area and each network segment contains up to 62 host IP addresses. Routes in the four network segments can be aggregated on the core server and only one network segment route with a class C IP address is announced.

During IP address assignment, consecutive IP addresses are allocated to each service area and certain IP addresses are reserved for future network expansion. In a service area, consecutive IP addresses are allocated to the servers providing the same services and functions.

An enterprise plans the IP addresses of the internal networks in the data center. When the servers that provide services for external users use private IP addresses, a network address translation (NAT) device must be used on the egress router to translate private IP addresses to public ones. If address resources are adequate, public network IP addresses can be directly allocated to the servers.

## Design Principles

- Uniqueness  
The same IP address cannot be used by two hosts in an IP network even when the MPLS/VPN technology supports address overlap.
- Continuity  
Continuous IP addresses facilitate path overlap in a hierarchical network. This reduces entries in a routing table and improves routing algorithm efficiency.
- Scalability  
Reserve certain addresses in each hierarchy to ensure the continuity of overlapped addresses during network expansion.
- Meaningfulness  
Ensure that an IP address has a meaning, that is, an IP address enables you to determine the device that it belongs to.

## Design Elements

- IP address mode  
Currently, the Internet adopts the IPv4 protocol in most cases; however, public IPv4 addresses are in shortage and will be replaced by the IPv6 addresses. Therefore, the compatibility with the IPv6 protocol must be considered during planning of IP addresses so that the transition and upgrade to IPv6 is prepared.
- Two types of IPv4 addresses for services
  - Private IPv4 address  
Huawei recommends you adopt a class B private IP address defined in IETF RFC1918 and use L3VPN/VLAN to isolate services using private IP addresses. Theoretically, each type of service can exclusively use a private address space; however, the address segments of services may be re-allocated to ensure efficient management.  
Private IP addresses can be allocated to the services that do not require external resources.  
If a large number of IP addresses are required, private IP addresses must be allocated to the services which need to access public resources. In this case, a NAT device is used to translate the private IP addresses to public ones.
  - Public IPv4 address  
If public IP addresses of service providers (SPs) are allocated to services, the services that adopt public IP addresses must be restricted because these address resources are in shortage. Do not assign public IP addresses for services that consume many address resources.
- Service IP addresses  
[Figure 4-6](#) describes the specifications for IP address design.

**Figure 4-6** Specifications for IP address design

Value specifications of IP address					
IP Address Structure	The 1 <sup>st</sup> Byte	The 2 <sup>nd</sup> Byte	The 3 <sup>rd</sup> Byte		The 4 <sup>th</sup> Byte
	8bits	8bits	4bits	4bits	8bits
Identifier	Identifier 1	Identifier 2	Identifier 3	Identifier 4	Identifier 5
Meaning	The number 10 indicates a class A private network.	Grade 1 institute	Grade 2 institute	Service area	Host bits

The following defines and allocates server IP addresses:

- Identifier 1: 8 bits, indicates a private IP address, such as 10.X.X.X/8.
- Identifier 2: 8 bits, indicates a grade 1 institute. Each grade 1 institute can apply for multiple class B IP address segments.
- Identifier 3: 4 bits, indicates a grade 2 institute. The active data center adopts 0000 to 0111 and the standby data center adopts 1000 to 1111.
- Identifier 4: 4 bits, indicates a service area in the data center, such as the production service area and the office area.
- Identifier 5: 8 bits, indicates a host or a server address.



**NOTE**

IP addresses are planned to meet customers' requirements. The server quantity and service types are customized as required. For example, the IP address of the production service is 10.100.0.0/17.

- Device management IP address
  - Assign one or more class C IP addresses for network management.
  - To facilitate management, assign consecutive IP addresses for L3 device management and consecutive IP addresses for L2 device management.
- Device interconnection IP address
  - An interconnection IP address is in the x.x.x.x/30 format. Device IP addresses are interconnected from bottom to top. Assign a small IP address to a device in the higher hierarchy. For the services in the same hierarchy, assign a smaller IP address to a device that has small loopback. For example, the address of a port that connects to the downlink is even, such as 10.1.1.2/30. The address of a port that connects to the uplink is odd: such as 10.1.1.1/30
  - Reserve adequate spaces for future upgrades and expansion.
- IP address of the server gateway
  - 1–9 specifies the IP address of a gateway or a gateway VRRP.
  - 10–254 specifies a service IP address, such as a server IP address.

### 4.3.3 DNS Design

#### Considerations

A user accesses a server using a domain name instead of the IP address because a domain name is easily understood and remembered, for example, the Web address of Baidu is [www.baidu.com](http://www.baidu.com). Deploy servers in the DMZ of the data center to provide FTP and Web services. The DNS holds the mapping between the domain names and the corresponding IP addresses and decouples the application system and the servers.



The DNS is very important for the data center. If the DNS failed, users cannot access the application servers, which severely affect the production of an enterprise. Therefore, multiple DNS servers must be deployed in the DNS. The following lists the roles of these DNS servers:

- **Master server**  
As the management server in the DNS, it can add, delete, and change the domain name. The changed information can be synchronized to the slave server. Generally, you can deploy only one master server.
- **Slave servers**  
These servers obtain domain name information from the master server, and provide DNS services as a cluster. They adopt hardware-based LBs to provide server cluster function. You can deploy two slave servers.
- **Cache servers**  
These servers cache the results of the DNS requests from internal users to speed up subsequent access. They are deployed on the slave servers.

## Design Principles

- **Easy-to-remember**  
The DNS holds the mapping between the domain names and the corresponding IP addresses. The domain names must be simple and easy-to-remember. The domain names must be closely related to services provided by the servers.
- **Hierarchical**  
The domain names must be hierarchically designed based on physical locations or logical service areas.
- **Intelligent**  
The DNS can intelligently distinguish carriers from broadband users, and map the domain names to the IP addresses of the carriers to speed up users' access.

## Design Elements

- **NAT mapping of the domain names on the FWs.**  
Implement NAT mapping for the Internet domain name. Map the virtual address of the slave servers to a public IP address and set the IP address to the access address for external Internet users.  
An internal user sends requests to the DNS, which then may send the requests to slave DNS1 and DNS2 servers. If the slave DNS1 server failed, all these requests are allocated to slave DNS2 server. If all slave DNS servers failed, the master DNS server handles these requests.

**Figure 4-7** NAT mapping of the domain names on the FWs

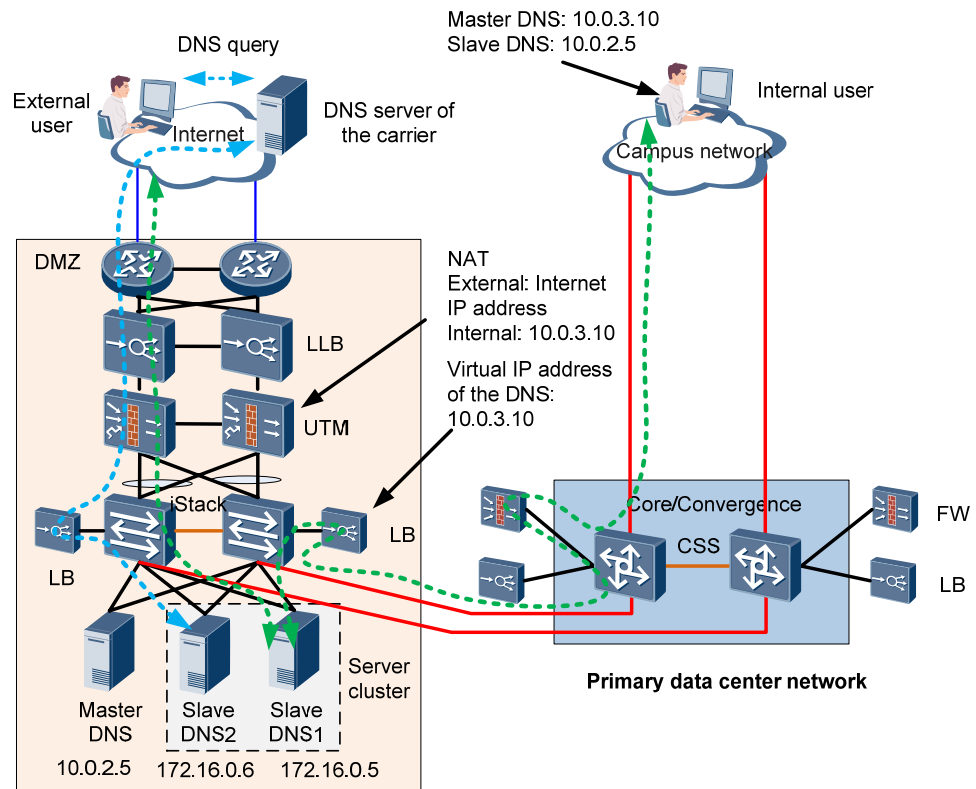


Table 4-2 describes the suggestions for network deployment.

**Table 4-2** Network deployment 1

DNS server type	<p>Deploy master and slave DNS servers. The following lists the deployment details:</p> <ul style="list-style-type: none"> <li>• Deploy a cluster of slave DNS servers to provide services for external Internet users. Cache servers can be deployed on these servers to speed up subsequent access.</li> <li>• Deploy a master DNS server and back up the server in the DMZ. The standby DNS servers providing services for internal users can be deployed in non-DMZ areas as standby ones.</li> <li>• Deploy a private IP address for the slave DNS servers. The IP address is displayed as a virtual one by the hardware-based LBs.</li> </ul>
Configuration of the client for an internal user	<p>Configure an active DNS server and standby DNS servers. The following lists the mapping between DNS servers on the client and those in the data center:</p> <ul style="list-style-type: none"> <li>• The active DNS server corresponds to the master DNS server.</li> <li>• The standby DNS server corresponds to the slave DNS server.</li> </ul>

Internal DNS service	The server uses an internal IP address and provides DNS services for internal users. Deploy FW protection and filtering instead of an IPS for internal users accessing the DNS servers.
External DNS service	Implement NAT forwarding on the FWs. This translates the IP addresses to public the IP addresses to provide DNS query services for external users. When Internet users access the DNS servers, the traffic passes through the IPS and the FWs in order. The IPS implements anti-attack protection and traffic cleaning, and the FWs implements protection and filtering.

- Deploy the intelligent DNS on the LBs to provide services for Internet users.

Internet users initiate DNS requests to the DNS server of a carrier. After receiving the requests, the intelligent DNS implements DNS resolution. The blue line in [Figure 4-8](#) indicates the preceding procedure.

The intelligent DNS distinguishes the users and resolves the domain names to corresponding IP addresses. If the user is from China Netcom, the DNS policy resolution servers resolve the Netcom IP address corresponding to the domain name for the user. If the user is from China Telecom, the DNS policy resolution servers resolve the Telecom IP address corresponding to the domain name for the user.

The intelligent DNS detects the quality of the links at the carrier side. When the links of a carrier are interrupted, the DNS returns the IP address of another carrier to ensure service continuity.

**Figure 4-8** Deploying the intelligent DNS to provide services for Internet users

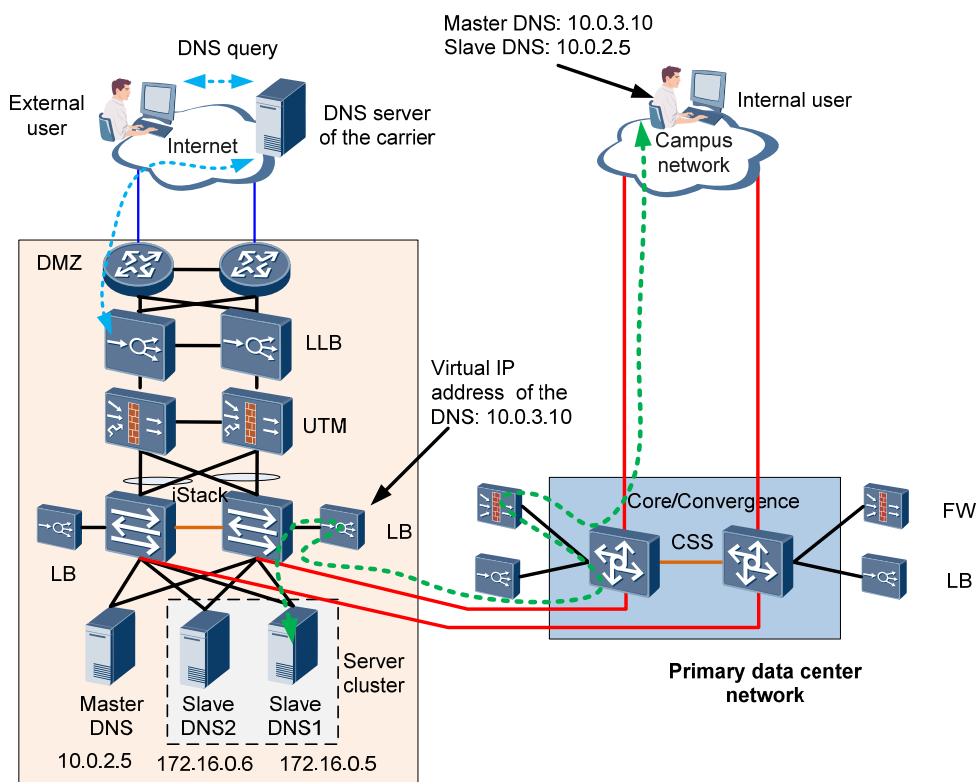


Table 4-3 describes the suggestions for network deployment.

**Table 4-3** Network deployment 2

DNS server type	Deploy master and slave DNS servers for internal users. Deploy the intelligent DNS on the link load balancer (LLBs) for Internet users.
Configuration of the client for an internal user	Configure an active DNS server and standby DNS servers. The following lists the mapping between DNS servers on the client and those in the data center: <ul style="list-style-type: none"> <li>The active DNS server corresponds to the master DNS server.</li> <li>The standby DNS server corresponds to the slave DNS server.</li> </ul>
Internal DNS service	The server uses an internal IP address and directly provides DNS services for internal users. Deploy FW protection and filtering instead of an IPS for internal users accessing the DNS servers.
External DNS service	Deploy the intelligent DNS on the LBs. the DNS provides the public IP addresses without NAT forwarding. The DNS services for Internet users are irrelevant to those for internal users.

## 4.4 Route Design

### 4.4.1 Overview

A Route is a relay used to forward data packets through the optimal path. A route provides two main functions: (1) Routing function: finds and selects the optimal path. (2) Forwarding function: forwards data packets to the destination address after selecting the path.

There are two types of routes: static route and dynamic route.

- Dynamic route includes BGP route and common IGP route, including RIP, ISIS, and OSPF route. Dynamic route completes route learning, selection, and maintenance by themselves. If the network typology changes after a dynamic route is configured, the dynamic route learns the change and modifies the router accordingly.
- Static route accurately controls paths for data packets forwarding; however, it cannot flexibly varies with network change due to static configurations.

Interior gateway protocol (IGP) is the base of an IP network. IGP is used to carry the Internet segment route and Loopback address route between routers. IGP design impacts on the configurations of critical parameters, such as network traffic model, convergence performance, reliability, and security.

### 4.4.2 OSPF Design

#### IGP Selection

ISIS and OSPF are the most mature and widely used IGP protocols. Although there are differences, the two protocols have little difference in function and performance.

In addition to technology considerations, the selection of the IGP protocol involves competitive marketing strategies. For example, Cisco wished to deploy ISIS to shield competitors. You can select one of the two protocols based on the following principles:

ISIS: used in IP bearer networks and for public routes. Most national large NSP bearer networks adopt the ISIS protocol.

OSPF: used in MANs and for private routes. Adopt OSPF dynamic route protocol in the data center to ensure network stability and rapid convergence of routes and to facilitate future management and maintenance.

#### Design Principles

- When a network is divided into areas, deploy all routes in Area0.
- When the network is allocated into multiple areas, the number of routes in each area must be considered.
- Generally, METRIC values of the two ends of a link need to be consistent.

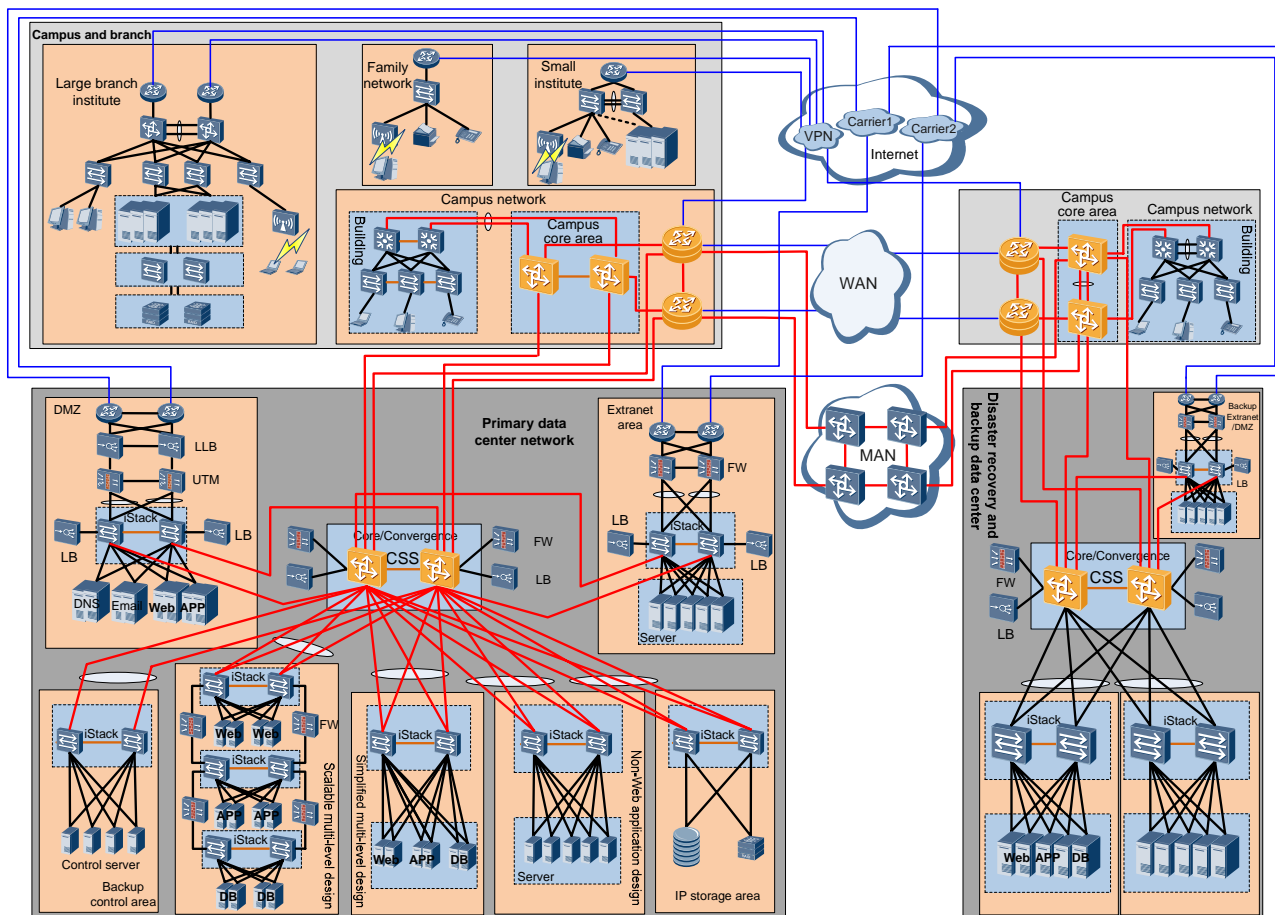
#### Area Design

Adopt OSPF dynamic route protocol in the data center. The internal routes are small in the data center; OSPF can be configured between the core/aggregation devices and the egress routers. In addition, OSPF is configured between the campus core switches and the egress routers, and configured between the core/aggregation devices in the disaster recovery center in the city. All the devices are allocated into backbone Area0. Allocate one or more network

segment addresses for each service area on the access layer. OSPF announces the routes on the core/aggregation devices to ensure reachable routes in the network.

Figure 4-9 shows the route planning for the data center.

Figure 4-9 Route planning



## COST Design

You can flexibly design COST based on the following principles:

- The distance between links and peer links
- Link bandwidth
- COST design determines network traffic direction (TE deployment is excluded), customer requirements for network traffic direction must be focused on. Before the COST design, you need to know network traffic directions in different end-to-end scenarios.

OSPF specifies the following two methods to set link COST value:

- Set COST values for the interfaces in the interface view.
- Calculate COST values based on the bandwidth in the system view.

The two methods are used for different commands. If the commands are run at the same time, the preferred method of setting COST values for the interfaces is in the interface. Reserve smaller COST values for large bandwidth to ensure that multiple 100 GE links that interconnect devices are bundled in future.

## Reliability Design

- OSPF rapid convergence

OSPF rapid convergence has the bidirectional forwarding detection (BFD) For OSPF feature.

The BFD For the OSPF feature enables to a rapid detection of link faults.

Table 4-4 describes the parameters and values for this feature:

**Table 4-4** Parameters and Their Values

Timer Parameter	Reference Value
Hello interval	10 ms
Dead interval	40 ms
spf-schedule-interval {intelligent-timer <i>max-interval start-interval hold-interval</i> }	5000 50 50
lsa-originate-interval {intelligent-timer <i>max-interval start-interval hold-interval</i> }	5000 0 20
LSA arrival interval	15 ms
Flooding-control	30 ms

- Security Design

OSPF supports the two authentication modes, interface authentication and area authentication.

- Interface authentication indicates the authentication and encryption of Hello packets sent and received by the interface.
- Area authentication indicates the authentication and encryption of Update packets in the area.

The preceding two authentication modes support simple password and MD5 authentication methods. MD5 authentication method ensures higher security.

## 4.4.3 BGP Design

### Design Principles

- Huawei recommends you use Loopback0 address for the BGP Router ID. The local interface for the Border Gateway Protocol (BGP) must be specified at the same time because the BGP Router is an indirect physical interface.
- If no router reflectors (RRs) are deployed, FULL MESH IBGP neighbor relationship between provider edges (PEs) must be created.

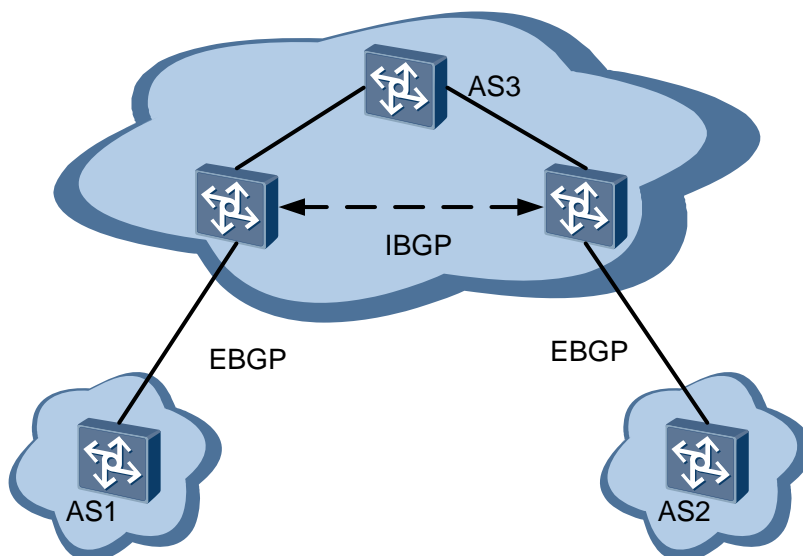
- If no router reflectors (RRs) are deployed, IBGP neighbor relationship must be created between each RR and all PEs.
- Do not use a dynamic IGP route.
- Huawei recommends you deploy BGP MD5 authentication to improve route security.
- Deploy multiple BGP peers in peer groups to simplify configurations and improve configuration efficiency.
- Aggregation routes can suppress route fluctuation and reduce BGP routers. Deploy the routes as required.
- The stability of network BGP routes can be maintained by deploying route fluctuation suppression technology.

## BGP Peer

The following lists the two types of BGP Peer:

- IBGP Peer: configured between two PEs in an authentication server (AS).
- EBGP Peer: Configured between two ASs.

**Figure 4-10** BGP Peer schematic drawing

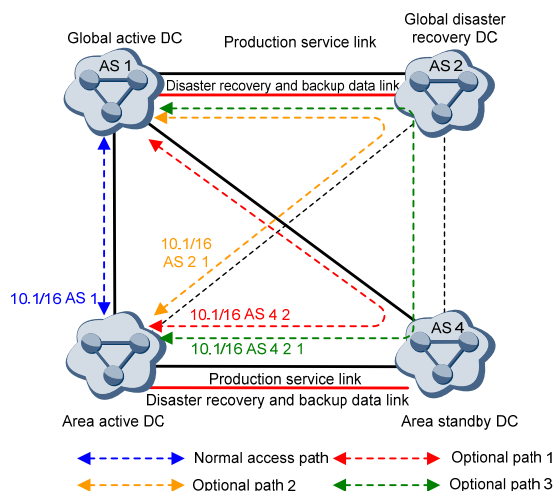


## BGP Route Notification

Route notification is implemented by deploying EBGP between the access routers in an internal network, branch data centers, and the disaster recovery center.

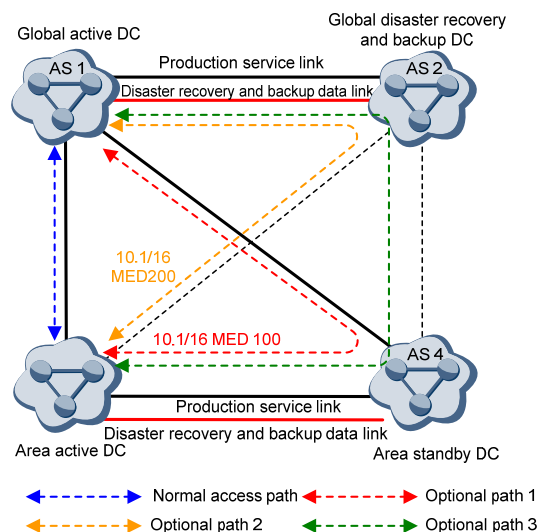


**Figure 4-11** BGP selecting an optimal path based on AS-Path



The route with short AS-Path is preferred by EBGP. In the preceding figure, AS3 receives route 10.1/16 from AS1, AS2, and AS4 respectively. The AS-Paths of the three routes are AS 1, AS 2 1, AS 4 1, and AS 4 2 1. Compared with the other two paths, the route AS-Path from AS1 is the shortest. Therefore, EBGP selects the route from AS1, that is, the route from AS1 has the highest priority. The route from AS4 2 1 has the lowest priority due to the longest AS-Path. The AS-Paths of routes from AS2 1 and AS4 1 are the same; therefore, the multi-exit discriminator (MED) property of BGP must be set to distinguish the priority.

**Figure 4-12** BGP selecting an optimal path based on MED



The MED property of route 10.1/16 from AS 4 is 100 which is smaller than that from AS2. Therefore, alternative path1 has the higher priority than alternative path2.

## Security

BGP MD5 authentication is recommended.

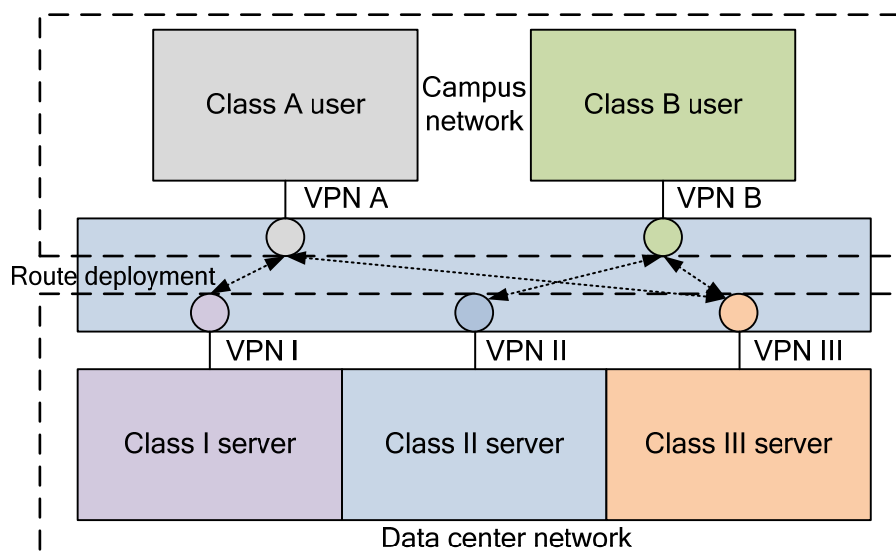
## 4.5 VPN Design

### 4.5.1 Overview

A lot of production and office servers are deployed in a data center. In most cases, the servers are deployed in the campus network in the headquarters of an enterprise. A campus network indicates the office network of an enterprise. As regional enterprise branches are excluded from the campus network, a campus network can be considered as a private network. The servers in an internal data center are classified based on information security levels. Enterprise users are classified based on responsibilities. VPNs can be created to plan the access relationship between the servers and the users.

In [Figure 4-13](#), the servers are classified into class I, class II, and class III. The users are classified into class A and class B. The users and servers are allocated into different VPNs and routes are deployed between VPNs.

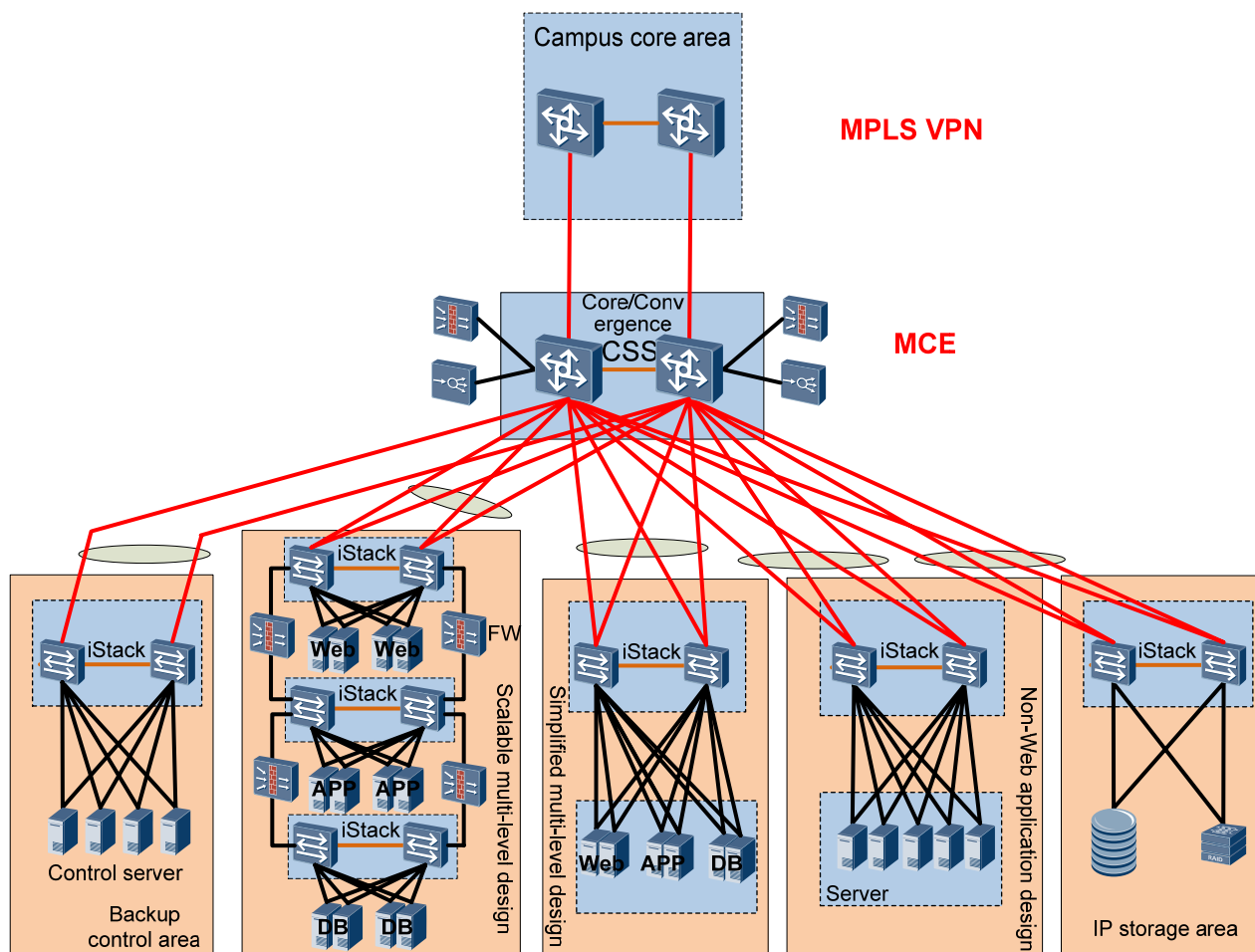
**Figure 4-13** Route-based isolation between the servers in the data center



### 4.5.2 VPN deployment

The core/aggregation devices in the data center are connected to the core devices in the campus network. In most cases, MPLS VPNs are deployed on the core devices in the campus network. MCE is deployed on the core/aggregation layer to implement service isolation. For details, see [Figure 4-14](#).

Figure 4-14 VPN deployment



## 4.6 Reliability Design

### 4.6.1 Overview

End-to-end reliability design includes the reliability design for device nodes, network topologies, and service systems. Currently, the reliability of nodes and network topologies is less important than that of service systems. The application of the service system reliability technology in specific scenarios is the focus of end-to-end reliability design.

The following lists the principles for reliability design:

- Networking layer principle: allocates a network into three layers, the core, aggregation, and access layers. Aggregate the core and aggregation layer of a small network as required.
- Backup principle: back up physical bases, such as links, devices, paths, and planes.
- Failover principle: adopt proper reliability technology to ensure failover and failover back with a minimum of packet loss when a fault occurs.

- End-to-end principle: deploy the reliability technology in an end-to-end and hierarchical mode to prevent faults from occurring.

## 4.6.2 Device Reliability

To increase reliability, a redundant device is set up for the following devices:

- Control boards
- Switching network boards
- Line cards
- Service cards
- Fan modules
- Power supply modules

All Huawei products meet the preceding reliability requirements. For details about reliability features, refer to the manuals of related products.

As an increasing number of users and devices access the data center, a single switch can no longer meet the increasing network reliability requirements. Huawei provides the following solutions to address this issue:

- Huawei S9300 core/aggregation device supports cluster switch system (CSS) function. This function connects two switches through private stack cables. The two switches are displayed as a logical switch.
- Huawei S5700 and S6700 access devices also support the stack function. A maximum of nine single devices can be connected together and displayed as logical switch to forward packets. Switch stack ensures that a large amount of data is forwarded with high network reliability.

CSS feature brings the following benefits:

- Helps customers maximize return on investment during network expansion.
- Virtualizes two physical devices to a logical one during network expansion, simplifying device configuration and management.
- Improves system reliability with device redundancy and backup.

Figure 4-15 shows the reliability design for devices.

**Figure 4-15** Reliability design for devices

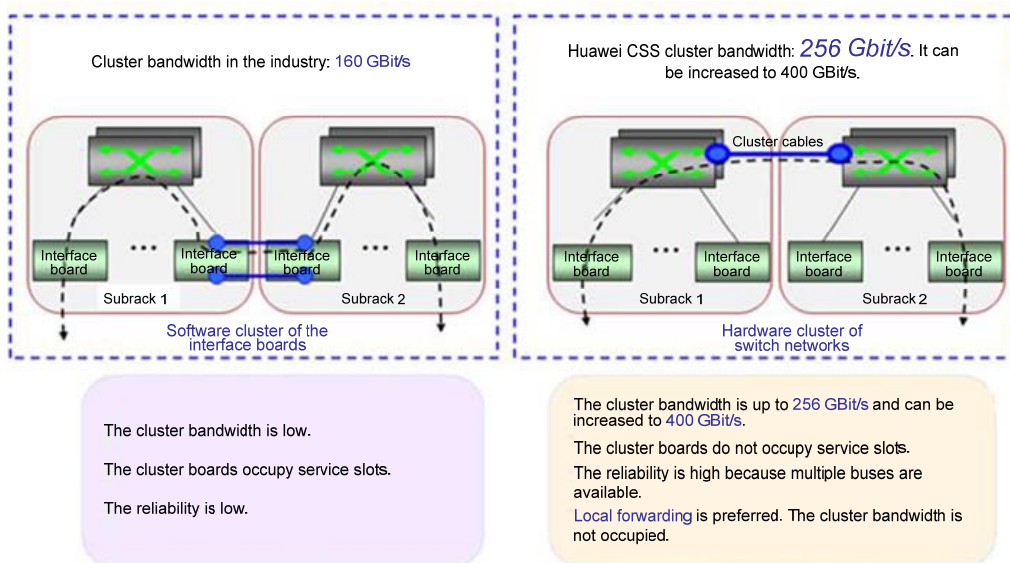


Table 4-5 describes the non-blocking stack counters of access switches.

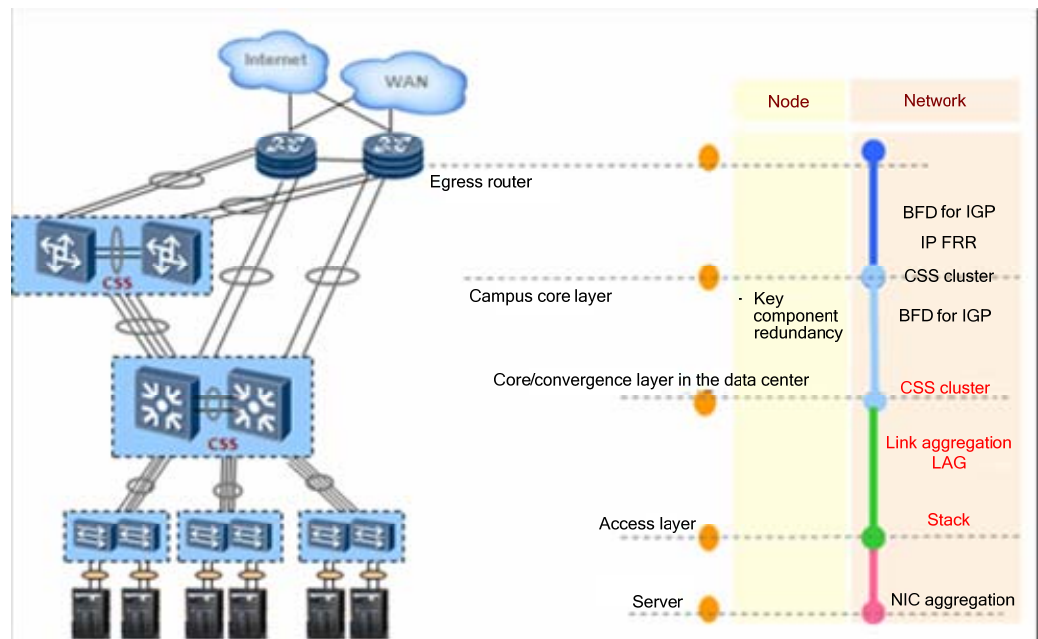
**Table 4-5** Stack counters of access switches

Item	S5700 Series	S6700 Series
Supportable topology	Link and loop	Link and loop
Stack switch capacity	48 Gbit/s	80 Gbit/s
Maximum number of stacked devices	9	9
Stack interface	Private stack interface	Configured common interfaces
Stack cable	57 private cable	Optical fiber

### 4.6.3 Network Reliability

Redundant network topologies are required for upper layer reliability, in addition to the network reliability. Select network topologies based on the reliability and resources available to meet customer requirements. Figure 4-16 shows a network topology.

Figure 4-16 Network topology



- The active and standby servers in the NICs are deployed.  
Dual uplinks are deployed for each server to ensure high availability.  
NIC teaming indicates that two NICs are bundled to a virtual one.  
The two NICs use the same IP address and MAC address and work in load balancing mode. They forward traffic at the same time doubling the data bandwidth.
- Link aggregation (LAG) is adopted between the core/aggregation device (as a server gateway) and access devices to protect the links.  
The following lists the advantages of link aggregation:
  - Increased link bandwidth
  - Link load balancing
  - Improved link reliability by the backup for links of group members
 IGP is configured between core/aggregation devices in the data center and the core devices and egress routers in the campus network. BFD for IGP or IP FRR technology is adopted to provide reliability protection. LAG is adopted to protect links.

## 4.6.4 Service Reliability

Service reliability covers service software reliability, server cluster reliability, and the reliability of service handover by translating a domain name to the corresponding IP address by DNS.

- Service software reliability: prevents a software operation fault from causing a task failure, or worse.
- Server cluster reliability: multiple servers are clustered and displayed as a server to provide specific services. The cluster can use multiple computers to operate in parallel resulting in higher speed computing. In addition, the cluster can use multiple computers as backup. When a device in the cluster failed, the cluster still operates and provides

services. The cluster operation can reduce service interruption due to single point failure and ensure high availability of cluster resources.

The reliability of service handover by translating a domain name to the corresponding IP address by DNS: adopted when all server clusters failed. For details, see section 4.3.3 "DNS Design".

## 4.7 Load Balancing Design

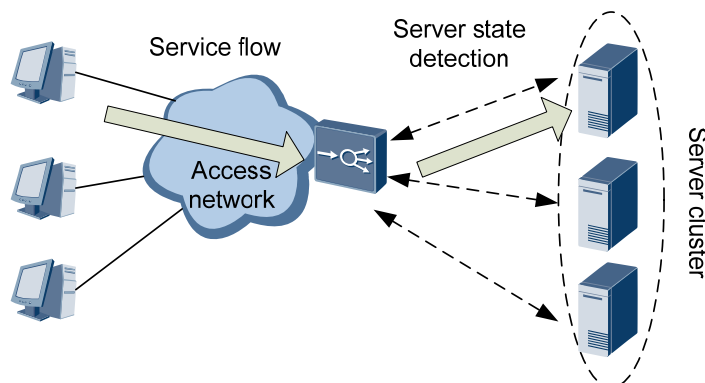
### 4.7.1 Overview

Generally, a LB is called as a L4 or L7 switch.

- A L4 switch analyzes the IP layer and Transmission Control Protocol/User Datagram Protocol (TCP/UDP) layer to ensure traffic load balancing on L4.
- A L7 switch supports L4 load balancing, in addition, it analyzes application layer information, such as HTTP URL and Cookie information.

Figure 4-17 shows the LB function.

Figure 4-17 LB function



### 4.7.2 Design Principles

The following lists the principles for LB design:

- As a solution for network bottlenecking, server LBs must support high throughput.
- LBs support multiple load balancing algorithms, including polling and IP-based and content-based Hash, to ensure the rationality of load balancing.
- LBs support reliability check and feature ease-of-expansion to ensure system redundancy.
- LBs forward the traffic of the loaded server clusters; they must be deployed on the core/aggregation layer to prevent from traffic bottleneck.

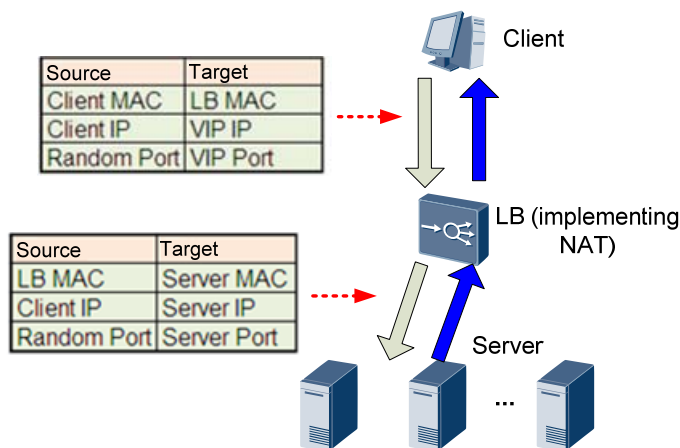
## 4.7.3 LB Deployment Modes

### Symmetry mode

The LB translates destination address of the traffic from the client to the server and translates the source address of the traffic from the server to the client.

Figure 4-18 shows the deployment schematic drawing.

Figure 4-18 LB deployment in symmetry mode



Advantage: The LB can be deployed on the aggregation or the core layer. The LB controls incoming and outgoing traffic and implements control policies in real time.

Disadvantage: The incoming and outgoing traffic may cause a bottleneck and requires high LB performance.

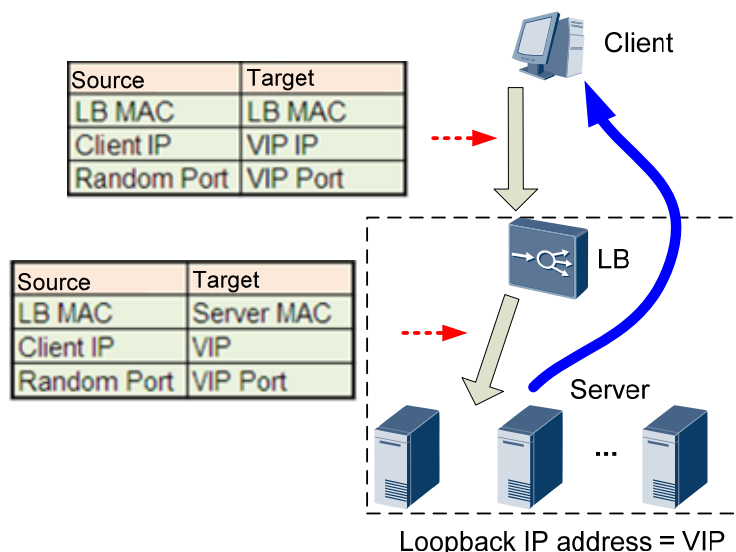
### Asymmetry mode

The LB changes the destination MAC address of the traffic from the client to the server to the server's MAC address, instead of translating the IP address. The traffic from the server to the client does not pass through the LB. The Loopback address must be configured as a VIP address on the server.

Figure 4-19 shows the LB deployment in asymmetry mode.



**Figure 4-19** LB deployment in asymmetry mode



**Advantage:** The traffic from the client to the server does not pass through the LB. Compared to the symmetry mode, this mode requires lower LB performance. This mode is suitable for large-throughput video distribution services.

**Disadvantage:** The LB must be deployed on the core switch layer, and the LB cannot count or charge the traffic because the traffic from the client to the server does not pass through it.

## 4.8 QoS Design

### 4.8.1 Overview

A data center requires a different QoS when bearing data services. The IP network must distinguish the data packets, color them, and provide congestion management, congestion avoidance, traffic policing, and traffic shaping. Using these methods, the network devices can provide specific services for each customer.


QoS services cover the three models, best-effort forwarding, integration service, and DiffServ service. Huawei uses the DiffServ model for the data center.

The data center network covers multiple types of services, including high-priority services and mid- and low-priority services. Congestion or delay may occur on certain nodes due to the bandwidth restriction. To ensure that high-priority services get prioritized handling when the network involves congestion or delay, a strategy for handling bandwidth and priority must be planned for each service type.

### 4.8.2 Service QoS

In most cases, a data center can endure burst traffic; QoS is not required.

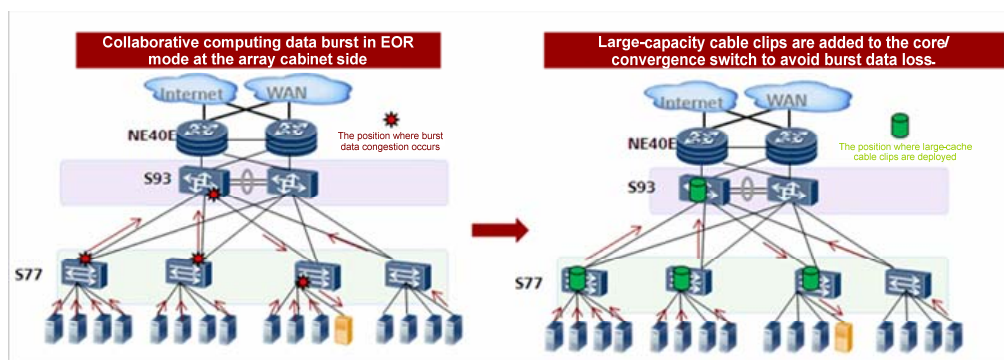
In collaborative computing service scenarios, such as search engine, oil exploration, and meteorological computing, the computing tasks must be collaboratively handled by multiple servers. Multiple servers may send computing results to the same server and the burst traffic may cause data packet loss on a port due to congestion.

The blue servers in [Figure 4-20](#) send response to the yellow servers. Congestion occurs at the points marked . If the forwarding queues on the network nodes are full, packets are lost.

The problem can be solved by using large-capacity line cards on the EOR switch and core switch. Large-capacity line cards cache burst data to prevent packet loss.

Currently, large-capacity line cards support a 48 GE optical interface (G48SBC) and a 48 GE electrical interface (G48TBC). Each line card has 1.2 GB memory and is deployed in the downstream access switch.

**Figure 4-20** Data burst



---

# 5 Network Management Design

---

## 5.1 Overview

### 5.1.1 NMS

An increasing amount of network resources are deployed in the data center due to the increase of enterprise services. As a result, the network scale is larger and larger. This requires efficient end-to-end network management. An NMS manages the network elements (Nes) and the overall network, including network resources, the topology, faults, configurations, performance, and reports. An NMS efficiently manages the network and reduces the workload for network maintenance personnel.

### 5.1.2 Network Scale

When evaluating network scale, consider the following concept:

- Equivalent network element (NE)  
NEs support different functions and features, different cross-connect capacities, and a different number of boards, ports, and channels, and occupy different NMS resources. The maximum number of NEs that can be managed depends on the NE types.  
The equivalent network element is a unified computing standard by converting NE types and the number of ports into equivalent NEs based on occupied system resources.
- Equivalent coefficient  
Equivalent coefficient = Resources occupied by physical NEs or ports/System resources occupied by equivalent NEs
- Network scale  
Refer to the NE equivalent coefficients to calculate the number of equivalent NEs in each type and the network scale. Reserve certain network capacities for future expansion. If there is no expansion plan, reserve capacities based on the 1:0.6 ratio. The calculation formula for network scale: Planned network scale = Current network scale + Reserved network scale.

The following lists the mapping between network scale types and the number of equivalent NEs:

- Small network: less than 2000 equivalent NEs
- Middle network: 2000 to 6000 equivalent NEs
- Large network: 6000 to 15000 equivalent NEs

- Very large network: 15000 to 20000 equivalent NEs

### 5.1.3 NMS Design

The NMS has a widely-used architecture. Inband or outband networking mode is adopted for the communication between servers and NEs. Therefore, a server fault does not affect the managed device networking or the services provided by these devices. The following lists the details of the two networking modes:

- Inband networking

Inband networking indicates that the NMS uses the service channels of the managed device to transmit NMS information to manage the network. The NMS interaction information is transmitted through the service channels of the managed device.

- Advantage: flexible and simplified networking and cost-efficiency.
- Disadvantage: When the network fails, the NMS cannot maintain the network because the information channels between the NMS and the managed network are interrupted.

- Outband networking

Outband networking indicates that the NMS uses the service channels provided by the other devices, instead of the managed device, to transmit NMS information to manage the network. Generally, the management interfaces on the control boards of the managed device are used as the access interface.

- Advantage: The NMS is not directly connected to the managed devices, but connected to the managed devices through other devices. Compared with the inband networking mode, the outband networking mode provides more reliable device management channels. When the managed device fails, the NMS can locate the device information in a timely fashion and monitor the device in real time.
- Disadvantage: The networking is costly. The NMS manages the devices by creating a network that consists of the non-managed devices. The created network provides maintenance channels which are irrelevant to the service channels.

Huawei provides eSight network management system based on the network scale and application features of enterprise data centers.

## 5.2 eSight System Design

### 5.2.1 Overview

eSight, the network management system for a data center, is in the B/S architecture. Therefore, the system is updated or maintained by only updating the software on the server. This reduces the cost and workload of system maintenance and upgrade. The B/S architecture has the following advantages:

- This architecture is in distributed mode. Users can query and view information anytime and anywhere.
- Services can be easily extended. Server functions can be added by adding Web pages.
- Maintenance is easy. The information can be updated for all users by modifying Web pages.

eSight supports the following functions:

- Security management
- Log management
- NE access
- Topology management
- Alarm management
- Performance management
- Report management
- Configuration file management
- Hierarchical NMS management

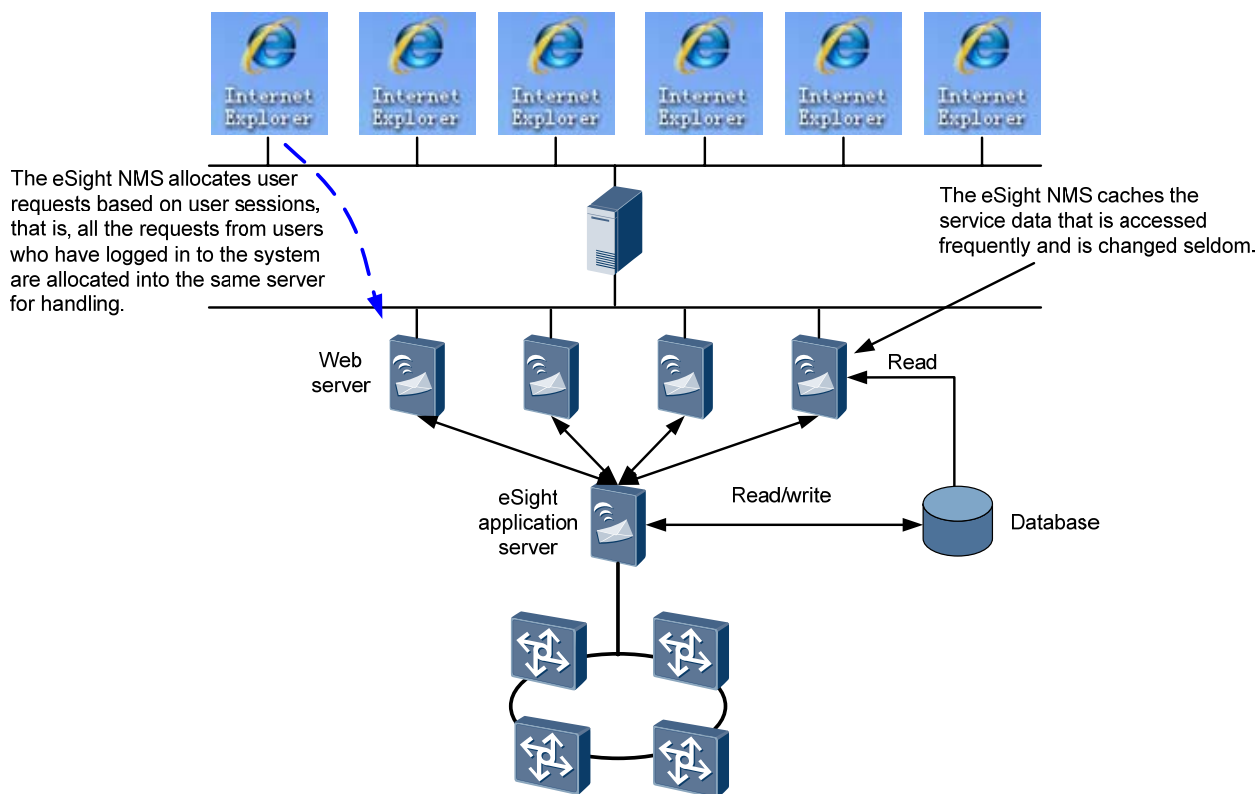
These functions meet the requirements for enterprise-class data center services.

## 5.2.2 Considerations

eSight supports the single-server mode and hierarchy-deployment mode. Generally, enterprise data centers are not managed based on areas or hierarchies. Therefore, Huawei recommends a single-server mode. Multiple browsers can access the eSight at the same time. eSight adopts LBs to handle requests from multiple users and allocates the requests to different Web servers. Service components of the eSight are deployed on the same server.

Figure 5-1 shows how the eSight works.

**Figure 5-1** Working mode of the eSight



eSight is deployed with a scalable architecture and multiple modules. It can manage each data network and the entire network. In addition, the eSight can manage the Huawei and non-Huawei devices commonly used in industry.

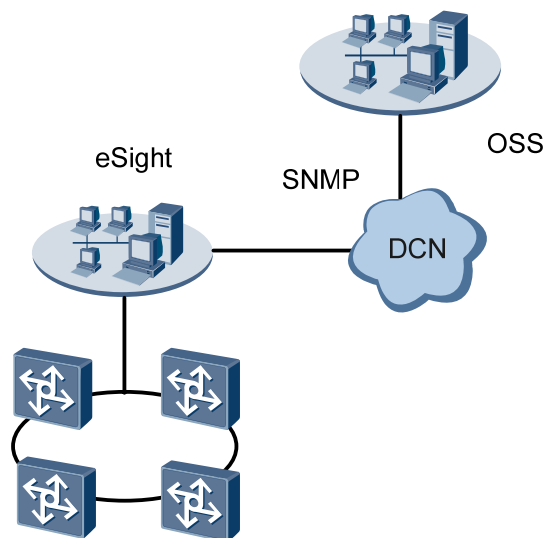
**Table 5-1** Fields and relevant devices

Field	Device
Switch	S9300, S7700, S2700, S3700, S5700, S6300, and S6700
NE series of routers	8090 series of routers: NE40E, NE40E-4, NE40E-X3, and NE80E 8011 series of routers: NE40 and NE80 8070 series of routers: NE20E-8 and NE20-2/4/8
AR series of routers	AR1220, AR1220 W, AR1240, AR1240 W, AR2220, AR 2240, and AR3260
Security device	Eudemon security devices: Eudemon8000E, Eudemon1000E, E200E-B/C/F, E100E, and E200S SRG security devices: SRG 2200 SVN security devices: SVN 3000
Third-party device	Pre-integrated third-party device, H3C, and Cisco, printer, and server

eSight can be integrated with the OSS. eSight adopts SNMP to report network alarms. This enables eSight to interconnect with the OSS alarm system.

Figure 5-2 shows the networking when eSight is integrated with the OSS alarm system.

**Figure 5-2** Networking used when eSight is integrated with the OSS



The following lists the advantages of integrating the eSight with the OSS:

- Improved network management capability
- Separation of NE management and network management
- Ensured enterprise O&M mechanisms

### 5.2.3 Design Principles

eSight network management capability varies with NE types. Consider the following factors when designing an eSight NMS:

- The number of used fiber cables and deployed services varies with NE type. In addition, NEs support different database capacity.
- eSight in base, standard, and professional versions have different management compatibilities.
- The management compatibility is affected by hardware platforms on which eSight is deployed.

Table 5-2 describes the differences between the management capabilities of eSight in the three versions on various hardware platforms:

**Table 5-2** Differences of management competency of eSight in different versions

Version	Subsystem	Management Capability	PC Server/PC
Base version	NE access	Less than 100 NEs	E5300
Standard version	NE access	Less than 2000 NEs	IBM X3650M3-2*Xeon-four CPUs (5405 2.26 GB) or more-16G(4*4G)-5*146G
	IPSec	Less than 500 NEs	
	WLAN	Less than 1000 APs	
Standard version	NE access	Less than 5000 NEs	IBM X3850X5-4*XEON-eight CPUs (7430 2.26G) or more-32G-8*73.4G
	IPSec	Less than 1500 NEs	
	WLAN	Less than 2500 APs	
Professional base	NE access	Less than 20000 NEs	IBM X3850X5-4*XEON-eight CPUs (7430 2.26 GB) or more-32 GB-8*73.4 GB
	IPSec	Less than 5000 NEs	
	WLAN	Less than 10000 NEs	

### 5.2.4 Design Elements

Table 5-3 lists the network management solutions provided by the three editions of eSight based on different network scales and functions.

**Table 5-3** Versions and corresponding functions

Version	Function
Basic edition	Supports the following functions: <ul style="list-style-type: none"><li>• NE access management</li><li>• Topology management</li><li>• Security management</li><li>• Alarm management</li><li>• Performance management</li><li>• NE management</li><li>• Configuration file management</li><li>• Device customization</li><li>• Inventory management</li><li>• Data dump and backup</li></ul>
Standard edition	Supports the following functions: <ul style="list-style-type: none"><li>• All functions of the basic version</li><li>• Report management, intelligent tool configuration</li><li>• The integration of service management components, such as WLAN service management</li><li>• Used as a lower-layer NMS</li></ul>
Professional edition	Supports the following functions: <ul style="list-style-type: none"><li>• All functions of the standard edition</li><li>• Used as an upper-layer NMS</li></ul>

## IP Address Planning for the NMS Servers

The IP address planning must be complied with the following principles:

- The IP address must be unique in the network.
- The servers can communicate with the managed devices.
- The servers can communicate with the clients.
- Only one IP address can be assigned to a network port.
- The IP address of a hardware control device, such as a disk array controller, must be planned based on the hardware on the NMS servers.
- The IP address of the NMS must be planned based on the NMS deployment solutions.

## Port Planning for the NMS Servers

FWs filter traffic based on the IP address and the port number of TCP/UDP.

The system separates data packets based on the port number of TCP/UDP and sends the data packets to the appropriate application programs.

The port number range (0–65535) of TCP/UDP is allocated into the following three segments:



- 0–1023: identify standard services, such as FTP, Telnet, and Trivial File Transfer Protocol (TFTP).
- 1024–49151: allocated by the Internet assigned number authority (IANA) to registered applications.
- 49152–65535: dynamically allocated to applications as private port numbers.

When FWs are deployed between the eSight server and NEs, clients, or the OSS, ports need to be deployed to facilitate the connection between the eSight server and these devices.

Table 5-4 shows the details for deploying ports on the device side of eSight.

**Table 5-4** Port details

Source IP Address	Destination IP Address	Protocol	Source Port	Destination Port	Description
U2000 server	Any SNMP device	UDP	Any port	161	U2000 server issues commands to port 161 of SNMP.
U2000 server	Any SNMP device	TCP	Any port	22 23	U2000 server issues Telnet request to port 23 of SNMP.
U2000 server	Any SNMP device	TCP	Any port	1400	Used to listen to the NMS. Used as the port through which a command line is issued to log in to the NMS.
U2000 server	Any SNMP device	UDP	Any port	1500	Used as the port through which the NMS automatically collects port information. Used to monitor the collection process.
U2000 server	Any SNMP device	TCP	Any port	5432	SSL command line is issued to log in to the NMS.

During eSight design, an appropriate version must be selected based on network scale.

Table 5-5 provides a reference for selecting a proper eSight edition.

**Table 5-5** Selecting NMS Versions

Item	Sub-item	Basic Edition	Standard Edition	Professional Edition
Management capability	Number of managed NEs	100	5000	20000
	Number of clients	10	100	200
Occupied resource	Memory usage	512 MB	1 GB	2 GB
	CPU usage		More than 30% occupancy within no more than 15 minutes	More than 50% occupancy within no more than 15 minutes
Storage capacity	Current alarm capacity	20000	20000	20000
	Historical alarm capacity		1.5 million	15 million
	Log data capacity	1 million	1 million	1 million
	Performance data capacity			60 million
Handling capability	Alarm response speed	No more than 30s	No more than 30s	No more than 30s
	Performance response speed		Collecting 30000 performance data within 15 minutes	Collecting 30000 performance data within 15 minutes
	Time of response to page operations	3s	3s	3s
Status update duration	Device status		No more than 30s	No more than 60s
	Link status		No more than 5s	No more than 5s